Small cEllS coordinAtion for Multi-tenancy and Edge services

Grant Agreement No.671596

Topic: H2020-2014-ICT-14
Advanced 5G Network Infrastructure for the Future Internet
Research and Innovation Action

Deliverable D3.1

CESC Prototype design specifications and initial studies on Self-X and virtualization aspects

Document Number:	H2020-5GPPP-GA No.671596*/WP3/D3.1/30.12.2016*
Contractual Date of Delivery:	*30.06.2016*
Editor:	Alan Whitehead – IP.Access Ltd. (IPA)
Work-package:	*WP3*
Distribution / Type:	Public (PU) / Report (R)
Version:	*1.0*
Total Number of Pages:	175
File:	*SESAME_Deliverable 3.1_v1.0_Final*

Abstract

SESAME aims to build a cloud-*enabled* platform that supports both radio access and edge computational services at one Point of Presence (PoP). Network Services are supported by Virtual Network Functions hosted in the Light DC, leveraging on technologies like SDN and NFV that allow achieving an adequate level of flexibility and scalability at the cloud infrastructure edge.

The CESC design splits radio functions in order to decouple software functions from the underlying hardware in order to realise the SESAME concept.

The contents of this deliverable will be used as a reference for the implementation of the Small Cell Prototype in D3.4.

In addition, it presents the results initial studies into Self-X and Virtualisation.

Version History

Version	Date	Comments, Changes, Status	Authors, contributors, reviewers
0.1		First Draft	
0.2	7th June	Contributions to sections 3.5, 3.6, 9.2	UPC
0.3	8th June	Contributions to multiple sections	IPA
0.4	9th June	Additional contributions	
0.5	10th June	Contribution from CreateNet (update)	CNET
0.6	17th June	Contributions from several partners	IPA
0.7	22nd June	Added PM description	IPA
0.8	22nd June	Additional detail on EMS and management	IPA
0.9	30th June	Restructured to remove duplication Added introduction and conclusion Added Figure and Table cross-refs	IPA
0.10	30th June	Final tidy-up	IPA
1.0	11th July	Final editorial and conceptual review performed by the project coordinator. Completion of the document and submission to the Commission.	OTE

Contributors

First Name	Last Name	Partner	Email
Alan	Whitehead	IP.Access	alan.whitehead@ipaccess.com
David	Brock	IP.Access	David.Brock@ipaccess.com
Kamesh	Kaul	IP.Access	kamesh.kaul@ipaccess.com
Kiran	Chackravaram	IP.Access	Kiran.Chackravaram@ipaccess.com
Jordi	Pérez-Romero	UPC	jorperez@tsc.upc.edu
Oriol	Sallent	UPC	sallent@tsc.upc.edu
Ramon	Agustí	UPC	ramon@tsc.upc.edu
Juan	Sánchez-González	UPC	juansanchez@tsc.upc.edu
Ramon	Ferrús	UPC	ferrus@tsc.upc.edu
August	Betzler	i2CAT	august.betzler@i2cat.net
Daniel	Camps Mur	i2CAT	dani.camps@i2cat.net
Pouria	Sayyad Khodashenas	i2CAT	pouria.khodashenas@i2cat.net
Cristina	Ruiz	i2CAT	cristina.ruiz@i2cat.net
Jordi	Ferrer Riera	i2CAT	jordi.ferrer@i2cat.net
Eduard	Escalona	i2CAT	eduard.escalona@i2cat.net
Jose Oscar	Fajardo	EHU	Joseoscar.fajardo@ehu.eus
Fidel	Liberal	EHU	fidel.liberal@ehu.eus
Daniele	Munaretto	ATH	Daniele.munaretto@athonet.com
Massimo	Gallina	ATH	Massimo.gallina@athonet.com
Shah Nawaz	Khan	CNET	s.khan@create-net.org
Babangida	Abubakar	UoB	B.abubakar2@brighton.ac.uk
Haralambos	Mouratidis	UoB	H.Mouratidis@brighton.ac.uk
Emmanouil	Panaousis	UoB	E.Panaousis@brighton.ac.uk
Nasr	Karim	UoS	k.nasr@surrey.ac.uk
Vahid	Seiamak	UoS	s.vahid@surrey.ac.uk
Moessner	Klaus	UoS	k.moessner@surrey.ac.uk
Leonardo	Goratti	CNET	leonardo.goratti@create-net.org
Ioannis	Chochliouros	OTE	ichochliouros@oteresearch.gr

Glossary

Acronym	Explanation
2D	Two-dimensional
3D	Three-dimensional
3GPP	Third Generation Partnership Project
4G	Fourth Generation of Mobile Communications
5G	Fifth Generation of Mobile Communications
ABS	Almost Blank Subframes
AC	Admission Control
ACK	Acknowledgement
ACM	Association for Computing Machinery
AI	Artificial Intelligence
AIAI	Artificial Intelligence Applications and Innovation
AIC	Akaike Information Criterion
AMBR	Aggregate Maximum Bit Rate
ANR	Automatic Neighbour Relations *Also referred to as Network Listen (NWL). A process by which a cell scans its radio environment to discover neighbouring cells.*
AP	Application Protocol
API	Application Programming Interface
ARIMA	Auto-Regressive Integrated Moving Average
ARM	Advanced RISC Machine
ARP	Allocation and Retention Priority
BS	Base Station
BW	Bandwidth
CDN	Content Delivery Networking
CDR	Call Dropping Rate
CESC	Cloud-Enabled Small Cells
CGI	Common Gateway Interface
CIR	Carrier to Interference Ratio
CLARANS	Clustering LARge Applications
CM	Configuration Management
CMAS	California Mutiple Awards Schedule
CN	Core Network
CORBA	Common Object Request Broker Architecture
CP	Control Plane
CPE	Customer Premises Equipment
CPRI	Common Public Radio Interface
CPU	Central Processing Unit
CQI	Channel Quality Indicator
CRE	Cell Range Expansion
cSON	centralised SON
CSP	Communications Service Provider
D2D, D-2-D	Device-to-Device
D-SON	Distributed SON
dB	decibel
DC	Data Centre
DDoS	Distributed Denial of Service
DL	Downlink
DN	Distinguished Name *An identifier that uniquely identifies a managed object instance within the*

	scope of the system.
DoS	Denial of Service
DRX	Discontinuous Reception
DSCP	Differentiated Services Code Point
dSON	distributed SON
E2E	End-to-End
E-RAB	E-UTRAN Radio Access Bearer
E-UTRA	Evolved Universal Terrestrial Radio Access
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
ECGI	Experiment Computing Grid Integration
EMS	Element Management System
eNB	Enhanced Node B
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
ETWS	Earthquake and Tsunami Warning Service
FAPI	Functional Auditory Performance Indicator
FFR	Fractional Frequency Reuse
FFS	Fast File System
FM	Fault Management
FPGA	Field Programmable Gate Array
FQDN	Fully Qualified Domain Name
FR	Frequency Reuse
FT	File Transfer
FTP, ftp	File Transfer Protocol
GA	Grant Agreement
GbE	GigaBit Ethernet
GBR	Guaranteed Bit Rate
GHz	Giga Hertz
GPRS	General Packet Radio Service
GPU	Graphics Processing Unit
GPZ	Geo-ProZones
GS	Group Specification
GSM	Global System for Mobile communications
GSMA	GSM Association
GTP	GPRS Tunnelling Protocol
GUI	Graphical User Interface
GUTI	Globally Unique Temporary Identity
GW	Gateway
H2020	Horizon 2020
HeNB	Home eNB
HO	Handover
HTTP, http	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
HW	Hardware
ICC	International Conference on Communications
ICIC	Inter-Cell Interference Coordination
ICT	Information and Communication Technology
ICWMC	International Conference on Wireless and Mobile Communications
ID, id	Identifier
IE	Information Element
IEEE	Institute of Electrical and Electronic Engineers
IFIP	International Federation for Information Processing
IMIX	Internet Mix

IP	Internet Protocol
IPsec, IPSec	Internet Protocol Security
IPSN	Information Processing in Sensor Networks
IPv4	Internet protocol version 4
IPv6	Internet protocol version 6
IRAT	Inter-Radio Access Technology
IRP	Integration Reference Point
IS	Information Service
ISP	Internet Service Provider
ISWCS	International Symposium on Wireless Communication Systems
IT	Information Technology
ITU	International Telecommunication Union
ITU-T	ITU –– Telecommunications Standardization Sector
ITW	Information Theory Workshop
IWF	Interworking Function
k-NN	k-Nearest Neighbour
KDE	Kernel Density Estimation
KPI	Key Performance Indicator
KVM	Kernel-*based* Virtual Machine
L1	Physical Layer
L2	Data Link Layer
LAN	Local Area Network
LOS	Line of Sight
LTE	Long Term Evolution
LTE-A	LTE Advanced
LTE-U	Unlicensed LTE
MAC	Medium Access Control
MANO	Management and Orchestration
MBR	Maximum Bit Rate
MCC	Mobile Country Code
MDT	Minimization of Drive Tests
MEC	Mobile Edge Computing
MIB	Management Information Base
MLB	Mobility Load Balancing
MLDM	Machine Learning and Data Mining
MME	Mobility Management Entity
MNC	Mobile Network Code
MNO	Mobile Network Operator
MOCN	Multi Operator Core Network -
MRO	Mobility Robustness Optimisation
ms	micro second
NAS	Non-Access Stratum
NFV	Nertwork Function Virtualization
NFV	Nertwork Functions Virtualization
NFVO	NFV Orchestrator
NNSF	Non-Access-Stratum Node Selection Function
NMS	Network Management System
NO	Network Operator
NOS	Network Orchestration System
NP	Non-Polynomial
NR	Neighbour Relation
NRM	Network Resource Model
NRT	Neighbour Relation Table

NS	Network Service
NSF	Node Selection Function
NSH	Network Service Header
OFDMA	Orthogonal Frequency-Division Multiple Access
OS	Operating System
OTT	Over-The-Top
P2P	Point-to-Point
PAM	Partitioning Around Medoids
PCI	Peripheral Component Interconnect
PCIe	Peripheral Component Interconnect Express
PDCP	Packet Data Convergence Protocol
PHY	Physical Layer
PLMN	Public Land Mobile Network
PM	Performance Management
PNF	Physical Network Function
PoC	Proof of Concept
PPP	Poisson Point Process
PPP	Public-Private Partnership
PRB	Physical Resource Block
PS	Packet Scheduling
QCI	QoS Class Identifier
QoS	Quality of Service
RAB	Radio Access Bearer
RACH	Random Access Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RAT	Radio Access Terminal
RB	Resource Block
REA	Research Executive Agency
REST	Representational State Transfer
RIA	Research and Innovation Action
RLC	Radio Link Control
RP-ABS	Reduce-Power ABS
RRC	Radio Resource Control
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
S1AP	S1 Application Protocol
S-GW	Serving Gateway
SC	Small Cell
SC-C-VNF	SC-Common-VNF
SC-PNF	Small Cell Physical Network Function
SCaaS	Small Cell as a Service
SCNO	Small Cell Network Operator. *The "landlord" operator that owns the network infrastructure.*
SCTP	Stream Control Transmission Protocol
SDU	Service Data Unit
SDN	Software-Defined Radio
SFC	Service Function Chaining
SFTP	SSH File Transfer Protocol
SI	System Information
SIB	System Information Block
SIB	System Information Broadcast

SIGMOD	Special Interest Group on Management of Data
SINR	Signal to Interference and Noise Ratio
SLA	Service Level Agreement
SNMP	Simple Network Management Protocol
SOAP	Simple Object Access Protocol
SoC	System on Chip
SOM	Self-Organizing Map
SON	Self-Organizing Network
SOTA	State-of-the-Art
SP	Service Provider
SRB2	Sonic Robo Blast 2
SS	Solution Set
SSH	Secure Shell
SVM	Support Vector Machine
SVR	Support Vector Regression
SW	Software
TA	Tracking Area
TAI	Tracking Area Identity
TB	Transport Block
TCP	Transmission Control Protocol
TEID	Tunnel Endpoint Identifier
TPS	Transactions Per Second
TR	Technical Report
TS	Technical Specification
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
UP	User Plane
URL, url	Uniform Resource Locator
UTRAN	Universal Terrestrial Radio Access Network
vCPU	virtual Central Processing Unit
VIM	Virtual Infrastructure Manager
VM	Virtual Machine
VNF	Virtual Network Function
VNFM	Virtual Network Function Manager
VNFO	Virtual Network Function Orchestrator
VSCNO	Virtual Small Cell Network Operator *A "tenant" operator that makes use of the network infrastructure owned by the SCNO. Note that the SCNO may also be a VSCNO.*
VTC	Vehicular Technology Conference
WAN	Wide Area Network
WiOpt	Modelling and Optimization in Mobile, Ad Hoc and Wireless Networks
WP	Work Package
WSDL	Web Services Definition Language *An XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information.*
X2AP	X2 Application Protocol
XML, xML	eXtensible Markup Language

Table of Contents

List of Figures

List of Tables

1 Introduction

To realise SESAME's core vision, that is "Cloud-Enabled Small Cells" (CESCs), will be designed, developed and implemented in order to offer access to network capacity coupled with mobile edge computing (MEC) resources in a single device. These resources will be offered on-demand to Communications Service Providers (CSPs), profiling both access and edge computation resources to satisfy the specific CSPs' needs.

The overall SESAME architecture has been defined in D2.2 [62] and the component interactions have been outlined in D2.3 [63]. This current document expands further on the functions contained within the CESC, from both the functional and implementation viewpoints.

This document represents the completion of task 3.1, covering the design of a Proof-of-Concept (PoC) demonstrator of the SESAME concept. It also contains an intermediate report of the work completed so far, regarding "how the Self-X features may be used in a SESAME environment" (task 3.2) and; regarding "how the virtualisation of radio resources may be used to further enhance the ability to slice the CESC resources between the tenant operators" (task 3.3). The remainder of this document is organised as follows:

Section 2 focusses on the CESC prototype design, based on splitting radio functions in order to decouple software functions from the underlying hardware. This is the "S1 functional split" proposed in D2.3, though in some areas other splits are explored for comparison and to show the future flexibility of the overall architecture.

Section 3 provides an update on the initial Self-X studies, exploring the possibilities enabled by the overall SESAME architecture, though in some cases beyond the capabilities of the current prototype implementation.

While *Section 2* covers the functional aspects of resource slicing between tenants and resource management within a tenant performed by the various VNFs, *Section 4* provides a brief description of the relationship to the VNFO, though as noted there, the bulk of this work is covered in other deliverables.

2 CESC Prototype Design

2.1 Architecture

The overall SESAME architecture is described in D2.3 and presented in Figure 1 below:

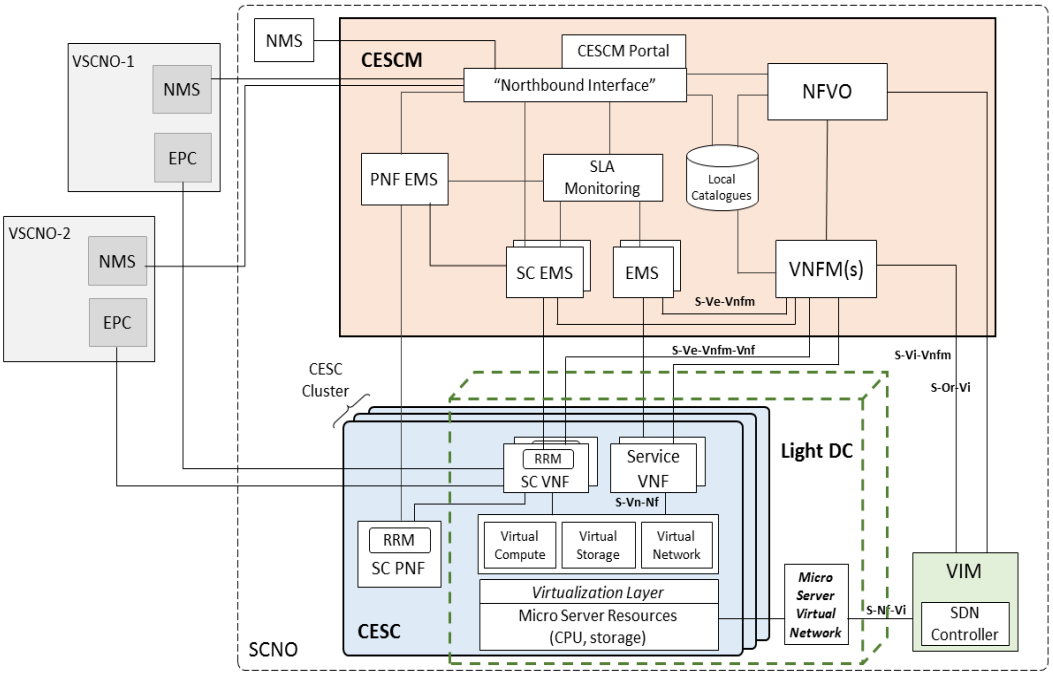


Figure 1: Overall SESAME Architecture

2.1.1 The CESC and the Light DC

The CESC is the fundamental building block of the SESAME architecture providing the virtualization platform for hosting SC-VNF, Service VNFs and an implementation of PNF corresponding to a particular chosen functional split. However, the hardware resources of a single CESC are low considering its low cost and the number of functional entities that the underlying processing power can support for multi-tenants is limited, thereby limiting the scope of MEC. However, the close proximity of small cells provides an opportunity to aggregate these resources into CESC clusters (Light DC) that can support a diverse set of MEC use cases. The Light DC is envisioned to provide the necessary virtualization resources to realize the objectives of SESAME.

To support the required network functions of a small cell and edge services in a virtualized environment, the light DC leverages the hardware (HW) and software (SW) resources of the CESC clusters. The light DC provides the necessary computing, networking and storage resources on top of a virtualization layer. The light DC is essentially a platform for realizing the benefits of SDN/NFV at the network edge. The network functions within a light DC are implemented as "VNFs" which may be hosted on top of virtual machines that are hosted on top of a hypervisor layer (e.g. KVM[1]-*based* solutions) or as "Containers".

[1] KVM is a full virtualization solution for Linux on x86 hardware containing virtualization extensions (Intel VT or AMD-V). More information can be found at: http://www.linux-kvm.org/page/Main_Page

The hardware architecture is based on a System on Chip (SoC) that includes multi core ARMv8[2] 64bit Central Processing Unit enabling not only efficient processing but also keeping the cost of the hardware low. The CPU is supported by dedicated hardware accelerators for networking and packet processing. The hardware architecture also includes PCI Express cards support, which can be used to enhance the capabilities of the light DC with dedicated hardware accelerators (FPGAs, GPUs, etc.).

The SESAME architecture facilitates deployment of multiple CESCs owned by a Small cell network operator (SCNO) that can be leased to multiple other operators, i.e. the Virtual Small Cell Network Operators (VSCNOs), willing to provide coverage at a given venue. These CESCs will consist of small cell physical network functions (PNFs) and small cell virtual network functions (VNFs).

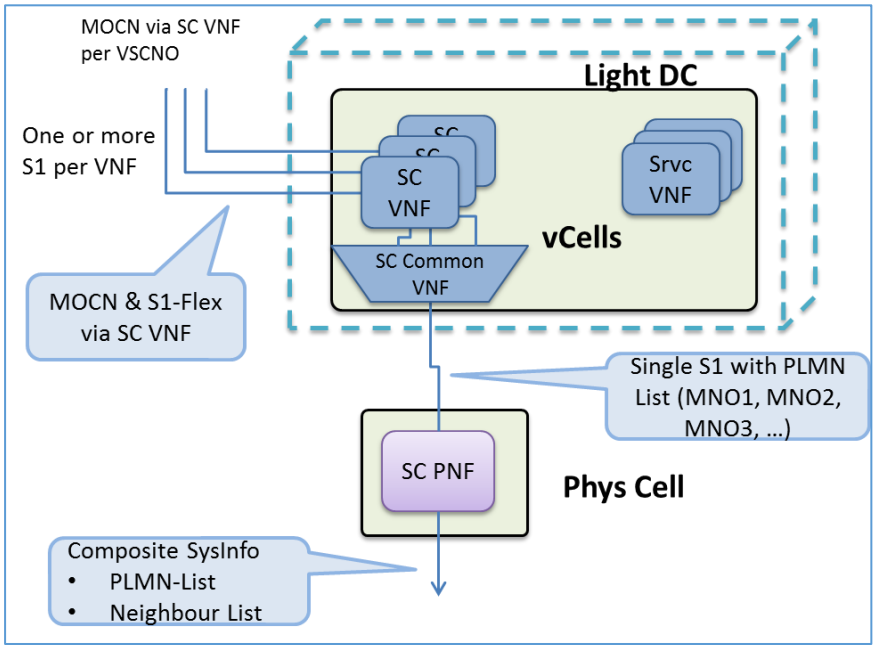


Figure 2: Small Cell VNFs

Each SC-PNF is associated with a single SC-Common-VNF plus an SC-VNF for each VSCNO using the SC-PNF. Additional Service VNFs provide further processing capabilities, in particular for User Plane processing.

The VNFs are hosted on a micro server platform residing in the Light Data centre. The SC-PNF provides a TR-069 [79] management interface to the PNF EMS to manage the cell specific TR-196 [80] parameters. The SC-PNF interfaces with an SC-Common-VNF that resides in the Light DC and terminates the S1 traffic belonging to multiple PLMNs over a single S1 link. The Light DC provides a virtualisation environment to deploy multiple VNFs. The SC-VNFs are located in the Light DC as a whole: they are not necessarily implemented in the micro server directly associat-

[2] The ARMv8 architecture introduces 64-bit support to the ARM architecture with a focus on power-efficient implementation while maintaining compatibility with existing 32-bit software. More related information can be found at: http://www.arm.com/products/processors/armv8-architecture.php

ed with the SC-PNF. The virtualisation mechanism for each of these VNFs is FFS[3], they may or may not be instantiated as formal VNFs[4].

The VNF Manager in CESCM manages the orchestration of VNFs based on the configuration set by the SCNO. The SC EMS in CESCM manages the tenant (VSCNO) specific configuration set by the VSCNO via northbound management interface[5]. The configuration associated with Edge Services hosted in Service VNF is managed via the (Service) EMS in CESCM. The SLA monitoring function in CESCM accumulates statistics from the PNF EMS and the SC EMS to present the KPIs to the VSCNO NMS over the north bound interface. The CESCM also allows the VSCNO to view the results of autonomous CSON and TR-069 configuration updates.

The SC-Common-VNF routes the traffic belonging to different VSCNOs (identified by their PLMNs) to the corresponding SC-VNFs and vice versa. It also performs SCNO admission control and congestion control as per the policies set by the SLAs agreed with VSCNO. The SC-VNF performs DSCP marking of packets per operator as per the configuration set by the SC EMS. It also performs real time congestion control based on the directives from SC-Common-VNF to regulate the Fronthaul (SC-PNF to SC-Common-VNF interface) and Backhaul usage (SC-VNF to VSCNO EPC interface) as per the SLA terms agreed with VSCNO.

This section primarily focuses on detailed design of CESC Proof of Concept (PoC) solution, although some functional blocks outside the CESC will also be discussed occasionally for CESC interface design.

2.1.2 Functional Split

The CESC proof of concept solution discussed in this document is based on a functional split of use cases based on the standard S1 interface [68] as shown below.

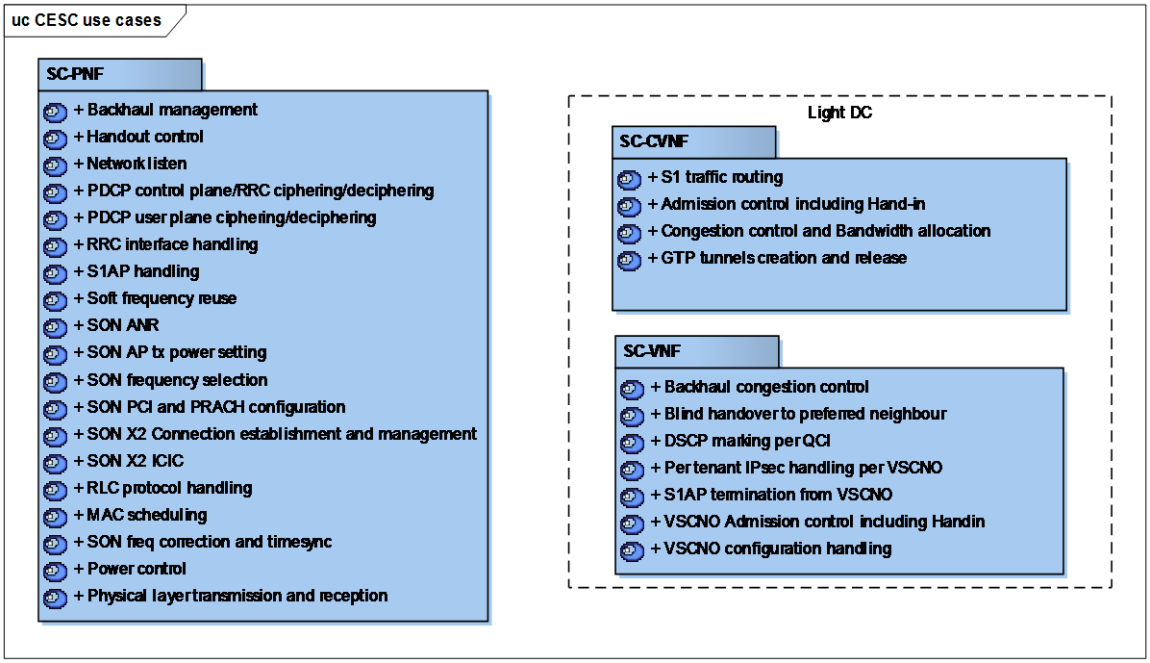


Figure 3: Split between SC-PNF and Light DC

[3] More Related information can be found, for example, at: https://en.wikipedia.org/wiki/Unix_File_System

[4] This actually depends on the functional behaviour and processing load being realised in the respective VNFs and is mainly a job for Virtual infrastructure manager (VIM). For Example, "containers" might be better approach than VMs to avoid duplication of OS resources across multiple VNFs.

[5] For more details see, *for example*: https://en.wikipedia.org/wiki/Northbound_interface

The use cases are split into SC-PNF, SC-CVNF and SC-VNF packages as shown above. This split offers basic Admission control, Bandwidth Management, Configuration Management and KPI monitoring aspects per VSCNO as virtual functions.

This functional split does not restrict the multi tenancy use case implementation in general; however, some advanced splits may offer advantages in dense deployments. D2.3 has already discussed the RAN virtualisation concepts and possible splits. This document explores the potential designs of various functional splits and compares them.

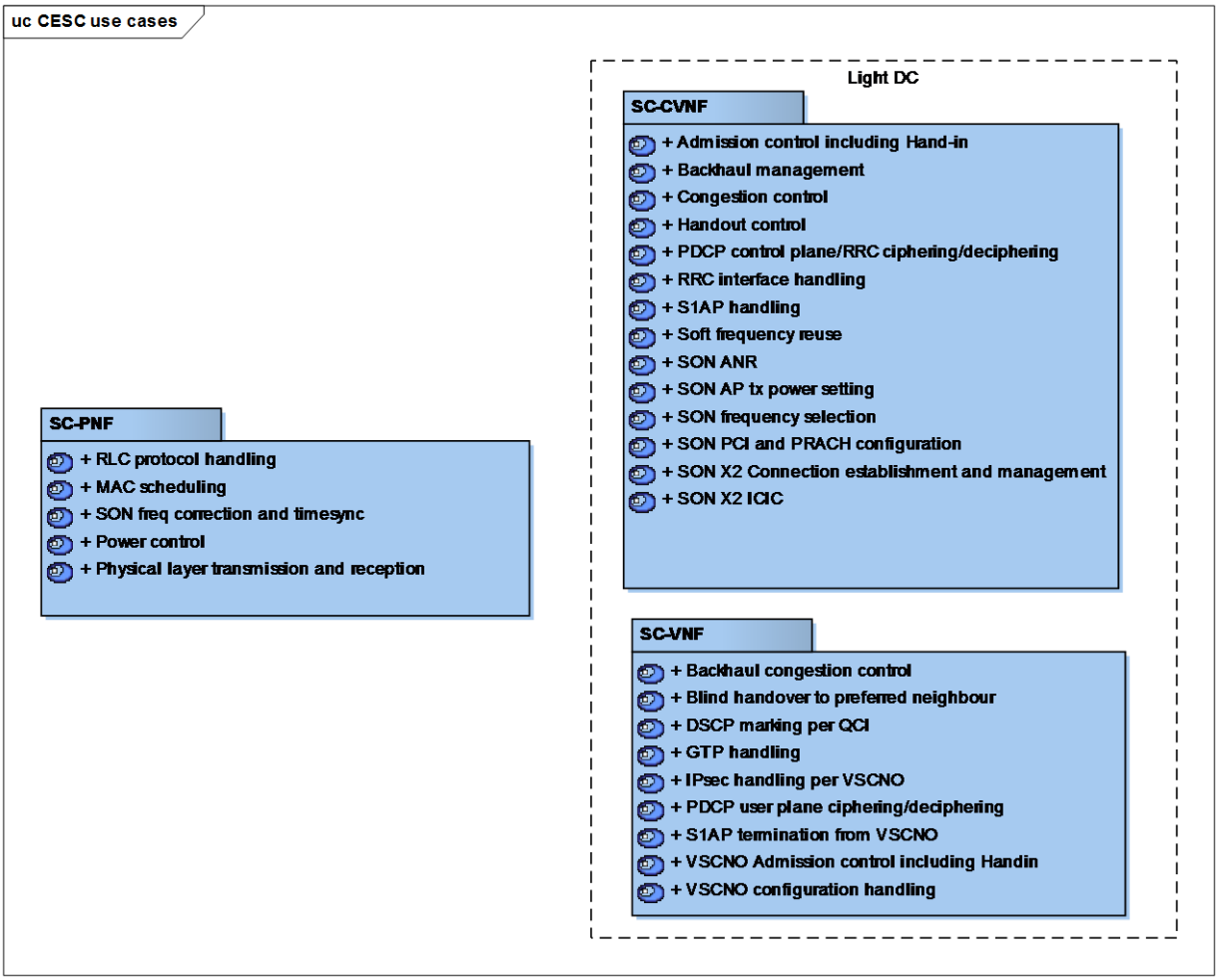


Figure 4: PDCP split between SC-PNF and Light DC

Following are the other possible functional splits discussed in this document for comparison, which are outside the scope of proof of concept solution.

- PDCP split (RRC-PDCP and PDCP-RLC) between SC-PNF and SC-C-VNF.
- FAPI split between SC-PNF and SC-C-VNF.

2.1.2.1.1 PDCP Split

The PDCP split can be realised as shown below:

This architecture is more suited for dense deployments with Fronthaul links that do not meet CPRI standard[6] (i.e. Fronthaul latency > 200 micro-seconds).

Detailed overview and latency requirements of PDCP splits is available in [86].

2.1.2.1.2 FAPI Split

The FAPI Split is as shown Figure 5 below:

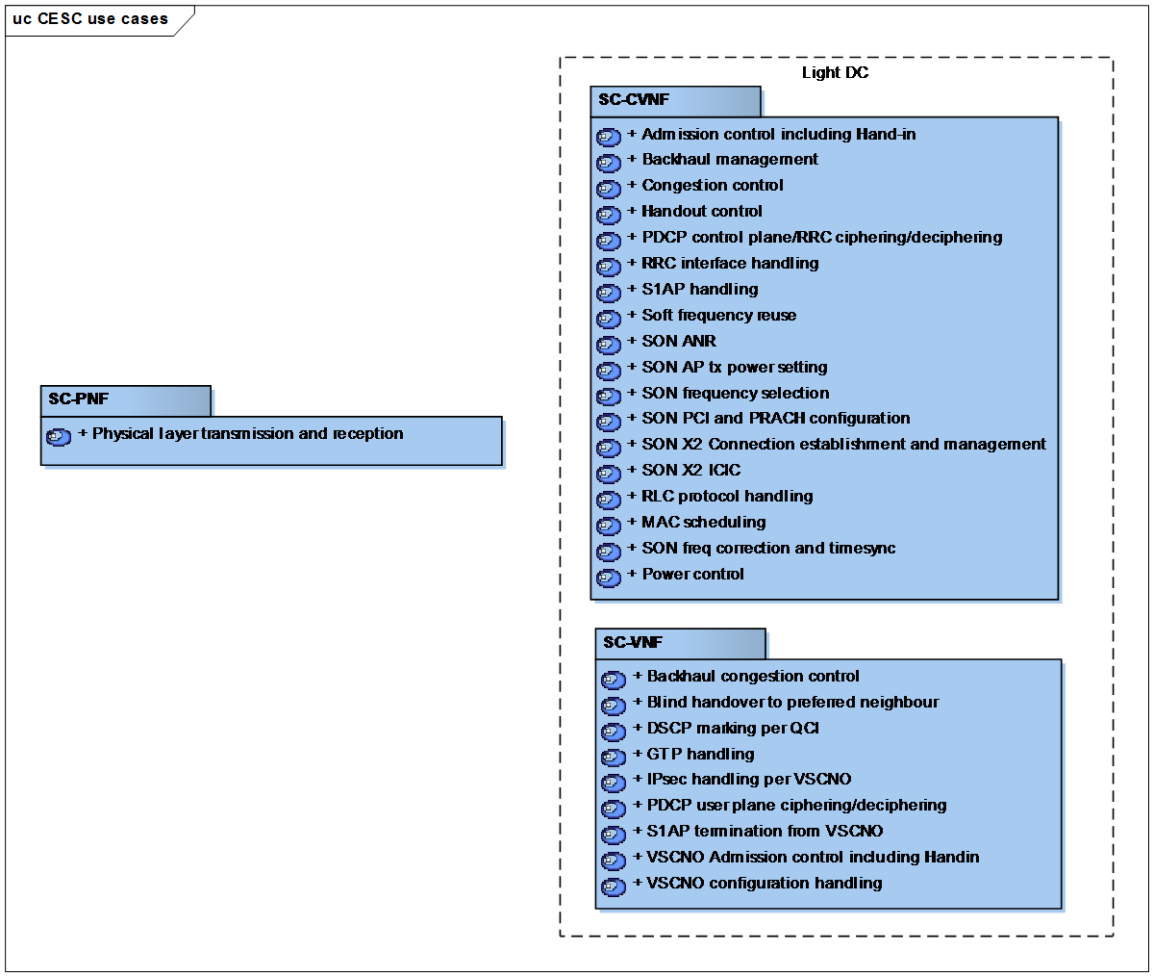


Figure 5: FAPI split between SC-PNF and Light DC

This architecture is more suited for dense deployments with Fronthaul links that meet CPRI standard (i.e. Fronthaul latency > 200 micro-seconds).

Detailed overview and latency requirements of FAPI split can be found in [86].

The following table presents a comparison between various functional splits:

[6] For more information see: http://www.cpri.info/

Differentiating criteria	S1 split	PDCP split	FAPI Split
MOCN and S1-Flex feasibility	Yes	Yes	Yes
Need for IP sec between SC-PNF and Light DC for user plane data	Yes	No	No
IP sec overhead on Fronthaul	Medium	Low	Low
IP sec encryption/decryption hops	2	1	1
Need for non 3GPP interface between SC-PNF and Light DC	No	Yes	Yes
Implementation complexity	Low	High	Medium
Fronthaul latency requirements	Low	Low	High
Access point resources (cost)	High	Medium	Low
Light DC resources (cost)	Low	Medium	Very High
Centralised RRM	No	Yes	Yes
VSCNO specific RRM	No	Yes	Yes
Effective Fronthaul utilisation based on CQI	No	Yes (via RRM<->MAC interface configuration)	Yes
Resource block/Spectrum split realisable	No	Yes (via RRM<->MAC interface configuration)	Yes
Centralised configuration of CESC resources at Light DC	No	Yes	Yes
More flexibility for centralised SON within clusters	No	Yes	Yes
Deployment ease and Self-X for large clusters	Complex	Very Easy	Easy
Control plane State machine complexity	Medium (due to distributed S1 and RRM functions)	Low	Low
Intra cluster handover Data forwarding limited to Light DC	No	Yes	Yes
Typical use cases	Small venue consisting of 5 to 10 Access points.	Venue consisting of more than 10 access points along with Better to have HeNBGW in Light DC.	Dense deployments in Stock markets, Defence, mass deployment of very low cost APs. Better to have HeNBGW in Light DC.

Table 1: Functional splits comparison

2.1.3 SC-PNF

The SC-PNF handles eNodeB [3GPP 36.300] functions that are common to all tenants, while the SC-VNF handles the tenant (i.e. the VSCNO) specific eNodeB functions. For example, in the SESAME PoC, the SC-PNF represents the end-point of S1 traffic coming to/from the tenant EPC through SC-VNF. In this instance, the PNF implements all the standard functionalities associated with an eNodeB or HeNB such as management and allocation of radio resources for the UEs under coverage and maintaining control and user plane connections with the core network of the VSCNO. This also implies that in the Proof of Concept (PoC) implementation, the SC-PNF is in essence agnostic to the presence of intermediaries between itself and the EPC or a collection of EPCs and works as if directly connected to the core network.

From the UE's perspective, the SC-PNF offers a standard Uu LTE interface[7] and therefore, no change is required on its part.

From the CESC side, the SC-PNF has a standard S1 traffic interface (for signaling and user plane traffic).

The SC-PNF is managed by the PNF-EMS via standard TR-069 interface [79] in line with the standard 3GPP practice. For additional management functions such as software updates or upload of performance management reports, an HTTP based file transfer is used. All the logical interfaces between the SC-PNF and CESC are carried over a single 1GbE physical interface and can optionally be secured using IPSec in tunnel mode[8]. The SC-PNF does not offer an explicit differentiation of the services between the PLMN-IDs.

Considering that the SC-PNF supports the standard 3GPP Multi-Operator Core Network (MOCN) functionality, it has to perform the additional functions required for this purpose, which include:

(a) Broadcasting PLMN-IDs: This function requires the SC-PNF to broadcast the PLMN-IDs of the operators (i.e. the VSCNOs) using the CESC, allowing UEs select their operator of choice from the broadcast list[9].
(b) Maintaining connectivity: This function has several associated tasks. It involves associating and maintaining the connectivity of the UEs with their respective PLMN, maintaining and broadcasting a neighbor list associated with multiple PLMN-IDs, and taking care of the mobility of the UEs by allowing handovers to a neighbor cell with the same PLMN-ID.

2.1.3.1 Detailed description of functions

The Small cell physical network function (SC-PNF) component performs the following functions.

- Physical layer procedures[10]
- Medium Access Control[11] (MAC)
- Radio Link Control[12] (RLC)
- Packet Data Convergence Protocol[13] (PDCP)
- Radio Resource Control[14]
- System Information Broadcast (SIB)
- Radio Resource Management[15] (RRM)
- S1AP protocol[16] handling using a single S1 link to SC-common VNF in the Light DC.

---

7 See, *for example*: https://en.wikipedia.org/wiki/EnodeB
8 IPSec operation will not be provided in the PoC implementation.
9 Note that as the CESC is an LTE cell and MOCN support is a mandatory LTE feature, all UEs using the CESC are, by definition, MOCN supporting.
10 For more detailed information see: http://www.3gpp.org/dynareport/36213.htm
11 For more related information see, *among others*: https://en.wikipedia.org/wiki/Media_access_control
12 More related information can be found at: http://www.3gpp.org/DynaReport/25322.htm
13 More related information can be found at: http://www.3gpp.org/dynareport/36323.htm
14 More related information can be found at: http://www.3gpp.org/DynaReport/25331.htm
15 For more related information see, *among others* https://en.wikipedia.org/wiki/Radio_resource_management

- Network listen
- Distributed SON (D-SON) functions[17]
- IP sec handling and
- Support for Multi operator core network (MOCN)
- TR-069 protocol [79] to configure the AP
- ETWS[18] and/or CMAS[19] scheduling

These functions accomplish following tasks as documented in the D2.3 deliverable:

- Managing the allocation and scheduling of radio resources towards the UEs.
- Maintaining signalling and user plane connections towards a single S1AP signalling termination point.

The MOCN capability in the proof of concept PNF is limited to:

- Broadcast of multiple PLMN IDs in SIB1[20] based on TR-196 [80] configuration from CESM.
- Associating an RRC connected UE context with the PLMN selected by the UE.
- Auto discover and broadcast the neighbour relations associated with all configured PLMNs.
- Facilitate broadcast of "cellReservedForOperatorUse" on a per PLMN basis based on configuration, so that a tenant (VSCNO) can inhibit its users from accessing the cell.
- Activate hand-out measurements on neighbour cells belonging to the selected PLMN ID from the UE.
- Generating and reporting per PLMN PM Counters uploaded to SLA monitoring function for KPI generation.

2.1.3.1.1 PLMN ID Broadcast

The EMS business logic described in *section 2.1.9.3.2* is responsible for collating the PLMNs of each SC-VNF and configuring the SC-PNF with the consolidated list. If an SC-VNF is not in an operational state (for example it has no S1AP connection to an MME) the corresponding PLMN ID broadcast is SIB1 is marked as unavailable by setting *cellReservedForOperatorUse* to "true". This ensures that the list of PLMNs broadcast by the SC-PNF do not change, even when they are temporarily unavailable. This is especially important for the first entry in the broadcast list which defines the home PLMN within which the SC-PNF's Global Cell Identity is defined. It is not possible to remove this PLMN from the broadcast list without effectively re-defining the identity of the cell.

2.1.3.1.2 Measurement Configuration

By its very nature, the neighbour cell information broadcast by the SC-PNF in System Information Blocks (SIBs) applies to all idle mode UEs, regardless of which PLMN they belong to. However, once a UE enters connected mode and has declared its *Selected PLMN* in the *RRC Set-up Complete* message, the SC-PNF performs PLMN-*specific* measurement configuration and instructs the UE to perform measurements only on the sub-set of neighbours that support its *Selected PLMN*.

2.1.3.1.3 Bandwidth management

The SC-PNF makes sure that the total uplink bandwidth demands of users does not exceed the UL front-haul capacity configured by EMS (*MaxUplinkBackhaulBitRate*).

---

16 A description of the S1AP can be found at: http://lteworld.org/specification/s1-application-protocol-s1ap
17 See: https://en.wikipedia.org/wiki/Self-organizing_network
18 See, *for example*: http://blog.3g4g.co.uk/search/label/ETWS
19 See, *for example*: http://www.dgs.ca.gov/pd/Programs/Leveraged/CMAS.aspx
20 For more information see: http://4g-lte-world.blogspot.gr/2012/10/system-information-block-1.html

2.1.3.2 Parameters, Alarms and PM Counters

The SC-PNF is managed by the PNF EMS by using the standard TR-069 protocol [79]. The data model is based on the TR-196 standard [80] with some SESAME-*specific* vendor extensions. HTTP based file transfer is used for software download and PM stats upload.

As part of CESC management and control, several SC-PNF configuration parameters, together with associated alarms and performance management aspects have to be addressed. Because the SC-PNF in the SESAME PoC implementation covers most of the functionality of a traditional eNB/HeNB, many of its attributes, alarms and performance management aspects are common with these legacy network entities. Femto cells and legacy Customer Premises Equipment (CPE) have been in operation for some time now together with some established management and control procedures. Because the CESC is also, fundamentally, a small cell with added benefits of cloud systems, its management and control aspects specific to SC-PNF are similar to the legacy network operator provided CPEs. This section describes some of these common and CESC specific attributes, alarms and performance management counters of SC-PNFs.

2.1.3.2.1 Parameters

Configuration parameters describe the general features set of the SC-PNF and may be functionality and control specific. The parameters may relate to the offered services, performed functions, and hardware capabilities of the specific entity i.e. the SC-PNF. The list below gives a brief description of the parameters associated with SC-PNF.

- Hardware parameters: Specifies all parameters related to the hardware capabilities of the SC-PNF. Most of these attributes are composite objects that may include other hardware-*specific* parameters such as maximum transmission power, supported systems/sub-systems (e.g. LTE, LTE-A[21]), maximum number of devices that can be attached, and etc.
- Service-*specific* parameters: The object representing the services provided by SC-PNF and the specific parameters associated with that capability.

As stated above, the SESAME PoC implementation of the SC-PNF implements the TR-196 data model version 2 for LTE an HeNB. The majority of the parameters in this data model are configured by the PNF EMS and are not described in this document. However, those parameters of special interest to SESAME are as follows:

TR-196 Parameter	**Description**
Device.Services.FAPService.{i}. CellConfig.LTE.EPC.PLMNList.{i}	This set of objects defines the list of PLMNs broadcast by the SC-PNF. There is one active instance for each supported PLMN. The first entry is of special significance as it defines the "home PLMN" of the cell. These objects are automatically populated by the EMS using the configuration of each SC-VNF hosted by the SC-PNF.
Device.Services.FAPService.{i}. CellConfig.LTE.EPC.PLMNList.{i}. PLMNID	Defines the identity of a PLMN supported by the SC-PNF.

[21] For more details, see: https://en.wikipedia.org/wiki/LTE_Advanced

TR-196 Parameter	Description
Device.Services.FAPService.{i}. CellConfig.LTE.EPC.PLMNList.{i}. CellReservedForOperatorUse	Set to "true" when the associated PLMN is not available to normal users.
Device.Services.FAPService.{i}. REM.UMTS.GSM.Cell.{i}.	These sets of objects are populated by the SC-PNF's SON ANR function. They capture the details of automatically discovered neighbour cells. Each such object requires a vendor extension to record up to five additional PLMNs supported by MOCN capable neighbours.
Device.Services.FAPService.{i}. REM.UMTS.WCDMA.Cell.{i}.	
Device.Services.FAPService.{i}. REM.LTE.Cell.{i}	
FAPService.{i}. CellConfig.LTE.RAN.NeighborListInUse. InterRATCell.GSM.{i}.	These sets of objects are populated by the SC-PNF's SON ANR function. They capture the details of neighbour cells that are currently being used; broadcast in System Information (SI) and used to configure measurement reports. Each such object requires a vendor extension to record up to five additional PLMNs supported by MOCN capable neighbours.
FAPService.{i}. CellConfig.LTE.RAN.NeighborListInUse. InterRATCell.UMTS.{i}.	
FAPService.{i}. CellConfig.LTE.RAN.NeighborListInUse. LTECell.{i}.	
Device.Services.FAPService.{i}. FAPControl.LTE.Gateway. S1SigLinkServerList	Configured by the EMS with the address of the SC-Common VNF.
Device.Ethernet.Interface.{i}. X_000295_ MaxUplinkBackhaulBitRate	Max Uplink Front-haul bit rate configured by the EMS.
Device.Ethernet.Interface.{i}. X_000295_ MaxDownlinkBackhaulBitRate	Max Downlink Front-haul bit rate configured by the EMS.

Table 2: SESAME Specific Parameters

2.1.3.2.2 Alarms

In general, alarms describe the specific conditions that trigger a particular EMS action such as resolution, reporting and prioritisation from the management and control framework of the system. Alarms in managed systems are usually represented as composite objects containing several parameters/attributes that describe the specific details of the alarm with some contextual information.

The SC-PNF also has to take care of several types of alarms including their representation (i.e. the data representation) and handling (the pre-defined set of operations). Being objects, the alarms are usually maintained in lists that represent the pending alarms as well as a history/log of handled alarms. All alarms have some common attributes such as alarm identifier, alarm type, associated event, probable cause of alarm, specific problem, notification type, perceived severity, and a brief description.

In CESC, the SC-PNF maintains the list of current alarms (Current Alarm List) that are not yet cleared. Newly raised alarms are added to this list and, if any change occurs during the course of action, the alarm parameters are updated. When an alarm is fully handled, it is cleared from the

current alarm list. A history of recent alarms raised by the SC-PNF is also maintained in an Alarm History List. These *CurrentAlarm* and *HistoryEvent* lists as part of the standard TR-196 data model [80] and are used by the EMS as part of its Fault Management view (see *section 2.1.9.1.3*).

2.1.3.2.3 PM Counters

PM counters are used to keep a track of the state of the CESC performance, its offered services and the quality of those services. Consequently, each network element has its associated PM counters that collectively represent the state of these aspects in that specific network element's scope.

In the scope of SC-PNF, it has to monitor the performance management counters and report them (e.g. as file upload to designated File Server). The file upload process has its own associated parameters such as upload interval, periodicity and authentication information. The SC-PNF collects periodic statistics for performance management purpose that are represented in Sample Sets with management attributes such as start time, finish times, intervals and etc. There is a large number of parameters that can be monitored/measured to generate performance related metrics. The list below gives a broad classification.

- Radio network related measurements,
- Hardware platform related measurements,
- Transport network related measurements.

The SC-PNF supports the standard 3GPP PM counters defined in Appendix A. In addition, as part of SESAME, it also supports a number of extended PM counter variants, also described in Appendix A. These report separate per-PLMN counts that allow a VSCNO to receive tailored PM reports relating to their particular network slice.

The SC-PNF supports the parameters defined in 3GPP 32.592 [66] that control the destination and periodicity of PM report file upload.

2.1.4 SC-Common-VNF (SC-C-VNF)

As already described, the multi-tenancy features in SESAME are realised at the CESC level through operator/tenant specific SC-VNFs. Because the tenant specific SC-VNFs are hosted on top of the same CESC hardware resources, certain coordination and control functions can be aggregated on behalf of all tenants. The SC-C-VNF is that aggregation point in the SESAME architecture. The SC-C-VNF acts as a helper function to support coordination of the tenant specific SC-VNFs.

The upstream traffic from the SC-PNF towards the CESC is an aggregated S1 flow representing the traffic of all the UEs of all the CESC tenants. The SC-C-VNF de-multiplexes this joint S1 flow into tenant specific S1 flows and directs them (based on the PLMN-ID) to their respective tenant specific SC-VNFs. The reverse function is carried out in the downstream direction from the tenants specific SC-VNFs to the SC-PNF. The S1 streams from the CESC hosted SC-VNFs are multiplexed together into a single Stream Control Transmission Protocol (SCTP) connection with the SC-PNF. This multiplexing and de-multiplexing for the S1 traffic between the SC-PNF and SC-VNFs has to be done for both control signalling and user plane traffic.

Being the receiver of the SC-PNF upstream and downstream traffic, the SC-C-VNF can maintain a view of the resources that are visible to all the SC-VNFs. It can monitor the RAB assignments[22],

[22] As of RAB assignments see, for example: http://www.3gpp.org/ftp/tsg_ran/wg3_iu/TSGR3_06/Docs/Pdfs/r3-99a09.pdf

modifications and releases for all tenants and for all purposes including handovers. It is, thus, able to apply control to these flows in order to apply the network shares defined by the SLAs of each SC-VNF.

2.1.4.1 Detailed description of functions

The SC-Common-VNF within the CESC performs the following functions.

- Multiplexing and de-multiplexing of S1AP signalling traffic towards SC-PNF and SC-VNF respectively,
- Admission control including Hand-in based on the resources of the SC-PNF as a whole,
- Congestion control during user admission,
- Bandwidth management.

The SC-Common-VNF is provisioned with parameters that reflect the SC-PNF capabilities and SLA agreements between SCNO and VSCNO for multi tenancy.

2.1.4.1.1 S1AP Traffic Routing

The SC-Common-VNF performs the following functions related to S1AP traffic routing:

- Accepts a single S1 connection request from the PNF,
- Creates up to six S1 connections to each SC-VNF,
- Routes S1 messages from the PNF to the appropriate SC-VNF,
- Performs a small amount of identity translation.

Note that the SC-Common VNF deals entirely with control plane messages and, apart from set-up and termination control plane messages, it is **not** involved with user plane traffic.

The SC-Common-VNF acts on S1AP messages as described below to facilitate S1AP routing between SC-PNF and SC-VNF.

S1AP Message	SC-Common-VNF handling
S1 Setup Request from SC-PNF	The SC-Common VNF extracts the Broadcast PLMNs IE and constructs up to six equivalent messages, one per entry in the Broadcast PLMNs list, each containing a single PLMN. It then establishes an SCTP connection to each of the appropriate SC-VNFs and forwards a customised S1 SETUP REQUEST to them. If there is a mismatch between the Broadcast PLMNs list and the list of VNFs with which the SC-Common VNF has been configured, it emits a Configuration Error alarm and continues, skipping the affected entry.
S1 SETUP RESPONSE from SC-VNF	On receipt of the S1 SETUP RESPONSE from each SC-VNF, the SC-Common VNF constructs merges the list of Served GUMMEIs into a single consolidated list. When all responses have been received or a procedure timeout has expired the SC-Common VNF sends a combined S1 SETUP RESPONSE to the SC-PNF. Failed S1 connections are retried periodically.
S1 INITIAL UE Message from SC-PNF	The SC-Common VNF extracts the Selected PLMN of the UE from the TAI IE of the message and uses this

S1AP Message	SC-Common-VNF handling
	to identify the SC-VNF to which the message should be forwarded. It extracts the value of the eNB UE S1AP ID IE and records this in a table for the associated SC-VNF. The message is then forwarded to SC-VNF. If the Selected PLMN cannot be matched against a connected SC-VNF a Configuration Error alarm is emitted and the message is discarded.
SC-VNF S1 Connection failure	If the underlying SCTP connection to an SC-VNF fails, the SC-Common VNF re-tries it periodically. On failure, it starts a hysteresis timer and, if this timer expires before the connection is re-established, the SC-Common VNF sends an MME CONFIGURATION UPDATE message to the SC-PNF with a revised list of Served GUMMEIs omitting those provided by the SC-VNF that cannot be reached.
SC-VNF S1 Connection Late Establishment or Re-establishment	If an S1 Connection to an SC-VNF is established after the SC-Common VNF has established its connection to the SC-PNF or it is re-established after the hysteresis timer has expired, then the SC-Common VNF merges the list of Served GUMMEIs from the S1 SETUP RESPONSE into its consolidated list and sends an appropriate MME CONFIGURATION UPDATE message to the SC-PNF.
S1AP MME CONFIGURATION UPDATE from SC-VNF	On receipt of an MME CONFIGURATION UPDATE message from an SC-VNF, the SC-Common VNF processes the list of Served GUMMEIs and merges them into its consolidated list. If the update results in a change to the consolidated list a corresponding MME CONFIGURATION UPDATE message is constructed and sent to the SC-PNF. The Relative MME Capacity IE is always omitted from such messages as the SC-PNF only has a single S1 connection. If more than one SC-VNF uses the same MME resulting in duplicate messages, the above processing ensures that only a single MME CONFIGURATION UPDATE is sent to the SC-PNF.
S1AP Initial Context Setup request from SC-VNF	The SC-Common VNF extracts the MME UE S1AP ID provided by the SC-VNF and allocates a new unique internal value. The range of this value (for example the most significant 4 bits) is chosen such that the associated SC-VNF can be easily identified. This new value is substituted in the message forwarded to the SC-PNF.
S1AP UE Context Release Request from SC-PNF	The SC-Common VNF extracts the MME UE S1AP ID provided by the SC-PNF and uses it to look up the associated table entry. It substitutes the original MME UE S1AP ID provided by the SC-VNF in the message forwarded to the SC-VNF. It then releases the internal MME UE S1AP ID value for re-use.
S1AP UE Context Release Complete	Handled as per UE CONTEXT RELEASE REQUEST

S1AP Message	SC-Common-VNF handling
from SC-PNF	above.
S1AP HANDOVER REQUEST	Handled as per INITIAL CONTEXT SETUP REQUEST above.
All Other S1AP Messages forwarded from SC-VNF to SC-PNF	For any S1 message that contains an MME UE S1AP ID IE, the SC-VNF looks up the value in a table associated with the SC-VNF and substitutes the stored internal value in the message relayed to the SC-PNF. All other IEs are forwarded as received.
All Other S1AP Messages forwarded from SC-PNF to SC-VNF	For any S1 message that contains an MME UE S1AP ID IE, the SC-VNF first identifies the associated SC-VNF instance (for example by inspection of the most significant 4 bits of the value) and then looks up the value in the table associated with the VNF. It substitutes the stored original value stored in the message relayed to the SC-VNF. All other IEs are forwarded as received.

Table 3: SC-Common VNF S1AP handling

2.1.4.1.2 Admission control

The SC-Common-VNF performs user Admission control whenever an S1AP *INITIAL UE MESSAGE* [68] is received from the SC-PNF. It also performs Admission control on bearer admission when it receives an S1AP *ERAB SETUP REQUEST* [68] message from the MME.

The SC-Common-VNF performs user admission control based on the SLA agreements between SCNO and VSCNOs. It also considers the overall SC-PNF capacity in terms of reserved number of UE contexts and reserved bandwidth before admitting a new user into the system.

The SC-Common-VNF detects congestion on the following events:

- SC-PNF reaches maximum capacity.
- When a bearer admission request comes with a guaranteed bit rate (GBR) greater than the available GBR bandwidth allowance for the SC-PNF.

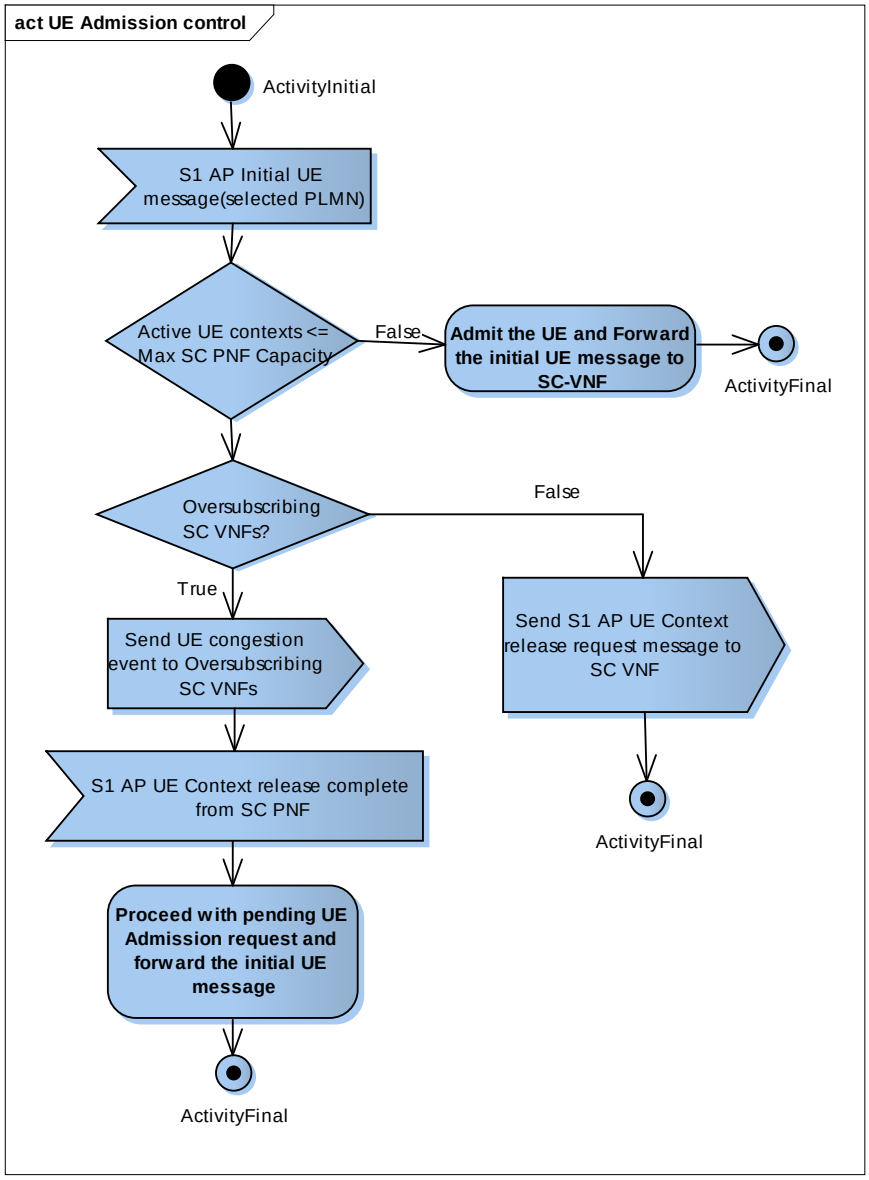


Figure 6: User Admission Congestion Handling

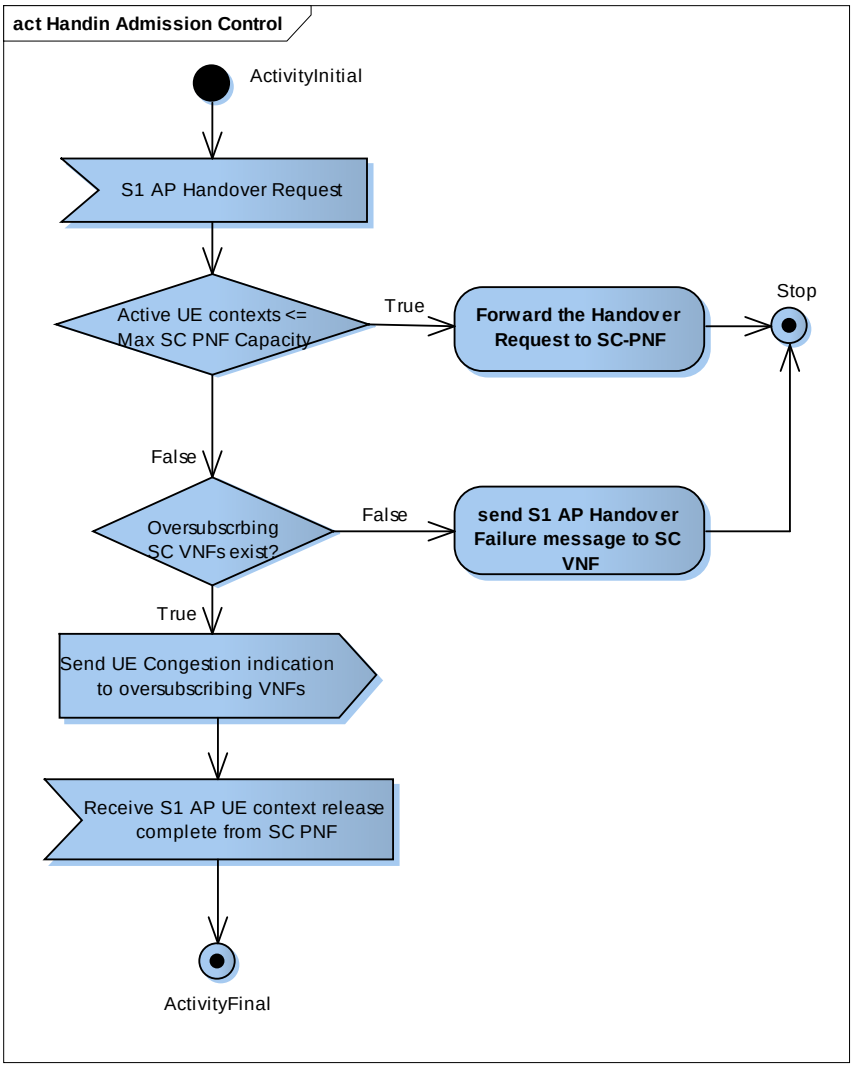


Figure 7: Hand-in Admission control

2.1.4.1.3 Bandwidth management

In SESAME, one of the key drivers for RAN sharing is to allow delivery of different bandwidth options as part of SLAs to different tenant VSCNOs. The SC-C-VNF makes sure, at all times, that the downlink bandwidth demands of tenant VSCNOs are met with the reserved bandwidth assigned to the VSCNOs as part of the SLA. The SC-C-VNF monitors the downstream bandwidth utilisation of tenant SC-VNFs and co-ordinates with the SC-VNFs to enforce policies to maximize the usage of the available front-haul link capacity towards the SC-PNF.

For PoC implementation, the SC-C-VNF policies for managing the available front-haul link capacity towards SC-PNF are listed below:

1) SC-C-VNF reserves a configurable percentage of the font-haul capacity for carrying management and control plane signalling traffic. After taking out the bandwidth reserved for management and signalling traffic SC-C-VNF makes the remaining bandwidth available for carrying user plane traffic.
2) SC-C-VNF marks an SC-VNF as available to handle traffic on receiving a "Status Indication" message with status "UP" (see 2.2.2) from the SC-VNF and configures the SC-VNF to report periodic bandwidth utilisation reports at a configurable granularity by sending a "Config Request" message to the SC-VNF.

3) On admission of the first user associated with the SC-VNF, the SC-C-VNF allocates a configurable minimum percentage of front-haul bandwidth to the SC-VNF. After reserving the minimum bandwidth allowance, the SC-C-VNF allocates the remaining front-haul bandwidth among other active[23] SC-VNFs based on the configurable oversubscription weights and the current bandwidth utilisation of each SC-VNF.
4) On release of the last user associated with the SC-VNF, the SC-C-VNF distributes the available front-haul bandwidth amongst the other active SC-VNFs based on the configurable oversubscription weights and the current bandwidth utilisation of each SC-VNF.
5) The SC-C-VNF receives bandwidth utilisation reports from SC-VNFs periodically after a configurable granularity period and distributes the available front-haul bandwidth among active SC-VNFs based on the configurable oversubscription weights and the current bandwidth utilisation.
6) The SC-C-VNF sets the operational status of the front-haul link towards SC-PNF as congested whenever the bandwidth utilisation of the front-haul link reaches the configurable congestion onset threshold limit of available front-haul link capacity.
7) On detecting congestion on front-haul link towards SC-PNF, the SC-C-VNF prepares a list of oversubscribing SC-VNFs from the bandwidth utilisation reports received from SC-VNFs. The SC-C-VNF then configures all the oversubscribed SC-VNFs with the reserved bandwidth limit to relieve the congestion.
8) On congestion abatement, the SC-C-VNF sets the operational status of the front-haul link towards SC-PNF to active whenever the bandwidth utilisation of the front-haul link towards SC-PNF falls below configurable congestion abatement threshold limit of available front-haul bandwidth.

The flow chart in Figure gives the details of actions performed by SC-C-VNF on receiving the bandwidth utilisation reports from all active SC-VNFs.

[23] SC-VNF is active if at-least one user associated with the SC-VNF is admitted. SC-C-VNF allocates configurable minimum percentage of the front-haul bandwidth to active SC-VNF even when current bandwidth utilisation of SC-VNF is 0 so that sudden increase in the bandwidth demand is met.

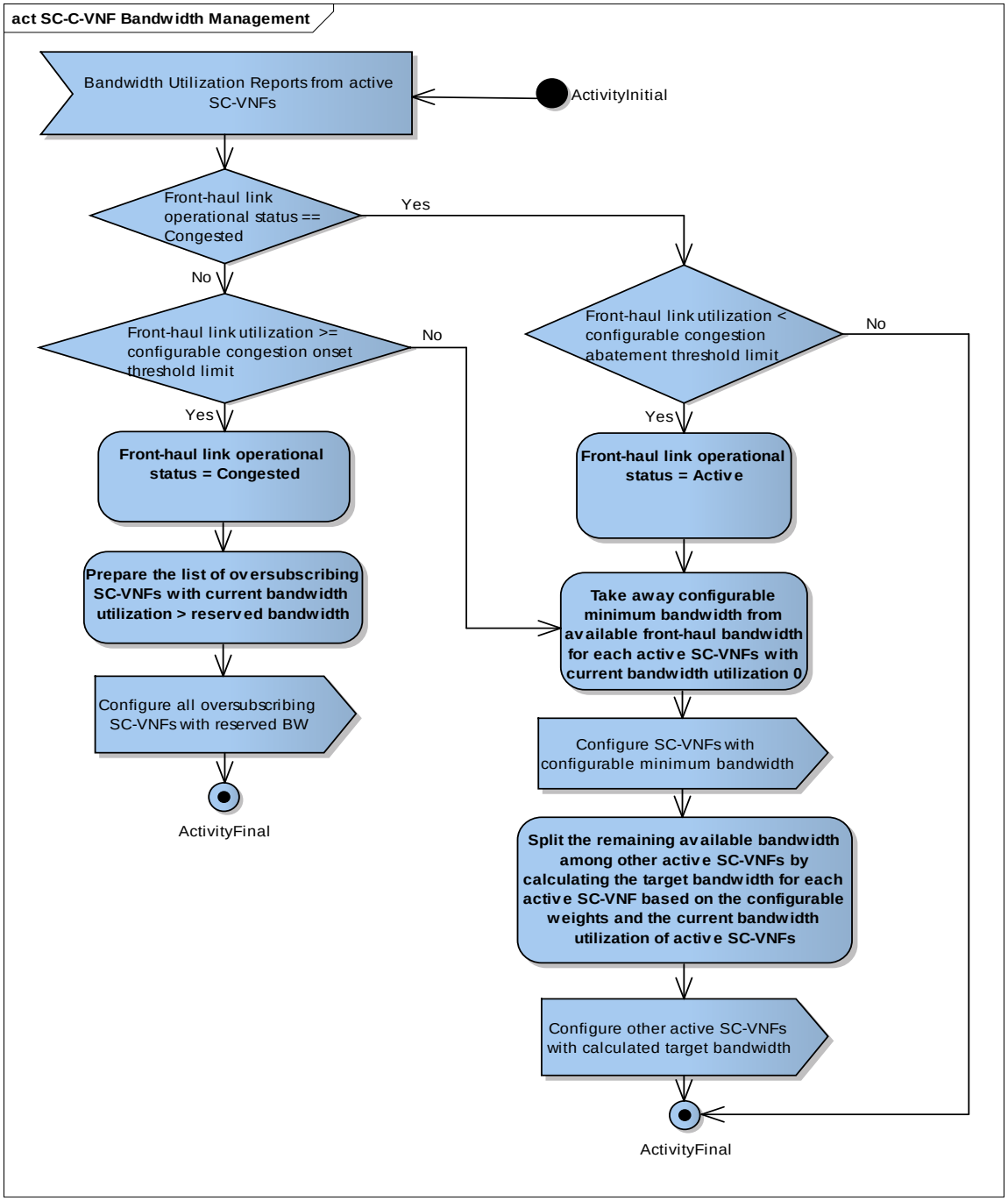


Figure 8: SC-C-VNF Bandwidth utilisation report handling

Equation (1) gives the calculation for distributing available front-haul bandwidth among active SC-VNFs.

$$TargetBWi = \frac{(Ui * Wi)}{\sum_{j=1}^{n}(Uj * Wj)} * (Available\_BW - m * MinBW) \quad (1)$$

$TargetBWi$– Is the target bandwidth calculated by SC-C-VNF for ith active SC-VNF with current bandwidth utilisation > 0.

Ui – Is the bandwidth utilisation factor of ith active SC-VNF.

$Ui = 0$ if current bandwidth utilisation is 0
$Ui = 1$ if current bandwidth utilisation > 0

Wi[24] – Is the configurable oversubscription weight of i[th] active SC-VNF where $0 < Wi \leq 1$.

$Available\_BW$ –Indicates the front-haul bandwidth available for distribution among active SC-VNFs after taking out allowance reserved for carrying management and signalling traffic.

n –Total number of active SC-VNFs.

m –Total number of active SC-VNFs with current bandwidth utilisation 0.

$MinBW$ –Configurable minimum bandwidth allocated to an active SC-VNF when the current bandwidth utilisation of SC-VNF is 0.

Figure 9 gives a view of front-haul bandwidth split for carrying control plane, management and user plane traffic.

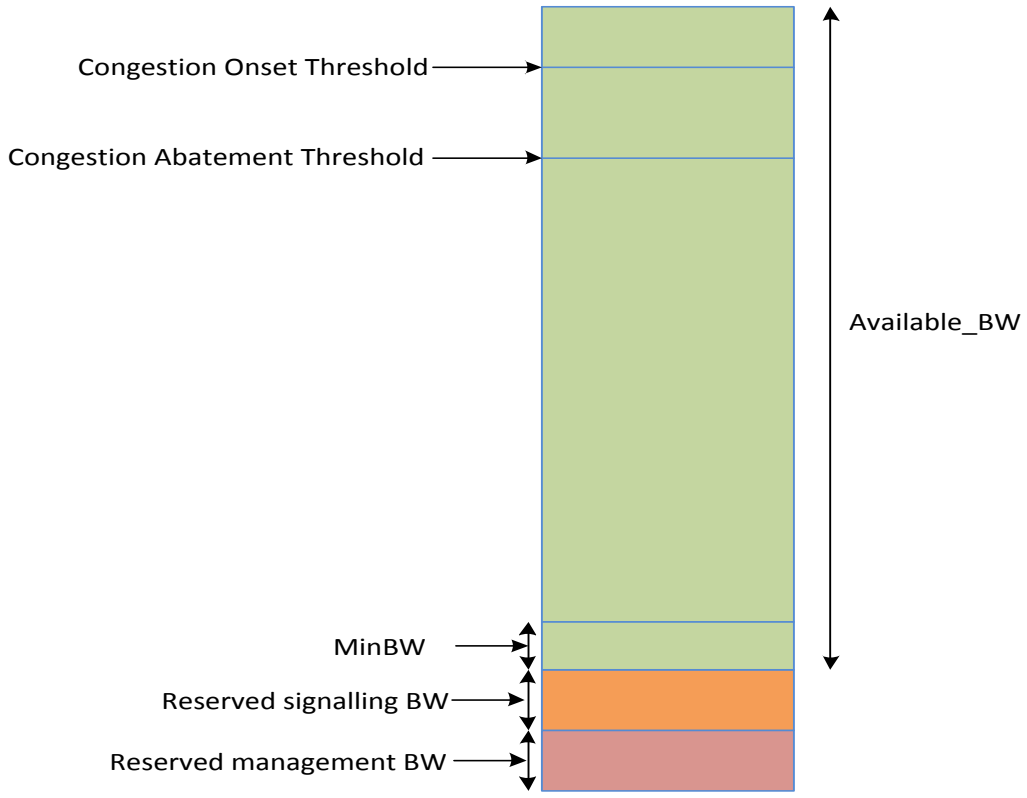


Figure 9: Front-haul Bandwidth Split

[24] Oversubscription Weight configured per SC-VNF associated with the VSCNO should be >= (Reserved Bandwidth for SC-VNF)/(Available_BW).

2.1.4.2 Parameters, Alarms and PM Counters

2.1.4.2.1 Parameters

Parameter	Description
MaxFronthaulBandwidth	Maximum fronthaul bandwidth including air interface limitations
MaxPNFCapacity	Maximum number of users supported by PNF
ReservedMgmtBandwidth	Reserved bandwidth available for carrying TR-069 management traffic, expressed as the percentage of the MaxFronthaulBandwidth
ReservedSignallingBandwidth	Reserved bandwidth available for carrying S1AP and X2 signalling, expressed as the percentage of the MaxFronthaulBandwidth
BwReportingGranularity	Periodicity at which Bandwidth reports are received by SC-C-VNF from SC-VNF
FronthaulCongestionOnsetThreshold	Congestion onset threshold specified in terms of percentage of the MaxFronthaulBandwidth
FronthaulCongestionAbatementThreshold	Congestion abatement threshold specified in terms of percentage of the MaxFronthaulBandwidth

Table 4: SC-C-VNF Parameters

2.1.4.2.2 Alarms

In the PoC implementation, the SC-Common VNF is able to report the following alarms:

Field	Value
3GPP Probable Cause	Link Failure
Specific Problem	S1 Link Down
Alarm Type	COMMUNICATIONS_ALARM
Permitted Severities	CRITICAL: Emitted when the link to the SC-PNF is down and the SC-C-VNF is no longer able to provide service CLEARED: Emitted when the link is restored
Additional Text	Describes any associated error code

Table 5: SC-C-VNF Link Failure Alarm

If IP Sec between the SC-C-VNF and SC-PNF is implemented, the following alarms may also be reported:

Field	Value
3GPP Probable Cause	Communications Protocol Error
Specific Problem	IPsec Client Authentication Failure
Alarm Type	COMMUNICATIONS_ALARM
Permitted Severities	CRITICAL: CLEARED:
Additional Text	

Table 6: SC-C-VNF Auth Failure Alarm

Field	Value
3GPP Probable Cause	Communications Protocol Error
Specific Problem	IPsec Dead Peer Detected
Alarm Type	COMMUNICATIONS_ALARM
Permitted Severities	CRITICAL: CLEARED:
Additional Text	

Table 7: SC-C-VNF Dead Peer Alarm

2.1.4.2.3 PM Counters

In the PoC implementation, the SC-C-VNF does not generate and PM reports.

2.1.5 SC-VNF

Whilst the SC-PNF serves as a network attachment point for the UEs of different CESC tenants, the network functions beyond the SC-PNF are differentiated through dedicated tenant-*specific* SC-VNFs. In other words, each SC-PNF is associated with an SC-VNF per tenant/VSCNO. However, a single SC-Common-VNF (see 2.1.4) provides the coordination service for all the SC-VNFs. As the SC-Common-VNF and SC-VNF lie on the path between the SC-PNF and the EPC of a particular tenant, they maintain a logical end-to-end S1AP interface.

From the overall functionality perspective, the SC-Common-VNF and SC-VNFs appear as an MME to the SC-PNF and as a standard operator-specific small cell (eNB/HeNB) to the operator's EPC.

For the PoC implementation, the chosen functional split between the SC-Common-VNF, SC-VNF and SC-PNF can provide a diverse set of functionalities. For example, an SC-VNF may provide support for multiple S1 connections towards an operator's EPC enabling S1-Flex[25]. Additionally, the SC-VNF provides a full implementation of the Non-Access-Stratum Node Selection Function (NNSF) and is responsible for MME selection for serving all the UEs of the tenant. As SESAME makes a distinction between SC-VNFs and Service VNFs, the set of functionalities associated with the SC-VNF covers all the basic features required in connectivity, control and management of

[25] S1-Flex functionality is not provided in the PoC implementation.

the CESC, which includes functions such as signaling, resource coordination among tenants, and user plane traffic control.

2.1.5.1 Detailed description of functions

The SC-VNF, within the CESC proof of concept solution, implements the following functions per VSCNO:

- Traffic shaping,
- Tenant specific Admission control based on limits applied to the specific tenant,
- GTP TEID Management within the CESC,
- Congestion control via blind handover,
- DSCP marking per QCI,
- S1AP routing to and from the Core Network.

2.1.5.1.1 Traffic shaping

The SC-C-VNF allocates the share of DL front-haul bandwidth to each active SC-VNF based on the bandwidth assigned to each tenant VSCNO as per the agreed SLA and the current bandwidth utilisation of each active SC-VNF. The SC-VNF regulates the flow of DL user plane packets based on the bandwidth configured by the SC-C-VNF. It does this by discarding or delaying the inflow of low priority traffic. The SC-EMS configures the SC-VNF with the VSCNO specific priority level for each traffic type identified by QCI.

The SC-VNF also regulates the flow of UL user plane packets on the backhaul link towards SGW of the respective VSCNO. EMS configures a maximum UL backhaul bandwidth limit per VSCNO. SC-VNF regulates the UL user plane packets by discarding or delaying the outflow of low priority traffic based on the configured UL backhaul bandwidth limit.

2.1.5.1.2 Admission control

The SC-VNF handles VSCNO tenant specific admission control, based on SLAs agreed with SCNO. For example, the SLA agreements between SCNO and VSCNO may allow oversubscription of resources if available beyond the reserved capacity purchased by the VSCNO. In the event of such oversubscription, if an overload event is detected by SC-C-VNF, it instructs the SC-VNF to perform congestion relief operation. The SC-VNF also performs VSCNO specific admission control during S1-AP hand-in operation based on configured PLMN and Handover restriction list. It also performs GBR bearer admission control based on reserved bandwidth procured by the tenant (VSCNO).

The following diagram shows an example embodiment of user Admission control algorithm within the SC-VNF.

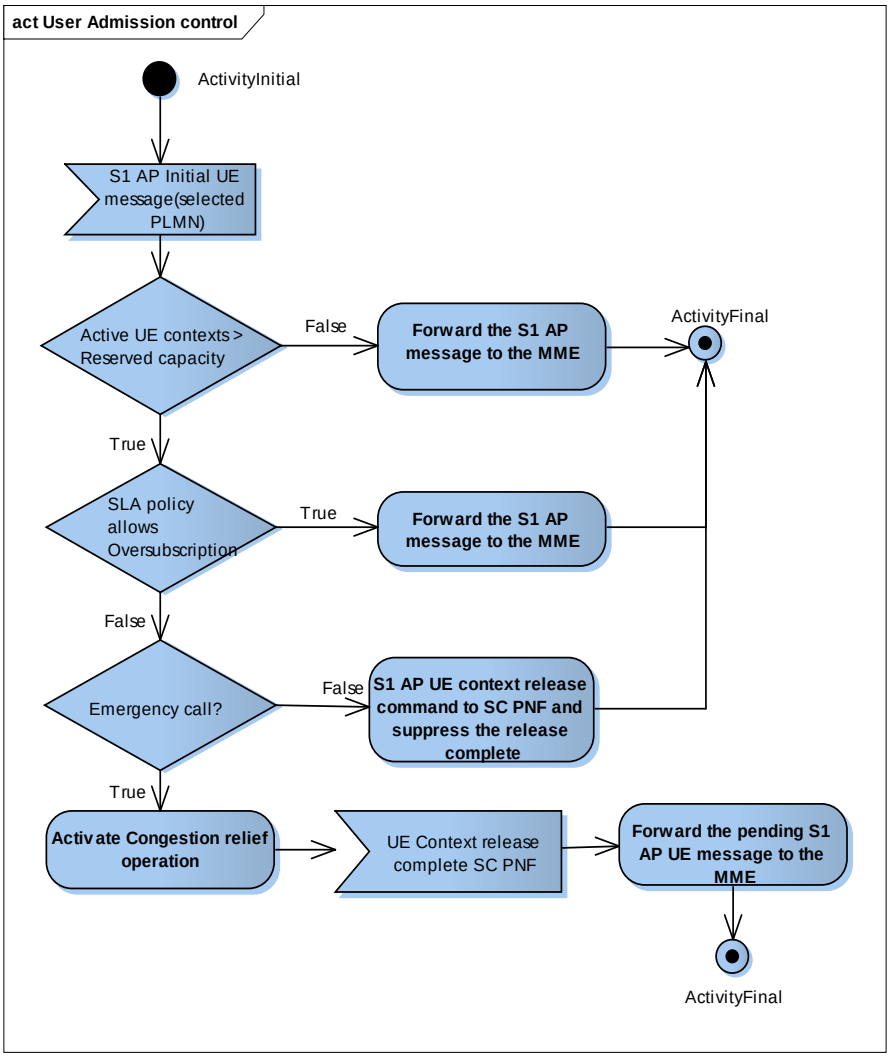


Figure 10: SC-VNF User Admission control

The SC-VNF performs UE admission control on S1 AP “Initial UE message” coming from SC-C-VNF or “Handover request” message coming from the MME. The Admission control policy, Reserved Bandwidth and Reserved user count parameters used in SC-VNF admission control algorithms are configured on the SC-VNF by the SC-EMS based on SLAs agreed between VSCNO and SCNO.

The following diagram shows an example embodiment of dedicated GBR Bearer Admission control at SC-VNF:

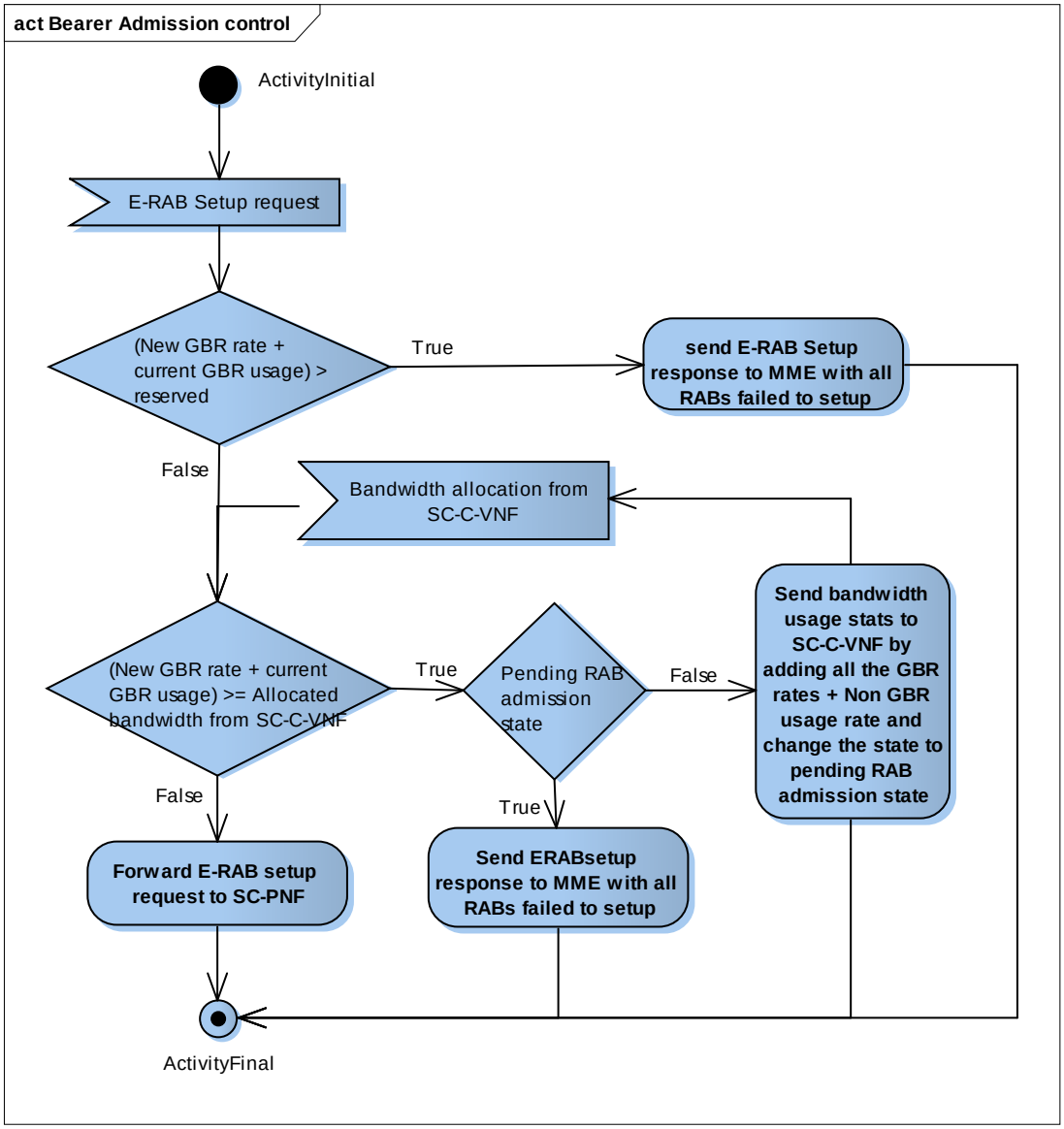


Figure 11 - SC-VNF Bearer Admission Control

The SC-VNF starts bearer admission control when it receives the S1 AP E-RAB setup request message from the MME. In the event of congestion caused by SC-C-VNF bandwidth allocation, the SC-VNF requests the SC-C-VNF for additional bandwidth. If the received bandwidth allocation still leads to congestion in Pending RAB admission state, the SC-VNF sends E-RAB setup response to MME with the list of RABs that failed to be admitted. If the new bandwidth allocation is sufficient to admit a new bearer, the SC-VNF forwards the E-RAB setup request to SC-PNF via SC-C-VNF.

2.1.5.1.3 GTP TEID Management within the CESC

In the traditional E-UTRAN architecture, S1 user plane GTP traffic flows are carried through GTP tunnels established between S-GW and eNodeB. In the context of SESAME, assuming the S1 functional split for a PoC demonstrator, the S1 GTP traffic flow is carried through multiple S1 GTP tunnel hops. In the downlink, one hop of the S1 GTP tunnel is established between the S-GW and the SC-VNF associated with the VSCNO. An optional second hop is via the Service Chain associated with the SC-VNF and the final hop of the S1 GTP tunnel is established between the SC-VNF and the SC-PNF. The reverse flow occurs in the uplink. The SC-VNF looks like an S-GW to the SC-PNF while handling UL GTP traffic whereas for DL GTP traffic originating from S-GW associated with VSCNO, the SC-VNF looks like and eNodeB.

Multiple GTP user traffic flows terminating at multiple GTP tunnel endpoints associated with a single transport layer address are identified by GTP TEID. Each per tenant VSCNO is associated with a unique IP address which imposes a requirement of allocating a unique GTP TEID to each traffic flow associated with the SC-VNF so that the traffic flow can be identified in the SC-VNF. The requirement of allocating unique GTP TEIDs is also imposed on the SC-PNF as the SC-PNF is also associated with single IP address. During default or dedicated bearer setup, the MME associated with VSCNO allocates a GTP tunnel endpoint. The SC-PNF and SC-VNF associated with the VSCNO each allocate a unique GTP tunnel endpoint for each bearer flow during default or dedicated bearer setup. SC-VNF binds the GTP endpoints allocated by VSCNO and SC-PNF together with its own GTP tunnel endpoint.

2.1.5.1.3.1 GTP TEID allocation within CESC

This section describes allocation and binding of GTP tunnel endpoints associated with the traffic flow inside the CECS during S1-AP default or dedicated EPS bearer setup control procedures. SC-VNF receives S1-AP “Initial Context Setup Request” or “E-RAB Setup Request” message from MME and allocates a new GTP TEID for each bearer and binds the “GTP TEID” and “transport layer address” received from MME with the allocated GTP TEID. SC-VNF then sends the S1-AP “Initial Context Setup Request” or “E-RAB Setup Request” message towards SC-C-VNF by setting its own IP address and newly allocated GTP TEID in the “transport layer address” and “GTP TEID” IEs respectively. SC-C-VNF forwards the message to the SC-PNF. SC-PNF on receiving S1-AP “Initial Context Setup Request” or “E-RAB Setup Request” message allocates a new GTP TEID for each bearer and sets its own IP address and newly allocated GTP TEID in the “transport layer address” and “GTP TEID” IEs respectively in S1-AP “Initial Context Setup Response” or “E-RAB Setup Response" message sent back to SC-C-VNF. SC-C-VNF forwards the message to SC-VNF. SC-VNF binds the “transport layer address” and “GTP TEID” received with the GTP TEID allocated by SC-VNF during the processing of S1-AP “Initial Context Setup Request” or “E-RAB Setup Request” message. SC-VNF sets its own IP address and GTP TEID in the “transport layer address” and “GTP TEID” IEs respectively in the “Initial Context Setup Response” or “E-RAB Setup Response" message sent back to MME.

2.1.5.1.3.2 GTP traffic routing within CESC

This section explains routing of UL and DL GTP-U packets associated with a traffic flow inside CESC. SC-PNF while sending the UL GTP-U packet sets the GTP TEID of the SC-VNF associated with the traffic flow in the GTP-U protocol header and forwards the GTP-U packet to the destination IP address of the SC-VNF. SC-VNF decodes the GTP-U packet and identifies the traffic flow from the received GTP TEID. SC-VNF after service chaining gets the GTP-U endpoint of the S-GW associated with the traffic flow and sets the GTP TEID in the GTP header before forwarding the GTP-U packet to the IP address of the S-GW associated with VSCNO.

SC-VNF on receiving the DL GTP-U packet decodes the GTP-U packet and identifies the traffic flow from the GTP TEID present in the GTP-U protocol header. SC-VNF after service chaining gets the GTP TEID associated with the SC-PNF for the traffic flow and forms the GTP-U packet by setting the GTP TEID associated with SC-PNF in the GTP-U protocol header and forwards the packet to destination IP address of the SC-PNF. SC-PNF identifies the traffic flow from the GTP TEID present in the GTP-U header and forwards the packet on the LTE-Uu interface.

2.1.5.1.4 Congestion Control via Blind Handover

The following diagram shows an example embodiment of congestion relief procedure at SC-VNF to offload the UEs in the event of user congestion.

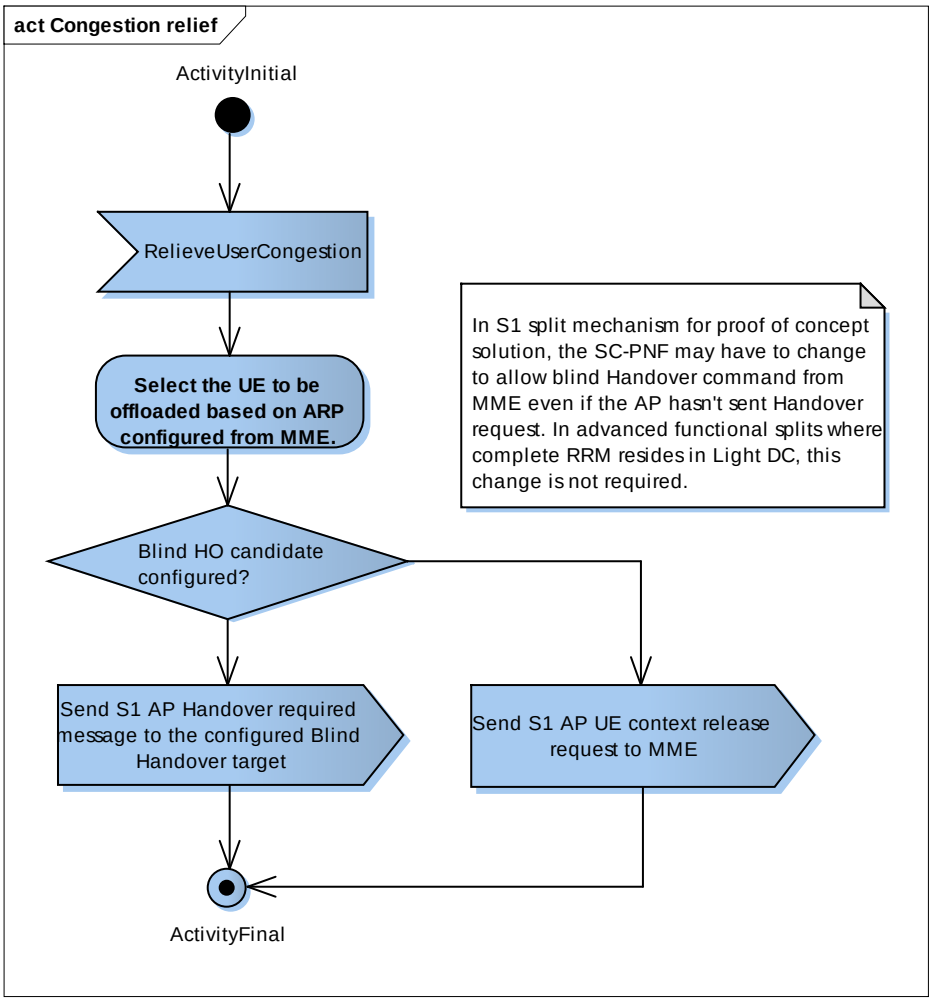


Figure 12: Congestion control at SC-VNF

The SC-VNF can perform UE context release procedure or Blind Hand-out of the calls based on congestion control policy agreed in SLA to a VSCNO specific redirection candidate cell.

2.1.5.1.5 DSCP marking per QCI

The DSCP mark for each QCI is configured on the SC-VNF by the SC-EMS. The SC-VNF has the bearer level QCI visibility and packs the appropriate DSCP based on E-RAB QCI.

2.1.5.1.6 S1AP Load balancing for S1-Flex

The SC-VNF is able to perform load balancing across multiple MMEs based on their relative *MME Capacity* as declared in the *S1 SETUP RESPONSE* message. When a UE requesting service does not indicate a valid serving MME, the SC-VNF uses a weighted random selection algorithm to select an MME from amongst the possible candidates.

2.1.5.2 Parameters, Alarms and PM Counters

2.1.5.2.1 Parameters

The following parameters control the behaviour of the SC-C-VNF:

Parameter	Description
BlindHOCandidateECGI	VSCNO specific Blind Hand-out candidate ECGI to offload the calls in the event of congestion.
DSCP per QCI	An array of value pairs defining the VSCNO specific DSCP marking per QCI.
AdmissionControlPolicy	VSCNO Admission control policy (Oversubscription allowed) from SLA.
ReservedBandwidth	VSCNO reserved bandwidth from SLA expressed as a percentage.
ReservedUserCount	VSCNO reserved UE count from SLA.
S1SigServerList	VSCNO specific S1 signalling server list.
S1ConnectionMode	VSCNO specific S1 connection mode (S1-Flex/Non flex).
PLMN ID	The PLMN ID served by this SC-VNF
OversubscriptionWeight	VSCNO specific weight (0 < weight < 1) attribute indicating share of front-haul bandwidth available for oversubscription.
MinBandwidth	VSCNO specific attribute indicating minimum percentage of available front-haul bandwidth assigned to VSCO in the absence of DL user plane traffic.
ConfiguredRemoteTrafficSelectors	Sec GW specific configuration not used in the PoC implementation.
DefaultIPsecEnable	
DefaultRemoteTrafficSelectors	
RemoteTrafficSelectorsInUse	
MaxUlBackhaulBandwidth	Maximum uplink backhaul bandwidth of the backhaul link towards SGW of the VSCNO.
Priority per QCI	An array of value pairs defining the VSCNO specific priority per QCI. Range of Priority values defined is from 1 to 15. Values between 1 and 14 are ordered in decreasing order of Priority and 15 indicates no priority.

Table 8: SC-VNF Parameters

2.1.5.2.2 Alarms

The SC-VNF is able to report the following alarm:

Field	Value
3GPP Probable Cause	Link Failure
Specific Problem	S1 Link Down
Alarm Type	COMMUNICATIONS_ALARM
Permitted Severities	MAJOR: Emitted when one of several MME links fail CRITICAL: Emitted when all MME links have failed or the SC-C-VNF link has failed and the SC-VNF is no longer able to provide service CLEARED: Emitted when all links are restored
Additional Text	Describes which link has failed and any associated error code

Table 9: SC-VNF Link Failure Alarm

2.1.5.2.3 PM Counters

In the PoC implementation, the SC-VNF does not generate and PM reports.

2.1.6 Call Flows

2.1.6.1 Default Bearer Setup

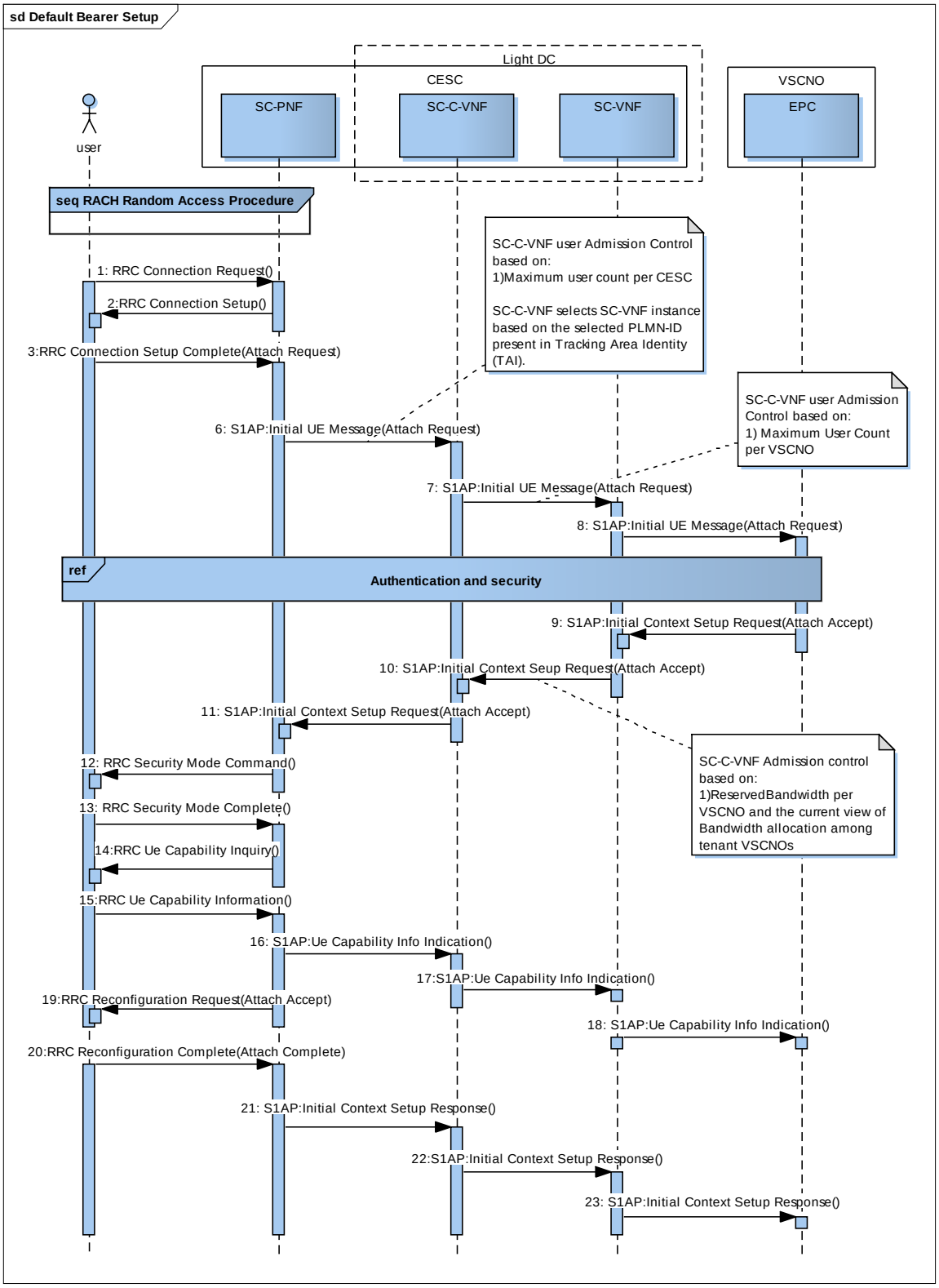


Figure 13: Default Bearer Setup Call Flow

Message	Description
1	UE sends [3GPP 36.331] RRC Connection request to establish a connection.
2	The SC-PNF sends [3GPP 36.331] RRC Connection Setup to the UE.
3	The UE transmits RRC Connection Setup Complete to the SC-PNF along with selectedPLMN-Identity, registered MME and dedicatedInfoNAS IE that carries the initial NAS message.
6	The SC-PNF transmits the S1 AP Initial UE message to the SC-C-VNF. The SC-C-VNF extracts the Selected PLMN of the UE from the TAI IE of the message and uses this to identify the SC-VNF to which this message needs forwarding. The SC-C-VNF performs PNF specific Admission control and forwards this message to the SC-VNF.
7	The SC-VNF forwards the Initial UE message to the VSCNO MME after performing VSCNO specific user admission control.
11	The MME performs authentication and security procedures with the UE and sends initial context setup request to the SC-VNF. The SC-VNF forwards the initial context setup request to SC-C-VNF. The SC-C-VNF performs bearer admission control with respect to allocated bandwidth per VNF and forwards it to SC-PNF.
12	The SC-PNF now sends RRC Security mode command which includes the AS security and integrity protection algorithms along with the START parameters for AS Security activation.
13	After successful security activation, the UE sends Security mode complete along with START Parameters to be used in uplink direction by the user plane.
14	The SC-PNF starts UE capability enquiry process following successful security mode procedure.
19	It then forwards the Attach accept message to the UE along with RRC Connection reconfiguration that establishes the SRB2 and the default bearer.
20	UE sends RRC Connection reconfiguration complete message followed by the Attach complete to SC-PNF.
21	The SC-PNF sends the Initial Context setup response confirming that the default bearer is established to the MME via SC-C-VNF and SC-VNF.

Table 10: Default Bearer Setup Messages

2.1.6.2 Admission control rejection of GBR bearer

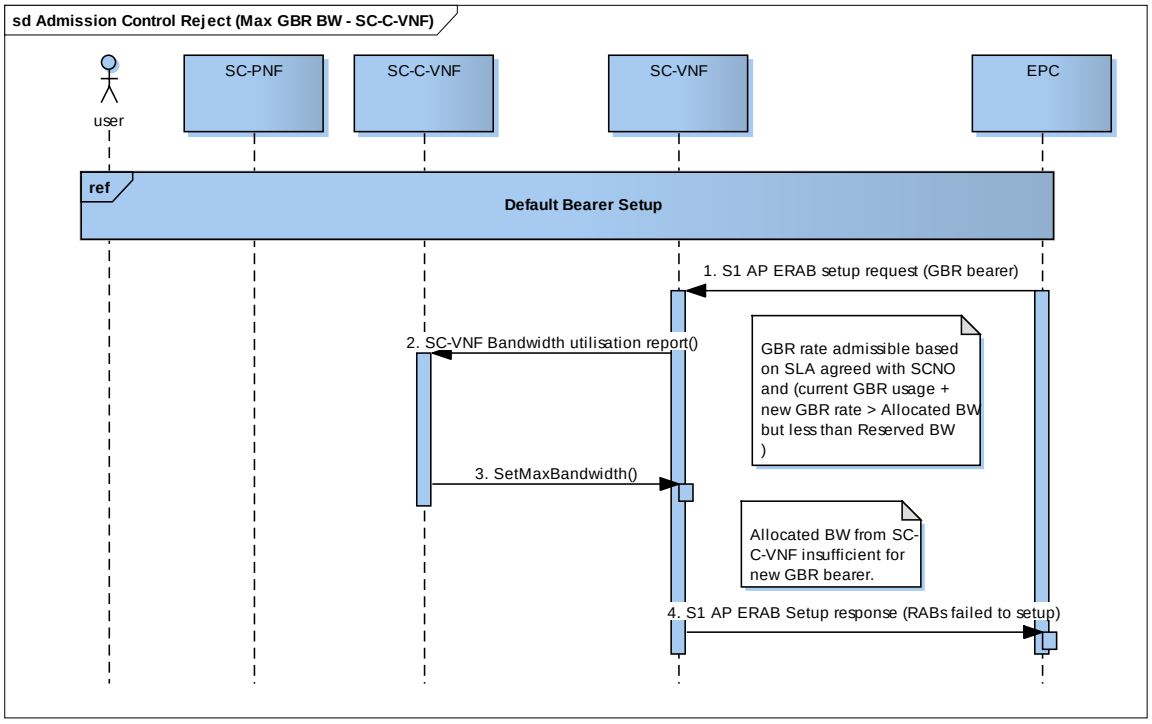


Figure 14: Admission Control rejection of GBR Bearer at Light DC

Message	Description
1	The MME sends ERAB setup request to SC-VNF for a UE, which has already established the default bearer. The SC-VNF performs bearer admission control by considering the GBR requirements of the new E-RAB from the VSCNO specific MME.
2	The SC-VNF sends a bandwidth utilization report to SC-C-VNF to indicate the new GBR usage requirement.
3	The SC-C-VNF allocates the new bandwidth to SC-VNF.
4	The SC-VNF sends S1 AP E-RAB setup response with the list of RABs failed to setup when the bandwidth allocated by SC-C-VNF is not sufficient to admit the new GBR bearer.

Table 11: Admission Control Rejection Messages

2.1.6.3 Admission control congestion at SC-C-VNF

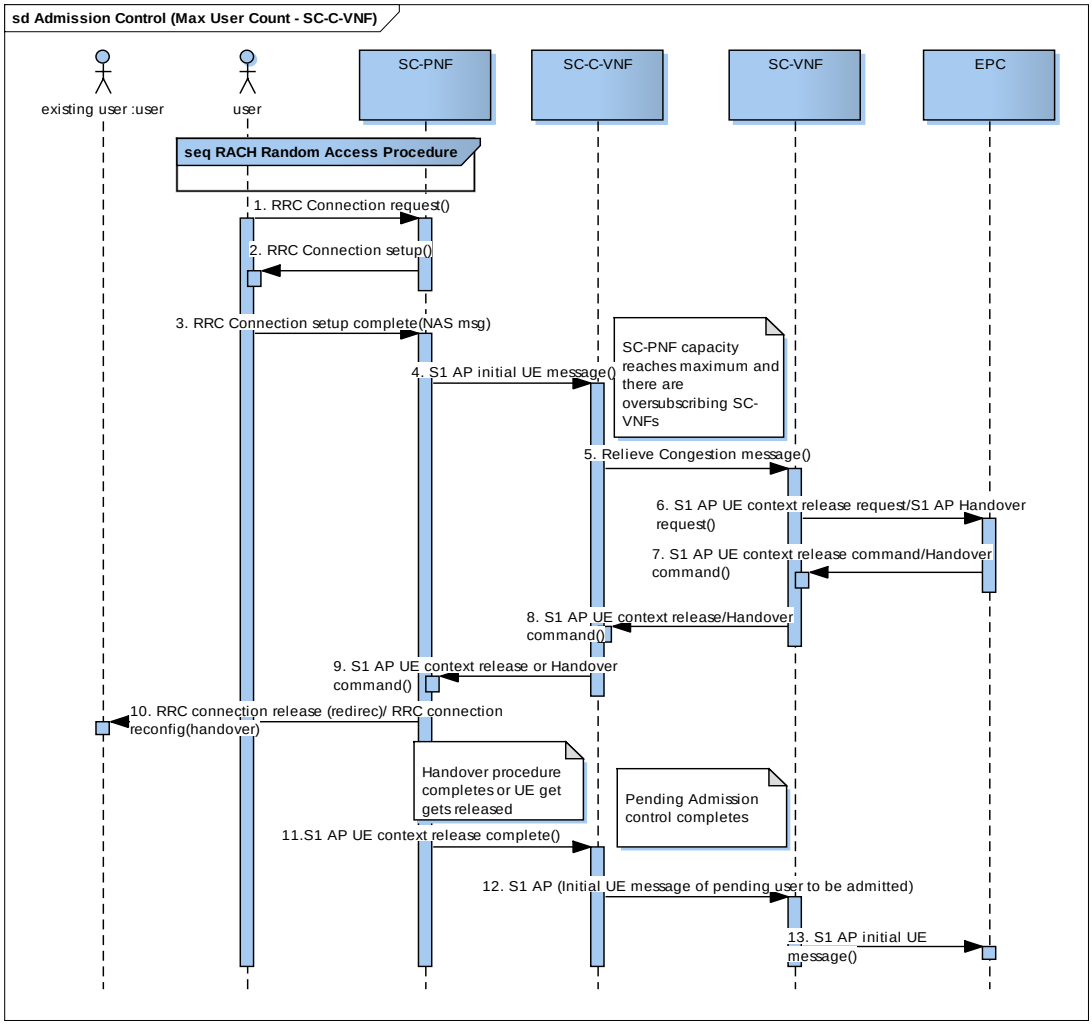


Figure 15: Admission Control Congestion at SC-C-VNF due to oversubscription of an SC-VNF

Message	Description
4	During default bearer establishment, SC-C-VNF detects SC-PNF congestion when it receives S1 AP Initial UE message.
5	The SC-C-VNF identifies that there exists an oversubscribing SC-VNF2 and sends Relieve congestion internal message to it.
6	The SC-VNF2 selects an existing UE to be offloaded based on Allocation retention priority and sends S1 AP UE context release request message to the MME.
7	MME sends S1 AP UE context release command to SC-VNF2 which gets routed transparently to SC-PNF.
11	When the S1 AP UE context release complete is received at the SC-C-VNF, the SC-C-C-VNF proceeds with pending UE context admission for SC-VNF1 as described in previous sequence diagrams.

Table 12: Admission Control Congestion Messages

2.1.6.4 User Admission control rejection at SC-C-VNF/SC-VNF

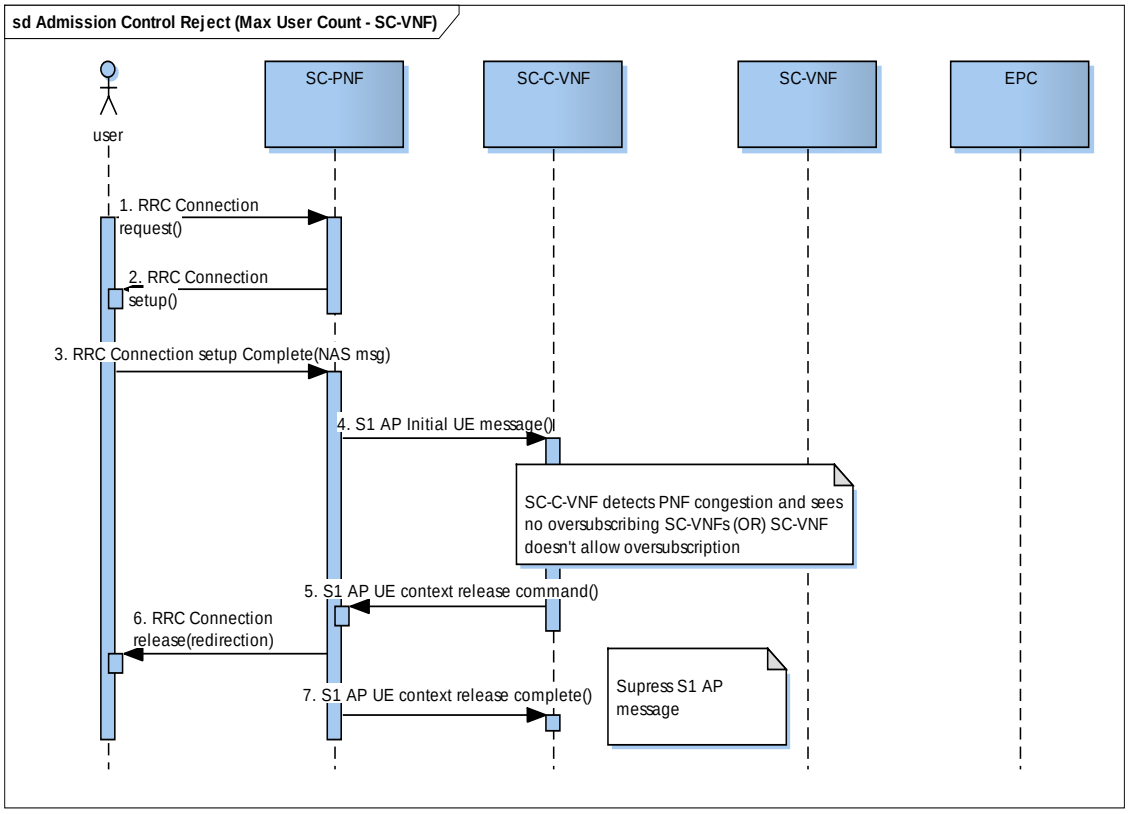


Figure 16: Admission Control Rejection at SC-C-VNF/SC-VNF due to PNF capacity breach

Message	Description
4	During default bearer establishment, SC-C-VNF detects SC-PNF congestion when it receives S1 AP Initial UE message.
5	The SC-C-VNF identifies that there is no oversubscribing SC-VNF in the system and sends S1 AP UE context release command to the SC-PNF.
6	The SC-PNF performs RRC Connection release with redirection to highest priority cell that matches the selected PLMN.
7	The SC-PNF sends S1 AP UE context release complete message to the SC-C-VNF. SC-C-VNF suppresses this message.

Table 13: Admission Control Rejection Messages

2.1.6.5 Call Release

The SC-VNF and SC-C-VNFs release the UE context resources by intercepting the standard S1 AP and X2 UE context release messages.

2.1.6.6 X2 setup via X2 GW

Note: The X2 GW is **not** part of the PoC implementation. The description in this section demonstrates that the architecture described in this document supports an X2 GW, should one be required.

The following diagram shows the X2 setup procedure using X2 GW procedures described in [4]:

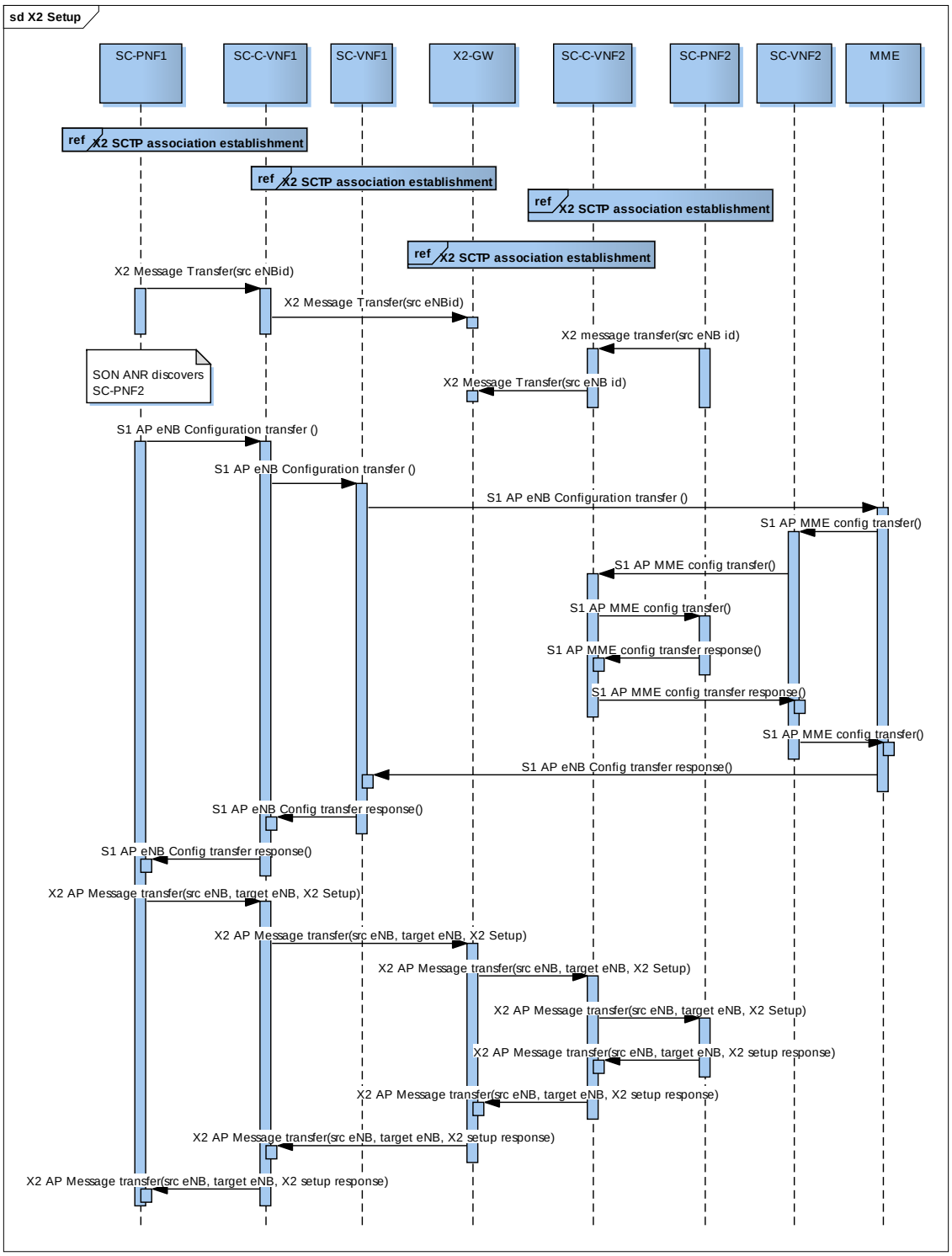


Figure 17: X2 Setup Procedure via X2 GW

In this method, the Light DC can implement the X2 GW functionality described in 3GPP 36.300 [4] for indirect X2 routing by using a single SCTP association per SC-PNF. As per this architecture, there will be one X2 GW for the whole CESC cluster in the Light DC. The PNF EMS configures the SC-PNF with the SC-C-VNF address as X2 GW address. The SC-C-VNF is configured with actual X2 GW address. The PNF uses the configured X2 GW address to establish X2 SCTP association with the SC-C-VNF. The SC-C-VNF creates another SCTP association towards the X2 GW. The SC-PNF then registers with X2 GW as per the protocol specified in [4] via SC-C-VNF.

The SON-ANR function in the PNF discovers the neighbour cells and initiates S1AP eNB configuration transfer message with X2 GW address included in ""eNB indirect X2 Transport Layer Address". In response to this message, the peer SC-PNF includes its X2 GW address in "eNB indirect X2 Transport Layer Address" IE in S1AP MME configuration transfer response.

The SC-PNFs now use the X2 GW address exchanged to setup an indirect X2 association, using X2AP Message transfer IE containing X2 setup message along with (source eNodeB id, target eNodeB id) IEs. Once the X2 setup succeeds, both eNBs use X2AP Message transfer mechanism to exchange X2 messages using standard X2 protocol by including source and destination eNB IDs always over the single SCTP interface.

X2 GW procedures in [4] inform the eNB when SCTP connection to target eNB is unavailable using X2 release message. This can clear the X2 association with target eNB that is unavailable in source eNB. For X2 U the GTP tunnels must be appropriately established for intra cluster routing of traffic.

2.1.6.7 Intra-Cluster (CESC to CESC) X2 handover via X2 GW

The following sequence diagram shows the X2 handover procedure via X2 GW. The basic protocol for handover remains the same as described in 3GPP 23.401 [85], except that the source SC-PNF to target SC-PNF messages are routed via SC-C-VNF and X2 GW. The SC-C-VNF performs Hand-in admission control considering the SC-PNF resources as a whole. The SC-C-VNF and SC-VNF communicate by using lightweight internal messages for vendor (VSCNO) specific admission control. The SC-VNF performs vendor specific admission control based on SLAs agreed with VSCNO.

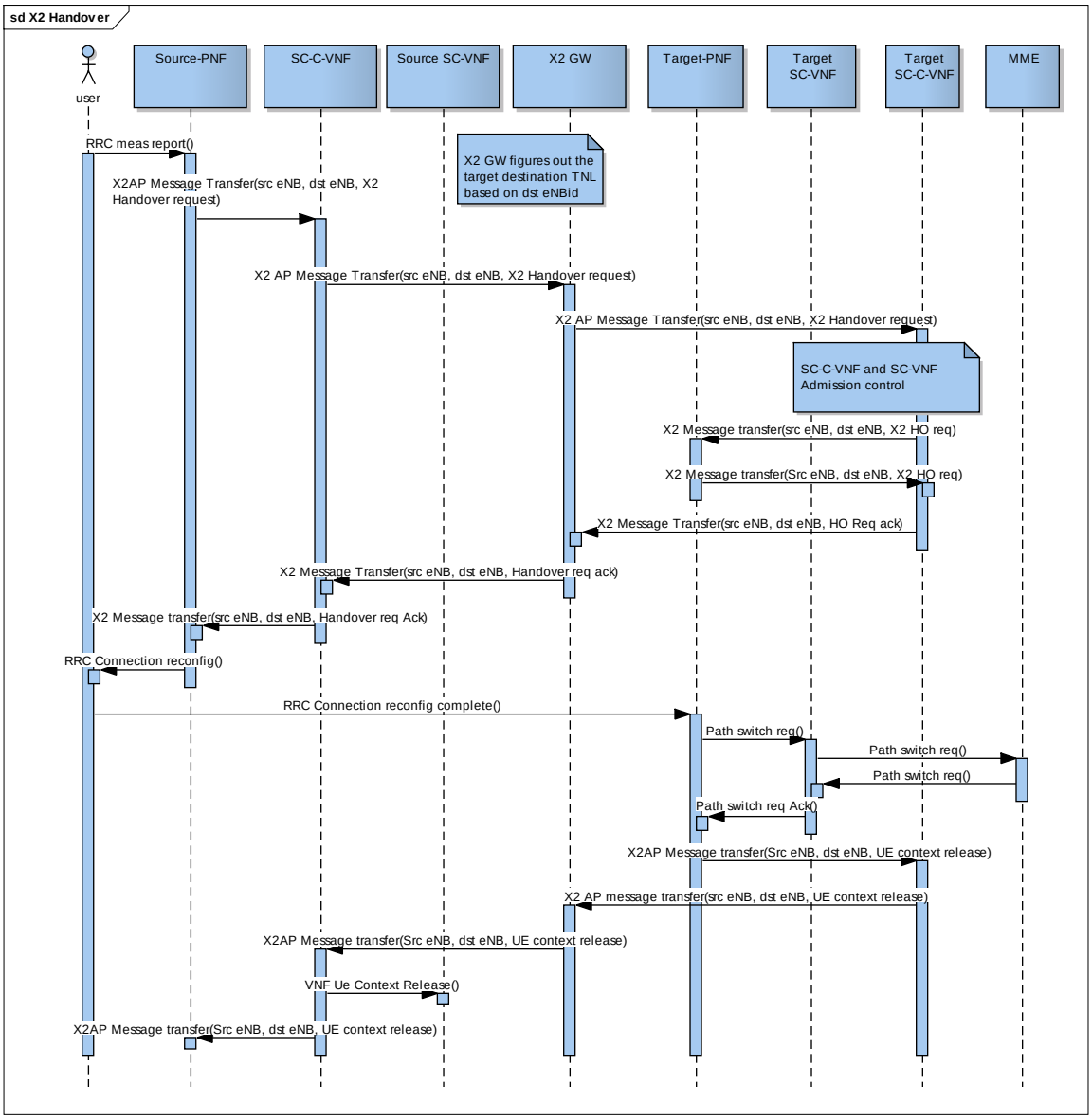


Figure 18: X2 Inter CESC Handover Procedure via X2 GW

2.1.6.8 X2 Setup without GW

The following diagram shows the X2 setup procedure without GW functionality which relies on internal routing of packets within light DC between SC-C-VNF's.

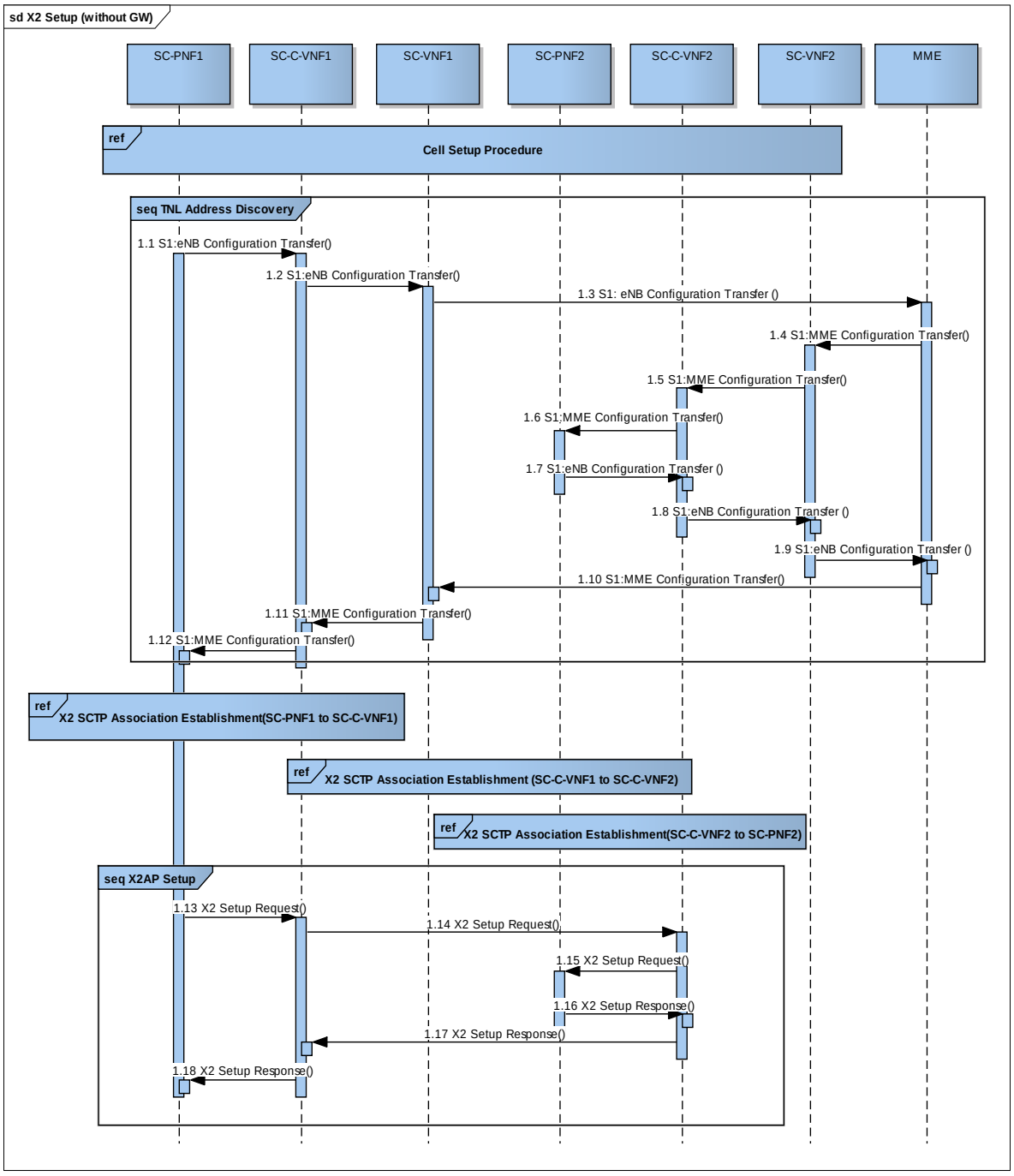


Figure 19: X2 Setup Procedures without GW

2.1.6.9 Intra-Cluster (CESC to CESC) X2 handover without GW

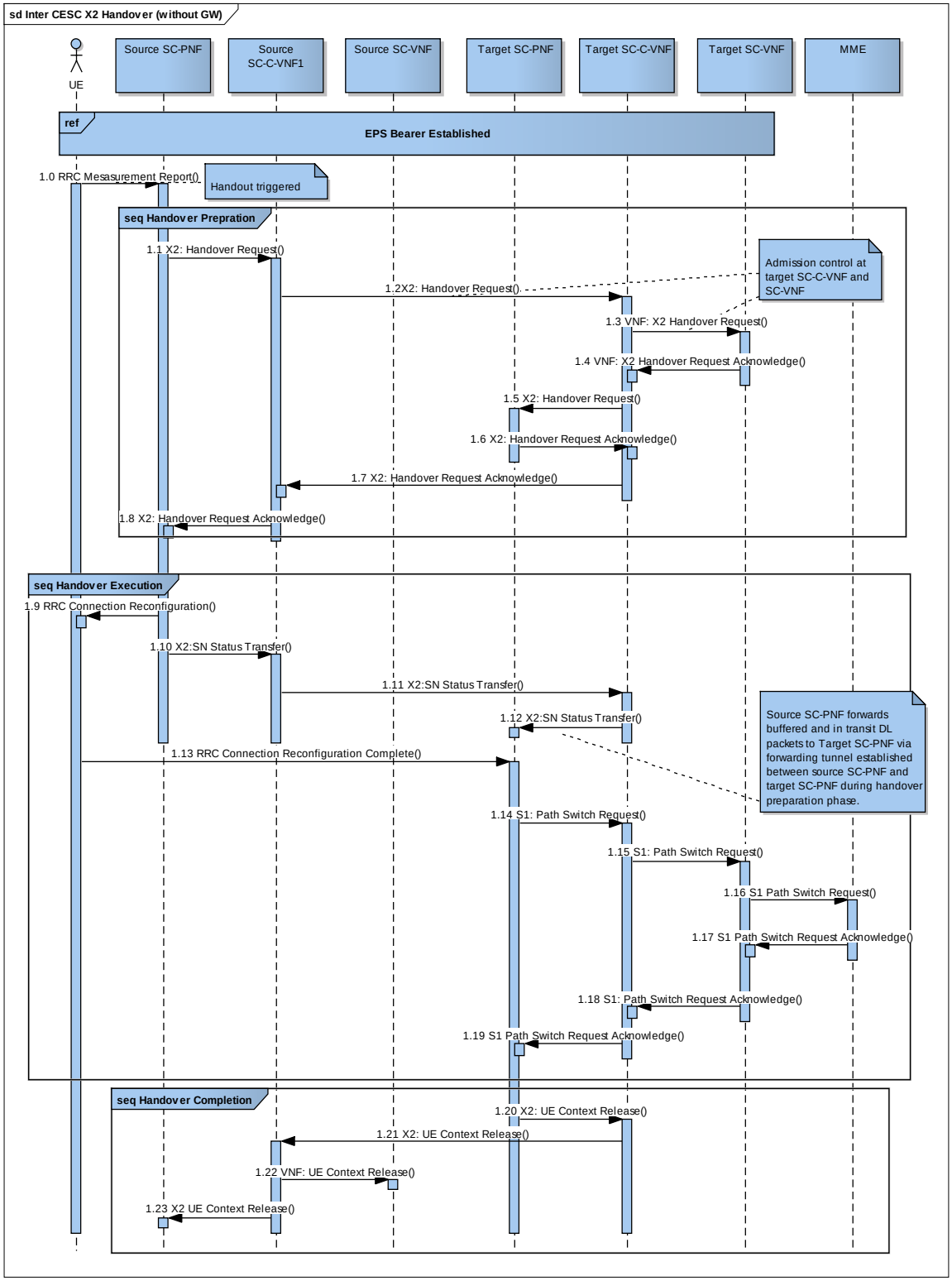


Figure 20: Inter CESC X2 Handover Procedure without GW

This sequence diagram shows the intra cluster X2 handover procedure via SC-C-VNFs within a cluster. This architecture assumes that direct communication is possible between source CESC

and target CESC for intra cluster Handover within a cluster. The basic protocol for handover remains the same as described in [85], except that the source SC-PNF to target SC-PNF messages are routed via SC-C-VNF and X2 GW. The SC-C-VNF performs Hand-in admission control considering the SC-PNF resources as a whole. The SC-C-VNF and SC-VNF communicate by using lightweight internal messages for vendor (VSCNO) specific admission control. The SC-VNF performs vendor specific admission control based on SLAs agreed with VSCNO.

2.1.6.10 Handin from macro cell to CESC

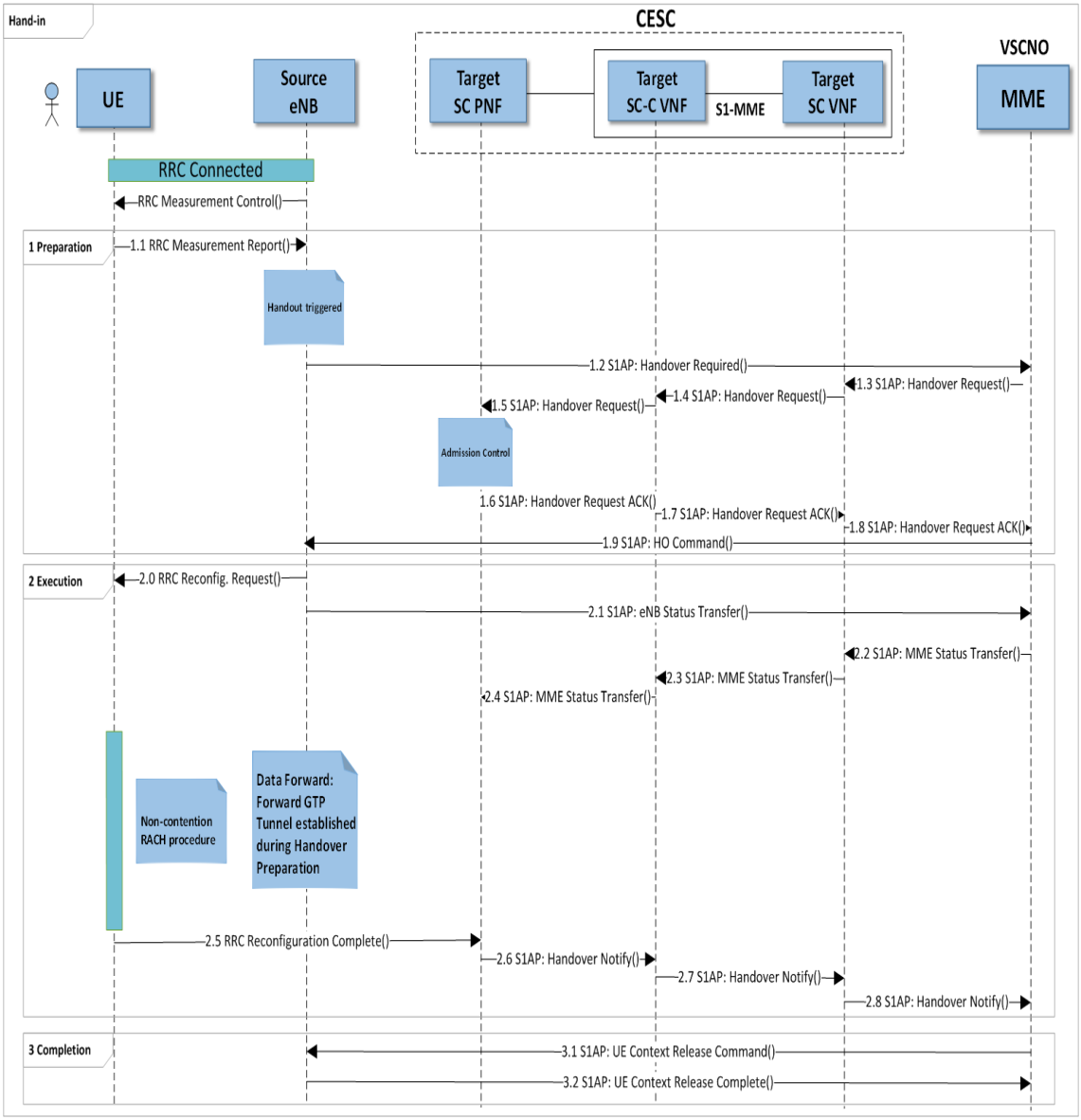


Figure 21: Hand-in from Macro Cell to CESC

This sequence diagram shows the S1 AP Hand-in sequence from macro cell to CESC. The basic protocol for S1 AP handover remains the same as described in[85], except for Hand-in admission control at target SC-C-VNF and SC-VNF.

2.1.6.11 Handout from CESC to macro cell

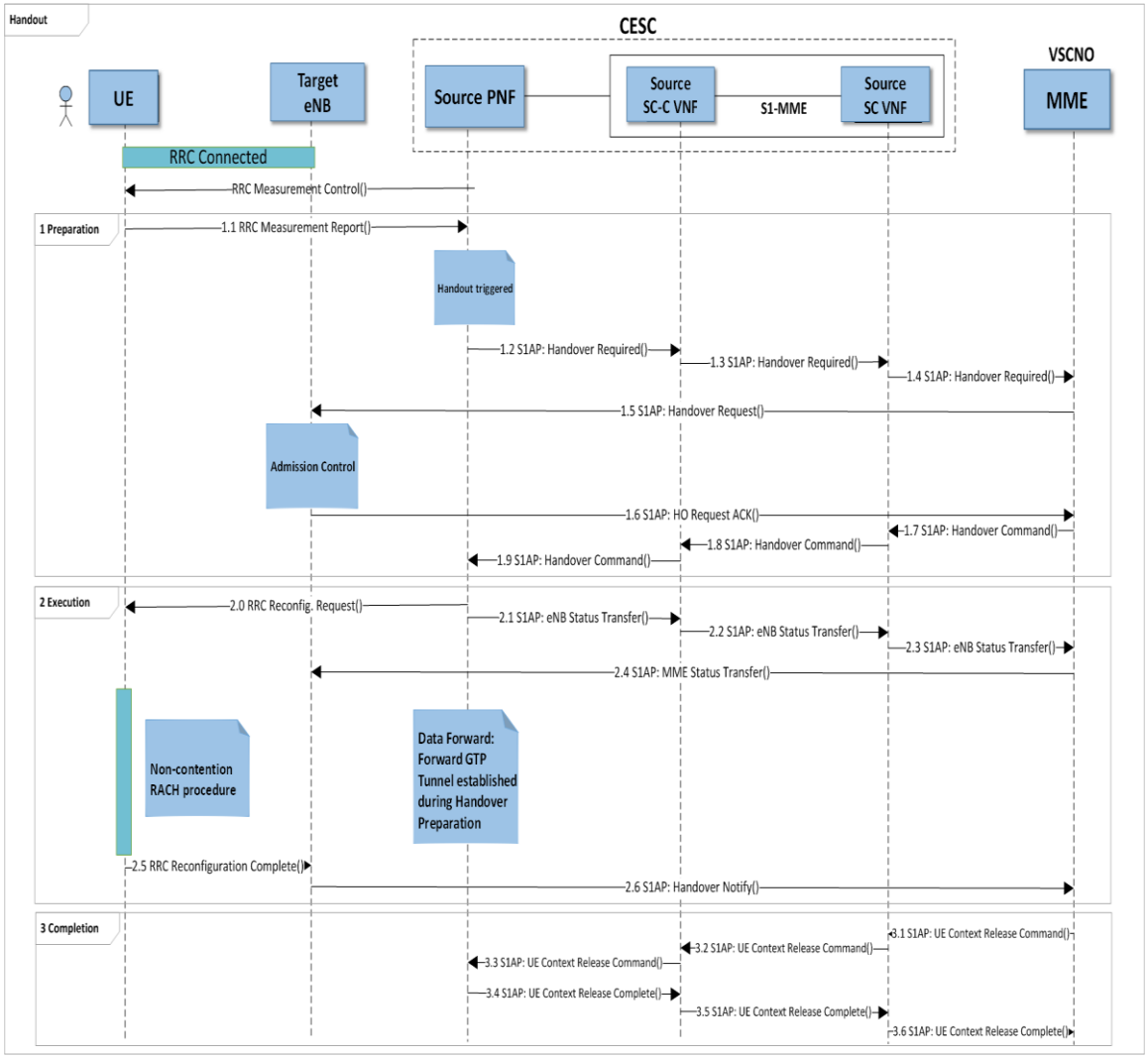


Figure 22: Handout from CESC to Macro Cell

Figure 22 shows the S1 AP Hand-out sequence from CESC to VSCNO macro cell. The basic protocol for S1 AP handover remains the same as described in [85].

2.1.7 Service Chains

A service chain is an ordered list of one or more functions that operate on the user plane data of a VSCNO. Each service in a service chain performs a specific task such as video transcoding or web caching and it is provided by a VNF running in the CESC. Service chains are attached to the SC-VNFs such that each VSCNO sharing the CESC may implement different service chains. They may be global to the SC-VNF and, *therefore*, apply to all of the VSCNO's user data or may be specific to certain classes of E-RAB, as distinguished by the QCI value.

The detailed implementation of service chains is not described in this document and will be provided as part of WP6. However, the current working assumptions are as follows:

- Both ends of the service chain are anchored in the SC-VNF; on receipt of a GTP-U packet, the SC-VNF pushes the packet into the service chain and receives it back again, once it has been processed by the entire chain. The SC-VNF then forwards the processed packet to its final destination. This anchoring allows calls and their associated service chain to be handed over from one CESC to another in the same Light DC without the need to deconstruct the service chain and without any loss of state.
- The beginning of a service chain is identified by a combination of IP address and port number that is recognised by the SDN and which is then mapped to the appropriate forwarding graph.
- There are separate service chain endpoints for uplink and downlink user plane traffic.

Figure 23 below illustrates the possible user plane traffic flow, for a VSCNO service chain:

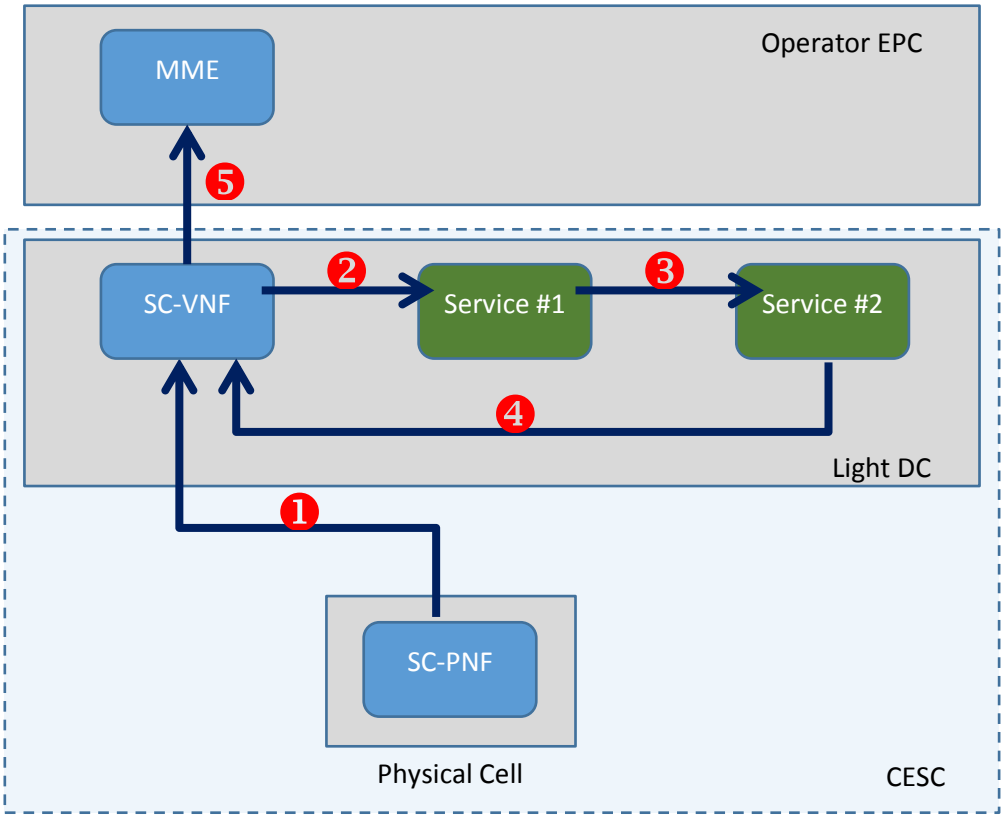


Figure 23: Service Chain Traffic Flow

2.1.8 SLA Definition and Monitoring

2.1.8.1 Definition of the SLA between the SCNO and the VSCNO

Legal, financial, technical and operational aspects for the implementation of a Small Cell as a Service (SCaaS) model between a SCNO and a VSCNO are captured through a specific SLA, as commonly done to formalise contractual agreements between service providers and customers. An SLA is a negotiated agreement that records a common understanding about the service and/or service behaviour offered by the SCNO, together with the measurable target values characterising the level of the offered service. The SLA for the SCaaS can be articulated around the following categories [18].

2.1.8.1.1 Service Scope

This category specifies both the geographical and temporal scope of the provided service:

- Geographical scope: The area where the service is provided (e.g. an enterprise, a stadium, a mall, etc.).
- Temporal scope: The time when the service has to be provided (starting time, end time, periodicities, etc.).

2.1.8.1.2 Service Specification

Essentially, the SCNO delivers a RAN service to the VSCNO so that VSCNO's customers (e.g. mobile subscribers) can be connected through the SCNO's CESCs to the VSCNO's core network. However, different service provisioning models are envisaged in this respect, as discussed in the following.

a) Capacity Based Provisioning

In this model, the service intends to provide a certain capacity to the VSCNO's subscribers over the temporal and geographical scope specified in the SLA. This model could fit, e.g., for a MVNO or a service provider that contracts the SCNO's RAN to provide service to its users in a given area. The specification of the capacity could be done in terms of aggregated global values (e.g. in Mb/s) or as a percentage share of the available bandwidth[26]. Limits can also be placed on the number and characteristics of the E-UTRAN Radio Access Bearers (E-RABs) that can be simultaneously established. The capacity specification can be further detailed including:

- Capacity conformance: It further specifies the provisioned capacity in time and space. Constraints can be established at spatial level (i.e., maximum Mb/s or maximum percentage share over a certain area) and temporal level (i.e., maximum Mb/s or percentage that can be offered within a certain time window).
- Excess capacity treatment: It specifies how the tenant's excess capacity demand not meeting the capacity conformance will be treated.
- E-RAB attributes and Key Performance Indicators (KPIs) targets , such as:
    - E-RAB accessibility KPI: Probability that an end-user is provided with an E-RAB at request. Alternatively, it can be expressed in terms of the blocking probability, which would be the complementary value to the accessibility.
    - E-RAB retainability KPI (dropping ratio): Probability that an end-user abnormally loses an E-RAB during the time the E-RAB is used.

[26] This latter option may be more practical due to the fact that the available bandwidth in a cell is not constant and varies according to the quality of the radio link to each individual user.

- E-RAB Quality of Service (QoS) parameters: QoS Class Identifier (QCI), Allocation and Retention Priority (ARP), Guaranteed Bit Rate (GBR) and Maximum Bit Rate (MRB) for GBR bearers.
- Per UE Aggregate Maximum Bit Rate (UE-AMBR): Limit on the aggregate bit rate that can be expected to be provided across all Non GBR bearers of a UE. It has an uplink and a downlink component.

- Dynamic capacity negotiation. The SCNO can offer, by automatic means, spare capacity as on-demand capacity to its VSCNOs. Specific mechanisms for querying, requesting and granting capacity based on certain policies should be in place between the SCNO provider and VSCNOs.

b) Capacity with Mobility Support

This provisioning model considers that the infrastructure of the SCNO will supplement the VSCNO's own RAN (e.g. the SCNO can deploy infrastructure inside a stadium, while the VSCNO, who in turn is a MNO, has macrocells deployed outside the stadium). In this context, and in addition to the terms already included in the case of the "Capacity Based" model, the MNO can be interested to include in the SLA the support of mobility between the CESCs of the SCNO and the rest of cells of the MNO. This may involve that the SCNO offers X2 interface connectivity between the CESCs and the cells of the tenant, and the SLA specifies the type of services supported through this interface (e.g. exchange of load information, handover support, etc.).

c) Customised Service

In this model, in addition to the provision of a certain capacity, the SCNO also offers to the VSCNO certain capabilities for carrying out selected operations in the shared CESCs. This opens the door to a much deeper involvement of the VSCNO in the way that the SCNO's infrastructure is managed, up to the extent that a VSCNO might envisage the operation of the CESCs in harmony with its own RAN. Different aspects can be considered in this respect:

- A VSCNO can specify its own algorithmic solutions for some selected RRM and Self-X functions (e.g. the VSCNO specifies the scheduling algorithm with corresponding automated parameter configuration through self-x, the tenant specifies a certain admission control strategy, etc.).
- Certain small cell parameters can be exposed to the tenant so that it can configure them (e.g. through its own self-x functions running at the tenant side).

It is worth mentioning that the achievement of isolation among VSCNOs should be implemented in a way that the customised configurations and algorithms enforced for one VSCNO do not affect the performance observed by the other VSCNOs.

2.1.8.1.3 Service Level Management aspects

This category specifies different elements, such as:

- Monitoring capabilities: Performance measurements, KPIs and alarms that the SCNO delivers to the VSCNO.
- Service availability: It specifies the percentage of the time that the service should be available to the VSCNO.
- Response to service related incidents: This specifies the response time to service related incidents notified by the VSCNO. Usually, incidents will be classified according to a certain priority level (High / Medium / Low) and different time frames will be associated to each priority.

- Changes in the SLA: This specifies the procedure to request changes in the SLA, and the conditions related to these changes (e.g. time to response, etc.).
- Accounting information: the SCNO needs to collect events supporting the accounting of resource usage by the UEs of a VSCNO (e.g. start of service by a UE of the tenant, end of service, etc.). These events may be delivered to the VSCNO.

2.1.8.2 Realisation of SLAs in the EMS

2.1.8.2.1 SLA Definition

Within the SESAME system, a Service Level Agreement is represented by a managed object that captures the measurable aspects of a contractual SLA. The parameters of a Service Level Agreement object are used in two distinct cases:

- At provisioning time, they contribute to the configuration an SC-VNF and associated SC-PNF to define resource limits such as the maximum number of UEs that are supported by the SC-VNF or the maximum bandwidth that the SC-VNF can support.
- During VNF operation, they define a number of performance thresholds against which KPIs are compared. If a threshold violation is detected, a configured action is taken.

At any specific time, each SC-VNF has at most a single associated SLA. The scope of this SLA may vary in both space and time; it may encompass a single virtual small cell, a set of virtual small cells (such as a venue) or every small cell belonging to the VSCNO. Similarly, an SLA may have a limited duration covering, for example, a specific event or it may apply indefinitely.

The SLA management hierarchy is illustrated in Figure 24, below. There is a single "SLAs" collection object beneath which there may be an unlimited number of "SLA" objects. Each "SLA" object represents a single, specific, service level agreement and applies to a single VSCNO. Each VSCNO may have multiple service level agreements agreed with the SCNO. Where this is the case, every SLA operates in parallel.

Within the EMS, each SLA object has the following key scope parameters:

- Geographic Scope –this defines the set of virtual small cells over which the SLA applies. This set be one of:
    1. A single, specific, virtual cell as defined by the DN of the SC-VNF managed object,
    2. A list of specific virtual cells (typically less the 10) as defined by the DNs of the associated SC-VNF objects,
    3. *A set of virtual cells serving a specific geographic area. All cells falling within a defined geographic area are included in the SLA*,
    4. All of the virtual cells owned by the VSCNO. There can only be one such SLA and this is known as the VSCNO's default SLA.

  Note: Only options 1, 2 and 4 are implemented by the PoC.
- Temporal Scope – this defines the period over which the SLA operates. SLAs may have a start time and date, and end time and date and a recurrence period. Thus, the temporal scope of an SLA supports:
    - SLAs that operate on or after a specific start date,
    - SLAs that cease to operate after a specific end date,
    - SLAs that operate according to a pre-determined schedule such as, *for example*, every weekday, every month or every year.

As an SLA affects the configuration of the CESC, the EMS is responsible for ensuring that, at any one time, at most one SLA applies to a given SC-VNF instance.

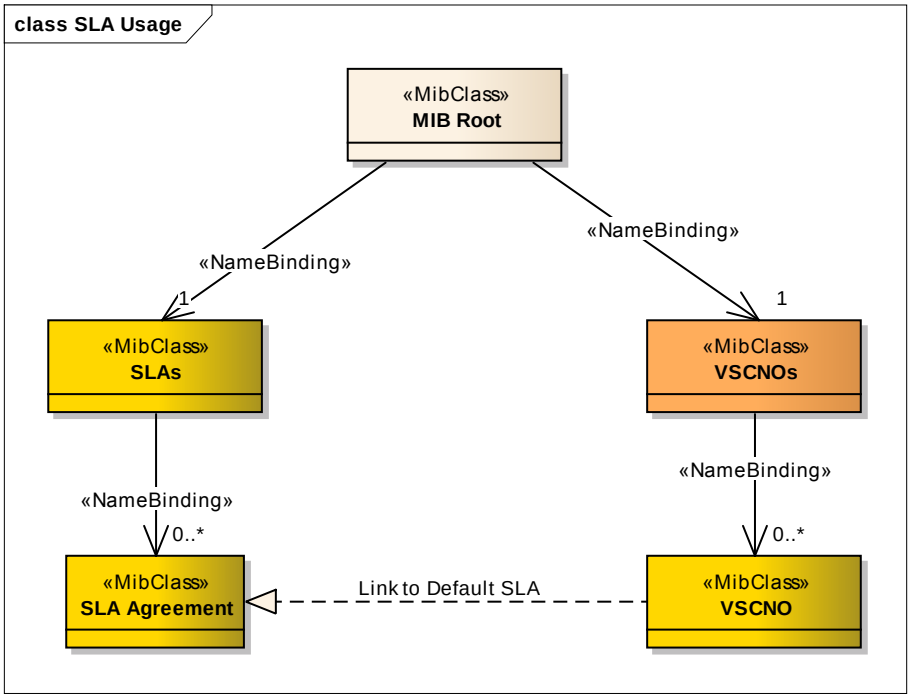


Figure 24: SLA Managed Object Hierarchy

SLA objects are created, as required, by the SCNO to represent new service level agreements and the parameters of each SLA object capture the details of the service level agreement such that they can be retrieved and acted upon by the Provisioning function described in *section 2.1.9.3.2.1* and the SLA Monitoring function described in *section 2.1.8*. In a given release of the system, each SLA object supports an identical set of parameters. The nature of a specific SLA is determined by the values of these parameters and, *where appropriate*, these include special "Not set" and "Not monitored" values. There are two broad classes of parameter:

- Parameters that define the resource share made available to the VSCNO; for example cell bandwidth or maximum number of UEs. Parameters of this class are utilised when configuring an SC-PNF to support a VSCNO's SC-VNF.
- Parameters that define a threshold related to a service level associated with the SLA; for example a minimum availability, throughput or delay. The values of these parameters are compared against a KPI calculated by the SLA Monitoring function and, in the case of an SLA breach, the specified action is taken. This action may, for example, comprise raising an alarm, scaling up the associated VNF or triggering a SON-*related* action.

2.1.8.2.2 General Parameters

The following parameters provide information about the SLA object:

Parameter	Access	Description
Object Name	Read-Write Max length 50 chars.	This is a common property of every managed object. In this context, it is used to provide each SLA with a unique user-friendly name and, thus, there is a requirement that the Object Name is unique within the scope of the SLAs collection.

Parameter	Access	Description
SLA Description	Read-Write Max length 200 chars.	This is a free-form text field that allows a description to be provided for each SLA object. Other than maximum length and character set, there are no restrictions placed on the content of this parameter.
VSCNO PLMN ID	Read-only	The PLMN ID of the VSCNO to which this SLA applies.
Geographic Scope	Read-only enum	Possible options are: - "VSCNO" – The SLA applies to every virtual small cell owned by the VSCNO, - "List of cells" – The SLA applies to a list of one or more specific virtual small cells identified by the DN of their SC-VNF, - "Geographic Area" – The SLA applies with coordinates that fall within the area defined by a polygon of latitude – longitude points.
Temporal Scope	Structure	A structure comprising four elements: - Start Time – The time that the SLA comes into force (default current time). - End Time – The time that the SLA ceases to be enforced (default blank) - Number of Occurrences – The number of times that the SLA is activated (default 1). The value zero has special significance meaning indefinite. - Recurrence Period – The period between successive invocations of the SLA. Default daily.
Northbound KPI Server Address	Read-Write String: URL	If not blank, this specifies the address of a northbound server to which KPIs are to be uploaded.
Upload Protocol	Read-Write enumeration: "HTTP" or "HTTPS" Default: "HTTP"	Specifies the protocol to be used for northbound KPI upload.
Upload Method	Read-Write enumeration: "PUT" or "POST" Default: "POST"	Specifies the HTTP method to be used for northbound KPI upload.
SLA Version	Read-only Integer	This value is incremented by the EMS whenever a change is made to the configuration of an SLA object. As the configured values are likely to reflect a commercial agreement, the SCNO is not free to change them

Parameter	Access	Description
		without such changes being recorded.
Last Update Time	Read-only String: Date & Time	This value is set to the current date and time whenever a change is made to the configuration of the SLA.
Last Validation Time	Read-only String: Date & Time	This value is set to the date and time that the SLA was last checked.
Last Outcome	Read-only enumeration: "Met", "In breach"	Indicates the result of the last SLA check operation.

Table 14: General SLA Parameters

2.1.8.2.3 Configuration Related Parameters

The following parameters are used by the provisioning and configuration changes processes when instantiating or updating an SC-VNF.

Parameter	Access	Description
Priority	Read-Write integer Range: 0-99	This provides a relative priority for the SLA compared to other SLAs with the value zero representing the highest priority. When the system needs to take action under congestion conditions such as dropping calls, throttling bandwidth or dropping users it uses the SLA priority of each VSCNO as input to the load-shedding decision.
Minimum UEs	Read-Write integer	This provides the guaranteed minimum number of UEs belonging to the VSCNO that can be simultaneously in "connected mode" on a virtual small cell. The sum of this parameter across all of the VSCNOs using the SC-PNF cannot exceed the capacity of the SC-PNF.
Maximum UEs	Read-Write integer	This provides the maximum number of "connected mode" UEs that a VSCNO with this SLA may have on a given virtual small cell. Even if there is spare capacity, the VSCNO may not exceed this value.
Minimum GBR Bandwidth %	Read-Write integer Range: 0-100	This provides the minimum guaranteed bandwidth for GBR bearers provided to VSCNOs with this SLA, expressed as a percentage of the capacity of the SC-PNF. The sum of this parameter across all of the VSCNOs using the SC-PNF cannot exceed 100%.
Maximum Bandwidth %	Read-Write integer	This specifies the maximum bandwidth share that the VSCNO can consume on a CESC, re-

Parameter	Access	Description
	Range: 0-100	gardless of whether or not there is spare capacity.

Table 15: VNF Configuration Related SLA Parameters

2.1.8.2.4 Monitored KPI Value Parameters

The following parameters are used by the SLA Monitoring function described in *section 2.1.8.2.5* to periodically validate that the SLA is being met:

The following KPIs are all derived from standard PM reports produced the SC-PNF. For a given SLA, KPIs are calculated as an aggregate value across the set of SC-VNFs encompassed by the SLA scope.

KPI Class	KPI	Description
Availability	Service Uptime (%)	The percentage of time that the monitored component, service or set of components (as defined by the SLA scope) has been able to provide service during the sample period. Managed elements are typically deemed to be providing service when they are administratively unlocked and operationally enabled (as defined by [82]).
Accessibility	RRC Connection Establishment success rate (%)	The number of successful RRC Connection Establishments expressed as a percentage of the total number of RRC Connection Establishment attempts.
	RRC Connection Re-establishment success rate (%)	The number of successful RRC Connection Re-establishments expressed as a percentage of the total number of RRC Connection Establishment attempts.
	Call setup success rate (%)	The number of successful E-RAB set-up attempts expressed as a percentage of the total number of E-RAB set-up attempts.
Retainability	Call Drop Rate (%)	The number of E-RABs released due to failure expressed as a percentage of the number of successful E-RAB set-up attempts.
	Inter-Enb Hand-Out Success Rate (%)	The number of successful inter-eNB handouts expressed as a percentage of the total number of inter-eNb handout attempts.
	Intra-frequency Hand-Out Success Rate (%)	The number of successful intra-frequency handouts expressed as a percentage of the total number of intra-frequency handout attempts.
	Inter-frequency Hand-Out Success Rate (%)	The number of successful inter-frequency handouts expressed as a percentage of the total number of inter-frequency handout at-

KPI Class	KPI	Description
		tempts.
Quality	Downlink Packet Drop rate	Where GTP-U sequence numbers are in use this counts the total number of GTP-U packets that were expected but did not arrive expressed as a percentage of the total number of packets sent. Note that this can only be calculated in the downlink and that the core network must perform the corresponding calculation for uplink traffic.
	Average Sub-band CQI	The average value of the sub-band CQI (Channel Quality Indicator) reported by UEs in the cell.
	Wideband CQI Distribution	The distribution of the Wideband CQI (Channel Quality Indicator) reported by UEs in the cell.
	Timing Advance Distribution	The distribution of the Timing Advance values transmitted by the LTE AP to UEs in the cell. Note that only non-zero counts are reported
Utilisation	Mean Active UEs	The average number of active UEs during the Monitor period
	Max Active UEs	The peak number of active UEs during the Monitor period
	Uplink Packets	The total number of uplink packets sent to the core network, indexed per QCI value.
	Uplink Octets	The total number of uplink octets sent to the core network, expressed in megabytes, indexed per QCI value.
	Downlink Packets	The total number of downlink packets sent to the core network, indexed per QCI value.
	Downlink Octets	The total number of downlink octets sent to the core network, expressed in megabytes, indexed per QCI value.
	DL Bandwidth Share %	The share of the downlink bandwidth consumed by the VSCNO expressed as a percentage of the total downlink octets received by the SC-PNF.
	UL Bandwidth Share %	The share of the uplink bandwidth consumed by the VSCNO expressed as a percentage of the total uplink octets transmitted by the SC-PNF.

Table 16: Monitored KPI Values

Note: Depending upon the nature of the item being monitored, some KPIs can be calculated separately for each individual VSCNO whereas others can only be calculated for the call as a whole.

Within the SLA managed object, each monitored parameter shares a common structure comprising the following elements:

Element	Type	Description
Display Name	String	A user friendly name for the KPI. For example "Call Drop Rate".
Formal Name	String Default: ""	Where the KPI relates directly to some other formally defined aspect such as a PM counter in the MIB, this element provides the formal name of the related element.
Monitored	Enumeration: "Yes", "No" Default: "No"	This element determines whether or not the parameter is monitored as part of the SLA. Initially, when an SLA object is created, all parameters are not monitored.
Threshold	Float	This element provides the threshold value for the KPI.
Threshold Trigger	Enumeration: "Above", "Below", "Equal" Default: "Above"	Specifies at what point the current value of the KPI is deemed to have reached or crossed the threshold, triggering the associated action.
Hysteresis	Float	Once a threshold trigger event has occurred, this value specifies the amount by which the current value must re-cross the threshold before the threshold trigger is cancelled.
Monitor Period	Enumeration: "1 hour".. "30 days" Default: "1 hour"	A value chosen from a set on a broadly exponential scale, in the range of one hour to 30 days. This value specifies the period over which the KPI is calculated. Note that the range of possible values is constrained to be a multiple of the Monitor period of the reporting network element.
Action on Breach	Enumeration: "Do nothing", "Log", "Raise Alarm" Default: "Log"	Specifies the action to be taken by the SLA Monitor when the monitored value meets or crosses the specified threshold. Note future actions might include actions such as "Scale-up", "Scale-down", "Throttle" and "Optimise".

Table 17: Common SLA parameters

Using the above definitions, the configuration of an SLA parameter for the "Availability" KPI might be:

Element	Value
Display Name	“Availability”
Formal Name	“”
Monitored	“Yes”
Threshold	0.99 (99%)
Threshold Trigger	“Below”
Hysteresis	0.01 (1 %)
Monitor Period	“Daily”
Action on Breach	“Raise Alarm”

Table 18: Example SLA parameter

2.1.8.2.5 SLA Monitoring Function

The SLA Monitoring function is effectively part of the EMS. For scalability reasons there may be multiple instances but, for the purposes of the following discussion, it is considered a single entity.

The SLA Monitor processes SLA objects. It runs periodically, according to a schedule aligned with the PM reporting interval of the system and operates as follows:

On each invocation, the SLA Monitor iterates over the set of extant SLA objects:

- For each *active* SLA object[27] it collects the PM reports from the set of managed elements defined by the Scope of the SLA.
- It aggregates the data into a set of consolidated values.
- It calculates the KPIs monitored by the SLA.
- If the SLA object specifies a northbound server for KPI upload, the calculated KPIs are uploaded to this server using HTTP/HTTPS PUT.
- For each monitored KPI defined by the SLA object, the SLA Monitor compares the current value to the specified threshold:
    - If a threshold breach is detected and the same threshold was not breached in the previous monitor period, the configured action is taken.
    - If the current value if within the configured threshold by the hysteresis amount and there was a threshold breach in the previous monitor period then the configured clearing action is taken.

[27] An SLA object is active when it is within its temporal scope.

2.1.9 EMS Functions

Note: For the PoC implementation, the EMS functionality will be based on the existing IP.Access Network Orchestration System (NOS) product. Much of the description in the following sub-sections "reflects" the features of this product. However, due the use of international standards based interfaces, the use of other EMS solutions are possible in the future.

The Element Management System (EMS) consists of a set of services (applications) responsible for the management of specific network elements and serves as part of an overall network management system (NMS). The EMS is generally responsible for the following set of management functions:

- Fault Management: This involves the detection of anomalies and the initiation of relevant recovery mechanisms.
- Configuration: Depending on the observed parameters in the associated network element, the EMS is responsible for configurations and re-configuration of the parameters that affect the performance of the network element (in this case the SC-PNF).
- Accounting: This involves keeping track of the offered service for a measure against predefined service level considerations that may be provided by other modules.
- Performance: Making sure that the network elements performs to the required service levels.
- Security: Ensuring that access to information and control parameters is secured and authorized.

In the scope of SESAME, the security aspect is considered as part of configuration and fault management tasks together with exception detection and raising alarms when security exceptions are detected. Additionally, accounting management is considered outside the scope of CESC management because the accounting is done on top of performance and certain required information is collected outside SESAME architecture. Therefore, the management scope in CESC is confined to configuration management, fault management and performance management and the aspects of security and accounting are abstracted.

Whilst the functionality of the EMS is divided into a number or discrete areas, in the PoC implementation, this functionality is provided by a single EMS with the interfaces between each areas being provided by shared data and internal APIs.

2.1.9.1 Common EMS Functions

The EMS aspects (PNF-EMS, SC-EMS and Service EMS) of the SESAME architecture illustrated in Figure 1 are all provided by a single, common, EMS. The different functions required by the architecture are delivered primarily by the roles of individual EMS users and the access rights they are assigned to different classes of component (see *section 2.1.10.3*). This approach has a number of advantages:

- Only a single EMS requires development for the PoC.
- Inter-EMS linkages and associated fragility are eliminated.
- Cross-EMS data duplication is eliminated.

Regardless of the class component being managed, the SESAME EMS provides a set of core functions. These include the following, as discussed in the subsequent sub-sections:

2.1.9.1.1 Client-Server Architecture

The EMS comprises two main parts:

- A server, providing database and file storage plus a set of services that implement the bulk of its functionality.

- A client that provides a graphical user interface (GUI) to users of the EMS. Multiple instances (at least 20) of the EMS client may be operational at any one time, each instance providing management facilities to an EMS user. The EMS client also provides supporting business logic such as the data-entry time validation of parameter values against definitions held in the Management Information Base (MIB).

The services of the EMS Server include the following:

Service	Description
GUI Client service	Authenticates and Connects clients to the server to provide the CM, FM and PM functionality to EMS client users.
TR-069 Management Service	Performs configuration management (CM) and fault management (FM) of physical small cells to deliver SC-PNF functionality.
Lightweight Management Service	Performs configuration management of "lightweight" components by the selected mechanism (see *section 2.1.15.2*).
FM File Service	Accepts fault management reports from "lightweight" components using HTTP(S).
PM File Service	Accepts performance management (PM) reports from components using HTTP(S). Stores the reports in the database and post-processes them into VSCNO specific variants for upload by the PM Upload Service.
PM Upload Service	Uploads VSCNO specific PM reports to the configured servers using HTTP(S).
Northbound CM Service	Provides the 3GPP compliant northbound CM interface.
Northbound FM Service	Provides the 3GPP compliant northbound FM interface.
Northbound PM Service	Provides the 3GPP compliant northbound PM interface.
SESAME Business Logic Service	Provides the bridge between SLAs, SC-VNFs and SC-PNFs. Merges SLA and SC-VNF configuration in order to generate the SC-Common-VNF and SC-PNF configuration.
SESAME SLA Monitoring Service	Periodically inspects SLA objects as described in *section 2.1.8.2.5* to determine whether the parameters of the SLA are being met.

Table 19: EMS Services

Another key property of the EMS is a Management Information Base (MIB). This is a highly structured XML document describing the CM, FM and PM properties of the system. The information contained in the MIB includes:

- The managed objects that may exist. The number of instances there may be of each object and their parent-child relationships (name bindings) that form the managed object hierarchy.
- The parameters (attributes) of each managed object, their permitted and default values.
- The management actions (such as lock, unlock and reinitialise) that may be performed on each managed object.
- The alarms that each managed object may raise.
- The performance management counters that each managed object supports.

2.1.9.1.2 Configuration Management View

The EMS Client Configuration Management View provides the user with a tree view of those parts of the system that they are permitted to see. According to the access rights that the client user has, they are able to perform configuration management operations that include:

- Viewing the parameters of any managed object to which they have access.
- Creating and deleting managed objects.
- Making changes to the configuration of a managed object by changing one or more parameter values.
- Invoking management actions, such as "lock" and "unlock" (see [82]) on managed objects.

EMS business logic supports these operations in the following ways:

- In controls the user's access according to their rights.
- It validates whether or not an object can be created or deleted according to rules in the MIB and the current state of the object or its parent.
- It validates the value of any parameter according to its MIB definition.
- It validates whether or not an action may be performed according to rules in the MIB and the object's current state.
- It propagates changes to the configuration of the managed object to the associated component by means of the appropriate management interface (see *section 2.1.15*).

2.1.9.1.3 Fault Management View

The EMS Client Fault Management View provides the user with a view of the alarms currently active[28] on the managed objects to which they have Fault Management access. Alarms handled by the EMS conform to the format described in X.733 [84] and comprise the following fields:

X.733 Alarm Field	Supported Values
Alarm Identifier	Uniquely identifies the alarm during its lifetime
Event (Alarm) type	Communications, Processing, Environment, Equipment, IntegrityViolation, OperationalViolation, PhysicalViolation, SecurityViolation, TimeDomainViolation
Probable cause	Text value describing alarm cause
Specific Problem	Unique text or integer value providing additional information
Perceived severity	Cleared, Indeterminate, Critical, Major, Minor, Warning, Cleared
Additional Text	Free form text field
Managed Object Class	Text defining class of object raising the alarm
Managed Object Instance	Text comprising the distinguished name of object instance raising the alarm
Current Time	The time and date that the alarm event was generated
Additional Text	Appended to the alarm by the source component to provide additional information to help in diagnosis

Table 20: Alarm Fields Displayed in FM View

[28] An active alarm is one that currently has a severity other than "Cleared".

In addition, for each unique alarm, the MIB also contains the following textual information that is displayed when an alarm is viewed by the EMS client in Fault Management view:

X.733 Alarm Field	Supported Values
Behaviour	Provides a short description of the alarm and what it means.
Possible Fault	For each Permitted Severity, describes the conditions that could cause an alarm of the associated severity to be emitted,
Repair Action	Describes possible repair actions that could be undertaken to clear the alarm.

Table 21: Alarm Help Displayed in FM View

In fault management view, the EMS user may filter the list of active alarms based on a number of criteria that include:

- The managed object or sub-tree on which the alarm is raised.
- The severity of the alarm (for example “Critical”).
- The time window in which the alarm was raised.
- The alarm type (for example “Processing Error” or “Environmental”).

The EMS user is also able to perform the following operations on each alarm:

- Acknowledge it. This simply indicates that they have seen the alarm and is useful if they take action such as, for example, dispatching an engineer to rectify the fault.
- Add a comment to the alarm. This feature is useful for recording the investigative and corrective steps that have been taken to deal with the alarm.
- Delete the alarm. When an alarm is no longer of interest, the EMS user may delete it from the display. Note that “Cleared” alarms are automatically deleted after a configured time.

2.1.9.1.4 Performance Management View

The EMS Client Performance Management View allows the EMS user to view the KPI reports provided by the system and, where appropriate, generate an on-demand report.

2.1.9.2 PNF-EMS

In SESAME, depending on the chosen functional split, the SC-PNF may represent different levels of management and control requirements. For example, in the SESAME PoC implementation, the SC-PNF represents a standard eNB or HeNB functionality and therefore its management, at the very least, requires the same level of management and control as required for a standard eNB/HeNB. The PNF-EMS has a dedicated interface with the SC-PNF through which it can perform its management tasks.

In SESAME architecture, the PNF-EMS is conceptually part of the CESC manager and takes care of all the management tasks for the SC-PNF through its dedicated interfaces and internal data structures. The PNF-EMS has an interface with SC-EMS (described below), with SLA monitoring module and with CESCM portal through northbound interface. Because of the functional split between SC-PNF and SC-VNF, the respective PNF-EMS and SC-EMS have to jointly oversee the management tasks and therefore have a dedicated interface between them. The interface with the SLA monitoring module provides the necessary information required for management tasks such as configuration, accounting and performance.

2.1.9.2.1 Managed object model and hierarchy

PNF-EMS functionality is available to EMS users with access rights of "Operational" and above (see *section 2.1.10.4*) to the APs sub-tree as illustrated in Figure 25 below:

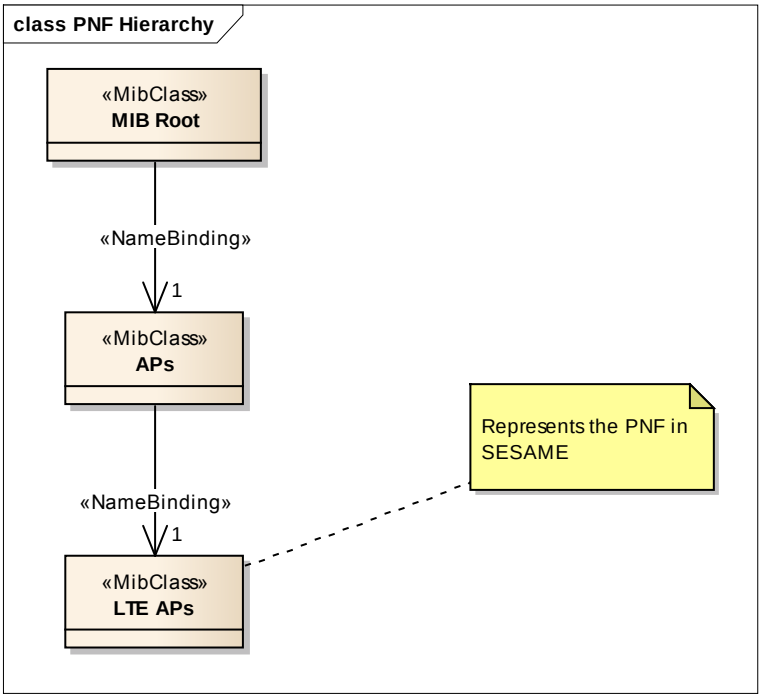


Figure 25: LTE AP sub-tree

The management operations that can be performed on these objects are:

- Provisioning – The managed objects representing a single specific physical cell are created. The parameters of the AP object are assigned initial values that include:
  - Its geographical location,
  - Its Cell Global Identity,
  - Its transmit frequency and power output.

  Note that, following initial provisioning, the physical cell is not enabled to transmit by the EMS until at least on SC-VNF has been assigned to it.
- Make configuration changes to those aspects that are common to every SC-VNF hosted by the cell.
- Take the cell in and out of service and reset it by means of the lock, unlock and reinitialise management actions.
- View alarms raised by the cell in Fault Management view.
- View PNF related KPI reports in Performance Management view.
- Decommission the cell when it is no longer required. Note that the EMS will prevent the user from decommissioning the cell if there are still SC-VNFs hosted by it.

2.1.9.2.2 PNF-EMS Business Logic

Following any change to the configuration of an SC-PNF, the EMS synchronises the configuration held in its database with that of the physical cell by means of the standard TR-069 [79] management interface (see *section 2.1.15.1*). Configuration changes may be applied immediately or may be deferred until the physical cell next routinely connects to the EMS.

2.1.9.3 SC-EMS

As mentioned before, the functional split results in the division of the required network functions into physical and virtual network functions. The lower layer functions below the chosen split are realised in the PNF through hardware implementation. The higher layer functions above the chosen functional split are realised through VNFs in the CESC. Together, the SC-PNF and SC-VNF operate in conjunction and provide the complete functionality of a small cell that is seen by the tenant EPC just like a regular small cell. While the PNF-EMS takes care of management of SC-PNF, the SC-EMS takes care of the management of SC-VNFs. The broad classes of the SC-EMS management functionalities are similar to the ones described for PNF-EMS, i.e. configuration, fault and performance management.

Residing in the CESCM, the SC-EMS has interfaces with PNF-EMS, SLA monitoring module, CESCM portal, VNF Manager, and with the managed SC-VNF. The SC-EMS also provides the observation and control window to the VSCNO through which it can observe the state of managed SC-VNFs in the CESC.

The SC-VNFs are tenant specific thus each tenant/VSCNO may view only those managed objects that relate to their own network slice and are not aware of the rest of managed objects in the CESC. It should be noted that since the SC-VNFs provide a sub-set of the functions of a regular small cell (the rest being realised in PNF) therefore the SC-EMS functionality also relates to that subset.

2.1.9.3.1 Managed object model and hierarchy

SC-EMS functionality is available to EMS users with access rights of “Operational” and above (see *section 2.1.10.4*) to the VSCNO sub-tree as illustrated in Figure 26 below. Note that although there are multiple instances of the VSCNO sub-tree beneath the VSCNOs collection object (highlighted in pale red in the diagram) an individual VSCNO only has access to the single instance that represents their virtual network.

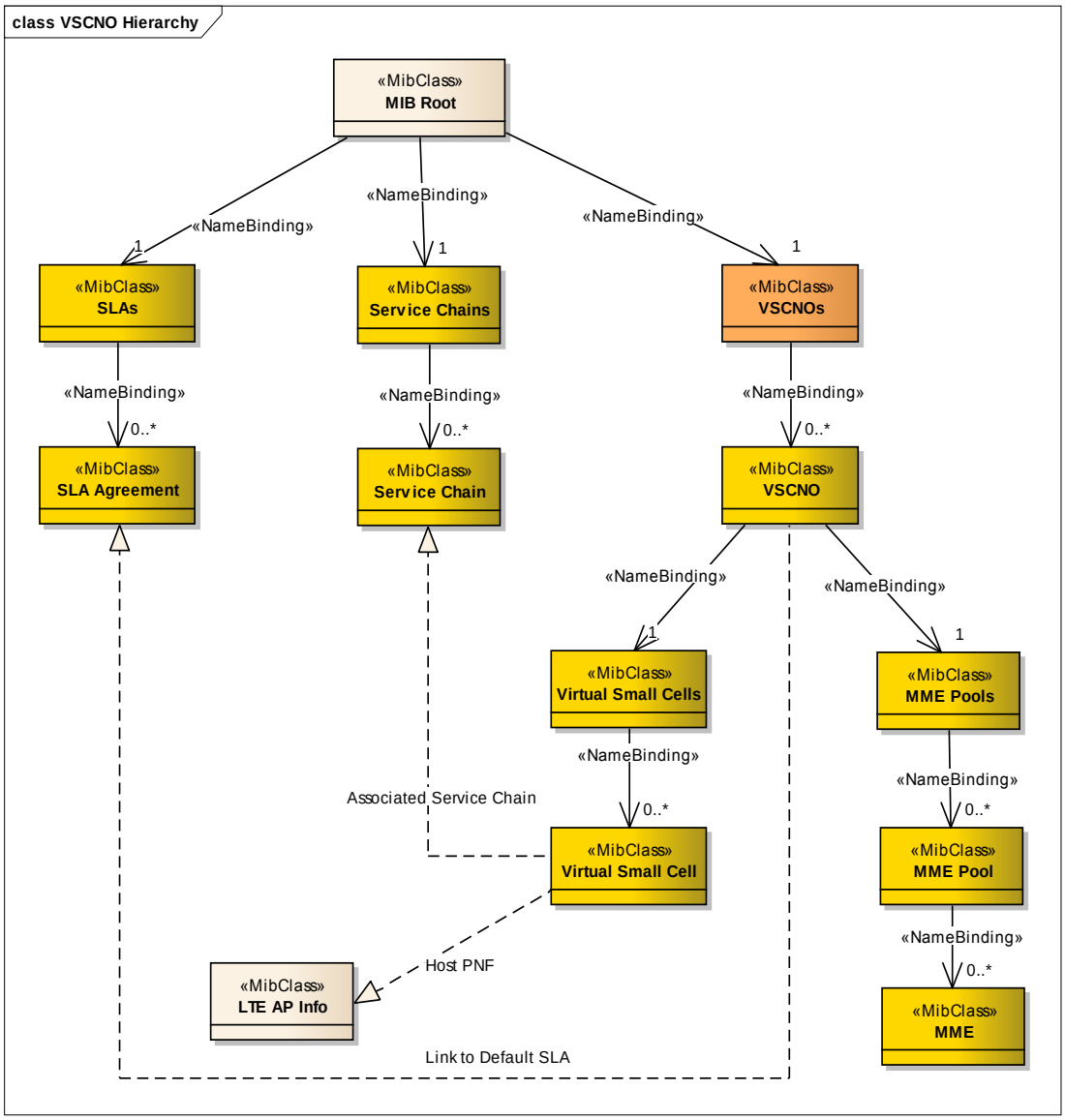


Figure 26: VSCNO Hierarchy

Configuration changes made to objects in the VSCNO sub-tree result in an invocation of EMS business logic that merges the contributions from individual SC-VNFs in order to generate the configuration of the SC-PNF hosting them. The details of the managed objects that comprise the VSCNO sub-tree are as follows:

2.1.9.3.1.1 VSCNOs Object

This is a collection object, beneath which child VSCNO objects may be created. A single instance is created automatically by the EMS and it cannot be deleted. Whilst the SCNO has full access to the object in order to create and delete VSNO objects, VSCNOs have no access.

2.1.9.3.1.2 VSCNO Object

Each instance of this object represents an individual VSCNO and captures the details of their network slice. The parameters of the VSCNO object include:

Parameter	Access	Description
Object name	Read-Write string	This is a standard property of every managed object. In the case of the VSCNO object, it is used to provide a user friendly name identifying the VSCNO.
PLMN ID	Read-Write string	This is the PLMN ID identifying the VSCNO. It is broadcast by each physical cell supporting a virtual cell belonging to the VSCNO and, therefore, must be unique within the scope of the system.
Default SLA	Read-Write DN	An optional distinguished name of an SLA managed object. If specified, this represents the SLA to be applied to virtual small cells belonging to this VSCNO when no other SLA applies.
Default Carrier Frequency List	Read-Write Array of Integer	Defines a list of carrier frequencies on which neighbour cells supporting the VSCNO's PLMN might be found. Propagated to the *Carrier Frequency List* parameter of an SC-VNF at provisioning time.

Table 22: VSCNO Object Parameters

Whilst each VSCNO has read-only access to their own VSCNO object and may create and delete a variety of child object types, only the SCNO is allowed full access.

2.1.9.3.1.3 Virtual Small Cells Object

This is a collection object beneath which child Virtual Small Cell objects may be created. A single instance of this object is created automatically by the EMS when the VSNO object is created and it cannot be deleted by a VSCNO EMS user.

2.1.9.3.1.4 Virtual Small Cell Object

Each instance of this object represents a single virtual small cell and provides the configuration of an SC-VNF instance. They are created, configured and deleted by the owning VSCNO. Their configurable parameters include:

Parameter	Access	Description
Object name	Read-Write String	This is a standard property of every managed object. In the case of the Virtual Small Cell object it allows the VSCNO to specify a user friendly name for the cell.
EMS Identity	Read-only String	This is a unique alphanumeric string generated by the EMS when the virtual small cell object is created. It is passed to the SC-VNF when it is instantiated enabling it to identify itself to the EMS.
Management Address	Read-only String	An FQDN or IP address by which the SC-VNF can be reached by the EMS for management. It is populated by the EMS when the

Parameter	Access	Description
		SC-PNF makes contact following start-up.
Host SC-PNF	Read-only DN	The distinguished name of the SC-PNF hosting this virtual small cell. This is normally assigned at object create time and cannot be changed directly by the VSCNO.
Administrative State	Read-only enumeration: "Locked", "Unlocked"	Set by the SC-VNF in response to the "lock" and "unlock" management actions. When locked, the SC-VNF does not have an S1 connection to an MME and the SC-PNF is configured to mark the VNF's PLMN as *reserved for operator use*, prohibiting normal UEs from attempting to access the cell.
Operational Status	Read-only enumeration: "Enabled", "Disabled"	Set by the SC-VNF to indicate its ability to provide service.
MME Address List	Read-Write string	A list of FQDNs or IP addresses of MMEs to which the SC-VNF should establish S1 connections.
Service Chain	Read-Write DN	The distinguished name of an, optional, Service Chain managed object. If specified, the referenced object defines a service chain that is linked to the SC-VNF.
Latitude	Read-Write Integer	The latitude (specified in millionths of a degree) of the *desired* coordinates of the virtual cell. Specified at provisioning time, the *desired* coordinates are used by the EMS in two cases: 1. To select an SC-PNF to host the SC-VNF 2. To identify cells falling within the geographic scope of an SLA.
Longitude	Read-Write Integer	The longitude (specified in millionths of a degree) of the *desired* coordinates of the virtual cell.
Location Tolerance	Read-Write Integer	The degree of tolerance (specified in metres) permitted between the desired location of the virtual cell and the actual location of the physical cell.
Carrier Frequency List	Read-Write Array of Integer	Defines a list of carrier frequencies on which neighbour cells supporting the VSCNO's PLMN might be found. Propagated to the SC-PNF to optimise its SON ANR scan.

Table 23: Virtual Small Cell Object Parameters

2.1.9.3.1.5 Service Chains Object

This is a collection object, beneath which child Service Chain objects may be created. A single instance of this object is created automatically by the EMS when the VSCNO object is created and it cannot be deleted.

2.1.9.3.1.6 Service Chain Object

An instance of this object class is created by the SCNO EMS user to represent a service chain to which the VSCNO has access. Service chains are realised by the NFVO and the primary function of the EMS is simply to capture the unique identifier by which the service chain is defined in the Local Catalogue and to quote this identifier to the NFVO when an SC-VNF is provisioned or modified. Hence, the Service Chain object has very few parameters:

Parameter	Access	Description
Object name	Read-Write String	This is a standard property of every managed object. In the case of the Service Chain object it allows the SCNO to specify a user friendly name for the chain.
VNFO Identifier	Read-only String	The unique identifier of the Service Chain that identifies it in the local catalogue.

Table 24: Service Chain Object Parameters

The details of how service chains are deployed and instantiated are defined by task 6.2 and are outside the scope of this document.

2.1.9.3.1.7 MME Pools Object

This is a collection object beneath which child MME Pool objects may be created. A single instance of this object is created automatically by the EMS when the VSCNO object is created and it cannot be deleted by a VSCNO EMS user.

2.1.9.3.1.8 MME Pool Object

An instance of this object class is created by the VSCNO EMS user to represent a specific MME Pool. An MME Pool is a collection of MMEs that service one or more complete tracking areas. Child MME objects define the members of the pool. The parameters of the MME Pool object include:

Parameter	Access	Description
MME Group Id	Read-Write Integer	The MMEGI is the unique identity of MME Pool within the context of PLMN as specified in [72].
TAC List	Read-Write Array of Integer	The Tracking Area Code list associated with the MME Pool

Table 25: MME Pool Object Parameters

2.1.9.3.1.9 MME Object

An instance of this object class is created by the VSCNO EMS user to represent a specific MME within an MME Pool. The parameters of the MME object include:

Parameter	Access	Description
Object name	Read-Write String	This is a standard property of every managed object. In the case of the MME object it allows the SCNO to specify a user friendly name for the MME.
MME Code	Read-Write Integer	The MME Code (in the range 1 to 255) is unique within the MME pool area and, if overlapping pool areas are in use, unique within the area of overlapping MME pools as described in [72].
MME Address	Read-Write String	The address (IP address or FQDN) of the MME.

Table 26: MME Object Parameters

MME objects are used during the Virtual Small Cell provisioning process and all the user to select pre-defined MMEs rather than have to enter the above details each time. If S1-Flex functionality is implemented, the EMS user may select one or more complete MME pools to which the virtual small cell connects.

2.1.9.3.1.10 LTE AP Info Object

An instance of this object class is created by the SCNO EMS user as part of SC-PNF provisioning. Its parameters contain information that the EMS needs to know about the physical small cell but excludes the actual physical small cell configuration. The parameters of the LTE AP Info object include:

Parameter	Access	Description
Object name	Read-Write String	This is a standard property of every managed object. In the case of the AP Info object it allows the SCNO to specify a user friendly name for the site.
Equipment Identity	Read-Write String	This parameter is used to control which small cells can connect to the EMS. It provides the identity of the physical cell managed by the EMS. The format of the Equipment Identity: <OUI>-<Serial Number> Where <OUI> is the Organisationally Unique Identifier and <Serial Number> is the unique identifier of the physical device.

Table 27: AP Info Object Parameters

2.1.9.3.1.11 LTE AP Object

The LTE AP object represents the configuration of the SC-PNF and corresponds to the standard TR-196 data model [80] for small cells. This data model contains several hundred parameters and is not reproduced here. For a description of those parameters of direct relevance to SESAME, please refer to the description of the SC-PNF in *section 2.1.3*.

2.1.9.3.1.12 SLAs Object

This is a collection object beneath which child SLA objects may be created. A single instance of this object is created automatically by the EMS and it cannot be deleted by any user.

2.1.9.3.1.13 SLA Object

An instance of this object class is created by the SCNO EMS user to represent the details of a specific service level agreement. There may be many instances of this object class that define:

- SLAs required by individual VSCNOs.
- SLAs that apply to specific geographic areas (such as a venue).
- SLAs that have specific temporal limits (such as some form of event).
- Any combination of the above.

The parameters of an SLA object are used by the EMS for two purposes:

- When configuring the SC-PNF, the EMS uses the parameters of the SLA to define the SC-VNF's share of network resources.
- The SLA Monitoring function described in *section 2.1.8.2.5* compares KPI reports with active SLAs to determine whether the SLA is being met.

Refer to section 2.1.8.2 for a list of the parameters of the SLA object.

2.1.9.3.2 SC-EMS Business Logic

The main aspects of the SC-EMS business logic are:

- Mapping an SC-VNF to a host SC-PNF at provisioning time.
- Merging the contributions from each SC-VNF hosted by a SC-PNF with its applicable SLA and propagating them into the configuration of the SC-C-VNF and SC-PNF.
- Receiving PM reports from physical cells and post-processing them into corresponding reports for individual VSCNOs.
- Running the SLA Monitoring Function, calculating KPIs values from received PM reports and comparing them against the associated SLA parameters.

2.1.9.3.2.1 SC-VNF Hosting

In the PoC implementation, physical cells and their associated SC-PNF must be provisioned before an SC-VNF can be provisioned. During the provisioning process, the VSCNO EMS user specifies the desired location of the virtual small cell and a permitted tolerance plus the SLA to be applied. The EMS uses these parameters to identify a CESC that provides the coverage in the desired area requested by the VSCNO. Such a CESC must have less than six SC-VNFs already assigned and not include one already serving the VSCNO's PLMN ID.

Note: The following steps are provisional at this stage and will be finalised in task 7.2:

- *The EMS selects the CESC that best matches the VSCNO's needs and attempts to complete the provisioning step by instructing the NFVO to instantiate the appropriate set of*

VNFs on the CESC by quoting to the local catalogue identifier referenced in the associated SLA.

- *In turn, the NFVO instructs the Virtual Infrastructure Manager (VIM) to instantiate the VNFs on the microserver forming part of the CESC.*
- *If the resource needs of the VNFs, as stated by the resource templates in the Local Catalogue, exceed the remaining capacity of the microserver, the provisioning request fails.*

If no suitable host CESC can be identified, the provisioning operation fails. The EMS user is informed of this fact and must contact the SCNO to request deployment of a new physical cell.

The ability to provision an SC-VNF without a pre-existing host SC-PNF is FFS.

2.1.9.3.2.2 Configuration Merging

The SC EMS business logic merges the following aspects of SC-VNF and SLA configuration:

- The PLMN ID of each SC-VNF is used to construct the *PLMNList* objects of the SC-PNF. If the associated SC-VNF is administratively unlocked and operationally enabled, the value of the *PLMNList* object's *CellReservedForOperatorUse* parameter is set to "false" otherwise, it is set to "true",
- The *Carrier Frequency List* parameter of the SC-VNF is merged with the values from other SC-VNFs hosted by the SC-PNF to derive the parameter values of the SC-PNF's *Device.Services.FAPService.{i}.REM.LTE.CarrierMeas* objects.
- The resource limits identified by the SLA associated with the SC-VNF are propagated to the SC-Common VNF so that it is able to police them,
- The SC-PNF is, optionally, instructed to perform an ANR scan in order to identify neighbour cells supporting a newly configured PLMN ID[29],
- If the VSCNO is permitted to configure their own neighbour cells, the EMS propagates the details of these cells to the SC-PNF. **Note:** This option is not available in the PoC implementation.

2.1.9.3.2.3 SLA Monitoring

See section 2.1.8.2.5 for a description of the SLA monitoring functionality.

2.1.9.4 Service EMS

Similar to the functions of the PNF-EMS and SC-EMS, the Service EMS is responsible for the tasks of fault management, configuration management and performance management. The scope of these functions in the Service EMS is limited, due to the fact that service VNFs are self-contained with very few configuration parameters exposed to the network management system. The Service EMS, **however**, takes care of monitoring the performance and fault management aspects (if any).

Service VNFs are not required for the fundamental operation of the CESCs and instead, are initialised as per requirements of the CESC tenants and their UEs, offering the benefits of Mobile Edge Computing. The Service EMS has therefore a comparatively simpler management role in the SESAME architecture. Apart from having an interface with the Service VNF, the Service EMS has interfaces with SLA monitoring module and with VNFM.

Full access to the Service EMS view is provided to SCNO EMS users who are able to create and delete Service Chain managed objects. Read-only access is provided to VSCNO EMS users so that they may reference a Service Chain when provisioning an SC-VNF.

[29] The SC-PNF will also be configured to perform period scans so this step can be omitted if immediate service is not required.

2.1.9.4.1 Managed object model and hierarchy

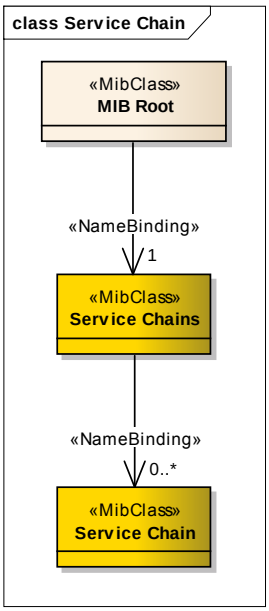


Figure 27: Service Chain Managed Object Hierarchy

2.1.9.4.1.1 Service Chains Object

This is a collection object beneath which child Service Chain objects may be created. A single instance of this object is created automatically by the EMS and it cannot be deleted.

2.1.9.4.1.2 Service Chain Object

Service Chain managed objects are created by a SCNO EMS user beneath the Service Chains collection object. There is almost no manageability associated with Service Chain objects. Their key purpose is to capture the identity of a set of resource templates in the Local Catalogue. These resources are used by the NFVO to instantiate an instance of the service chain and activate the necessary SDN functionality to connect it.

2.1.9.4.2 Service EMS Business Logic

Note: The following description is provisional and will be finalised as part of tasks 6.2 and 7.2.

A service chain is an ordered list of services, attached to an SC-VNF instance through which user plane traffic flows.

A service chain is first made available on the EMS by means of the following steps:

- *Its resource template(s) and a copy of the code images for each component service are installed in the local catalogue.*
- *An SCNO EMS user creates a new managed object representing the service chain. This managed object is available on a read-only basis to all EMS users and captures the key aspects of the service chain including the identity of its resource template(s) in the local catalogue.*

Once available in the local catalogue, a service chain is instantiated on a CESC when a VSCNO EMS user wishes to attach it to an SC-VNF. The overall process flow is as follows:

- *The service chain is attached to an SC-VNF either at provisioning time or subsequently as a configuration change.*
- *EMS business logic detects the configuration change and triggers instantiation of any new services by sending an instantiation instruction to the NFVO. In turn, the NFVO instructs the VIM resident on the target CESC to instantiate the service chain using the Lightweight Management protocol described in section 2.1.15.2. This instruction carries the identity of the resource template obtained from the Service Chain managed object.*
- *The VIM is then responsible for identifying a microserver within the CESC cluster with sufficient spare capacity and for instructing the SDN Controller to provide appropriate network connectivity.*

2.1.9.5 Cross-module EMS functionality

In addition to the discrete areas described previously, the EMS also provides a small amount of cross-module functionality in the areas of configuration, fault and performance management:

- According to the chosen functional split, it collects configuration data from each SC-VNF and associated SLA, such as the PLMN ID of the VSCNO, the list of neighbours for handout and the VSCNO's share of network bandwidth, and merges this data into the configuration of the SC-PNF.
- It receives alarm reports generated by the SC-PNF and forwards, as appropriate, copies of these to each SC-VNF for reporting to the tenant VSCNO.
- It receives performance management data generated by the SC-PNF and, if requested by configuration, forwards a PLMN specific copy to each VSCNO.

2.1.10 EMS Access Rights

2.1.10.1 Overview

The EMS uses a managed object containment based access rights system. The management hierarchy of the system is divided into a number of distinct sub-trees to which users of the EMS are assigned access. These access rights are not assigned on an individual basis. Instead, they are assigned to named groups and an individual's rights are determined by the set of groups to which they belong. In addition, EMS users may have one of three roles: Normal User, User Manager or Administrator.

2.1.10.2 EMS User Roles

Function / Operation	Normal User[1]	User Manager[2]	Administrator[3]
Can login to the EMS	✓	✓	✓
Can change their own password	✓	✓	✗[4]
Can assign "User Manager" role to any normal user	✗	✓	✓
Can reset the password of any other user except Administrator	✗	✓	✓
Can create, delete, enable and disable users except Administrator	✗	✓	✓
Can add components to User Groups and modify their access permissions	✗	✓	✓
Can have User Manager rights revoked	✗[5]	✓	✗

Table 28: User Roles and Associated Rights

Notes:

1. The majority of EMS users have this role. All VSCNOs will have this role with access to individual components defined by their group membership.
2. This role is effectively the same as the Normal User role with the addition of user management rights. Such roles are normally be reserved for employees of the SCNO.
3. The Administrator is a special user account that cannot be deleted and whose password cannot be changed by the normal EMS client.
4. The Administrator password cannot be changed via the EMS client and requires access to the EMS server console.
5. A Normal User does not have user management rights so they cannot be revoked.

2.1.10.3 EMS User Groups and Group Permissions

Each EMS User group has a unique name and a set of associated access rights. Individual EMS users belong to groups rather than groups having members. Access rights are "additive" such that a user belonging to multiple groups has access rights that are the sum of the rights obtained from each individual group.

A user group may have access rights to the following classes of managed object:

Object Class	Object or Sub-tree	Description
EMS Components	MIB Root EMS Server HW Platform Security GW NTP servers	These managed objects relate to the functioning of the EMS itself and the supporting services needed to provide an operational system. Typically, only EMS users that are employees of the SCNO are granted access to these objects.
Physical Cells (PNFs)	LTE APs	These collections of managed objects represent the physical small cells that host the virtual small cells. SCNO users have full access to these objects in order to provision, configure and manage them. Read-access *may* be provided to VSCNO users for information purposes only, but there is no requirement to do so.
Virtual Small Cell Network Operators (VSCNOs)	VSCNOs	The top-level VSCNOs object is created automatically by the system and cannot be deleted. SCNO users may create, configure and delete child VSCNO objects as a means enrolling, updating and de-enrolling virtual small cell network operators.
	VSCNO	SCNO users have full access to the VSCNO object representing an individual virtual small cell network operator. The VSCNO to which this object relates has read-only access to it, but is not allowed to make configuration changes. They are, however, allowed to create, delete and configure the child objects. VSCNOs have no access to the objects belonging to other VSNOs.
	Child objects	The child objects of the VSNO object represent the configuration of the VSCNO's network slice and comprise virtual small cells, MMEs and neighbour cells. A VSCNO has the necessary access rights to perform these operations themselves (with assistance from EMS business logic). A suitably privileged SCNO user may also perform these operations on behalf of the VSCNO.

Table 29: Managed Object Access Classes

2.1.10.4 Access Rights

The EMS provides the following classes of access rights that may be assigned to user groups. These rights are ordered such that each class of rights includes all of the capabilities of the previous class whilst adding new capabilities.

Access Right	Description
No Access	This is the default for certain classes of object and means that members of the group are unable to view the object or sub-tree.
Read Access	Members of the group may view the object but may not make configuration changes. They do not receive fault management reports relating to the object or sub-tree.
FM Access	This provides Read Access rights plus the ability to perform Fault Management operations. Members of a group with this right may view an object but may not make configuration changes. They may also view alarm reports relating to the object or sub-tree and acknowledge, comment or delete the alarm.
Operation Access	This provides the rights of FM Access **plus** the ability to administratively Lock, Unlock and Reinitialise (i.e. re-boot) an object to which a group member has this right.
Full Access	Members of the group may perform any legal CM, FM or PM operation on the object or sub-tree.

Table 30: Access Rights

2.1.11 Control Plane Protocol View

Figure 28 shows how the Control Plane is handled in the proposed CESC architecture for a UE.

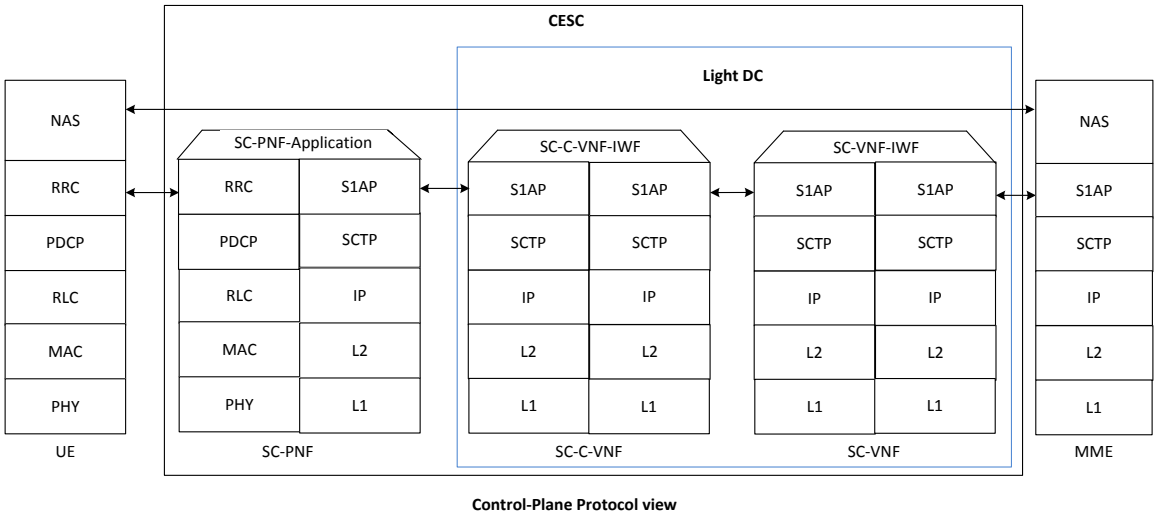


Figure 28: Control Plane Protocol Stack in the data path associated to one UE

Following the 4G LTE specifications, the mobile control procedures between the UE and the SC-PNF are managed by means of the RRC protocol over the radio interface. Likewise, the control operations with the EPC are implemented by the use of the S1AP protocol between the SC-PNF and the MME.

Following the specification of the SESAME PoC CESC provided in SESAME D2.3, the SC-PNF is able to broadcast different PLMN IDs (belonging to different tenant VSCNOs) and the control plane multi-tenancy support resides in the common VNF instance (SC-C-VNF). Therefore, each CESC needs to deploy one VNF instance of this type, which acts as the network selection function between the possible tenant VSCNOs.

The SC-PNF sends the S1AP traffic of all the connected UEs to the same IP address, and the S1AP traffic arrives to the SC-C-VNF. This VNF instance is configured with the specific association of UEs and VSCNOs, and therefore the SC-C-VNF is capable of forwarding the S1AP traffic to the IP address of the specific VSCNO SC-VNF.

In this way, the S1AP data arrives to the VNF instance (SC-VNF) specifically deployed at the CESC for the associated VSCNO. This VNF instance handles the S1AP negotiation with the VSCNO MME, being able to capture and modify the User Plane GTP TEID allocations for the UE.

From a deployment viewpoint, there is a different SC-VNF instance per VSCNO at each CESC. Each of these SC-VNFs terminates an S1AP connection to an MME. In this way, the VSCNO MME would see a collection of Small Cells in the same way as if they are physical Small Cells.

2.1.12 User Plane Protocol View

Figure 29 shows how the basic User Plane data path in the proposed CESC architecture for a UE.

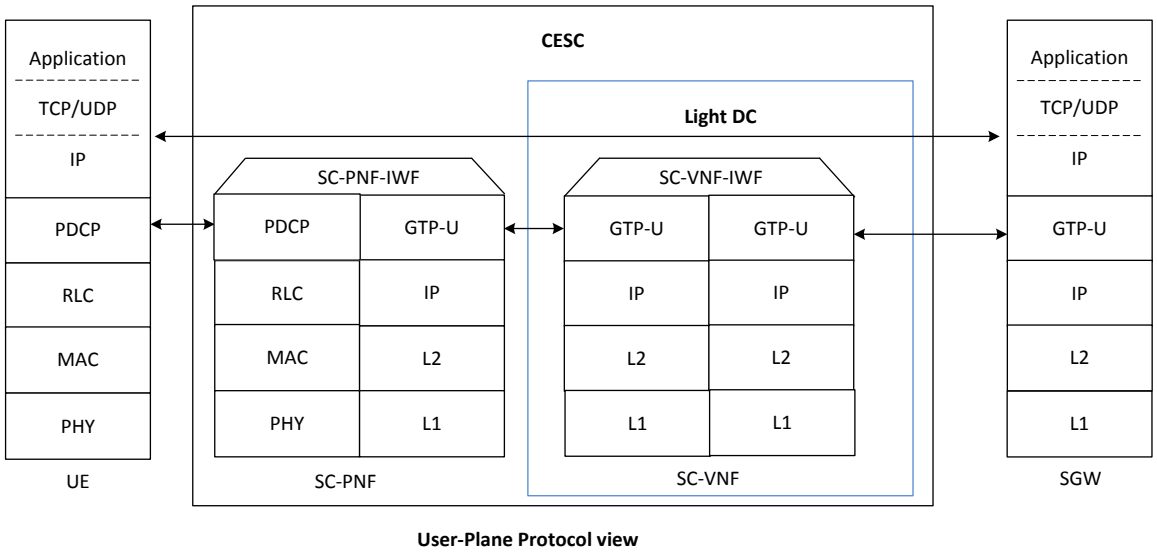


Figure 29: Basic User Plane Protocol Stack in the data path associated to one UE

From the Control Plane procedures, the Small Cell elements are able to learn from the MME the required forwarding information for the User Plane data of the specific UE, including the GTP TEID and the IP address of the assigned VSCNO SGW.

In the UL direction, the SC-PNF de-encapsulates the traffic incoming from the UE in PDCP over the radio interface and encapsulates the user data into the assigned GTP-U tunnel towards the respective SC-VNF associated with VSCNO. SC-VNF after relaying the user data through the service chain further encapsulates the user data into the assigned GTP-U tunnel towards the respective VSCNO SGW.

Since the SGW of the different VSCNOs has a different IP address, the User Plane traffic is forwarded to the specific SC-VNF instance deployed for each VSCNO at the CESC. The network selection function for the User Plane resides in the SC-VNF and the inclusion of the SC-C-VNF is not required.

The SC-VNF is then able to act as GTP-U tunnel endpoint to the SC-PNF and to the VSCNO SGW, which allows perform different processing operations over the user data in the RAN before forwarding the traffic to the EPC.

In the DL direction, the traffic coming from the VSCNO SGW is addressed through the associated GTP-U tunnel to the IP address of the SC-VNF, which is exposed to the VSCNO SGW by the SC-VNF instance associated to that VSCNO at the CESC. The DL user plane traffic arrives at the corresponding SC-VNF instance that may perform different processing operations over the IP packets before encapsulating them again into the GTP-U tunnel towards the SC-PNF.

As a result, the proposed User Plane data path will result as depicted in Figure 30.

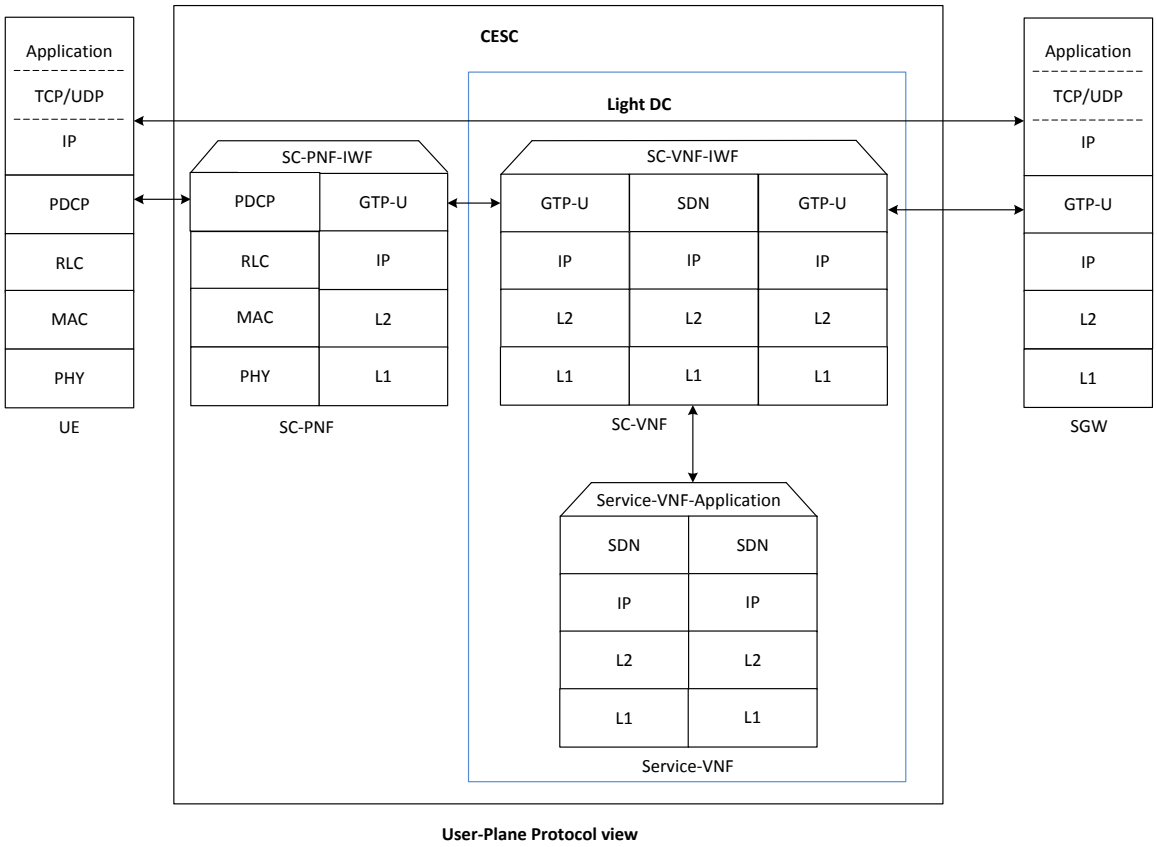


Figure 30: Basic User Plane Protocol Stack with edge computing capabilities in the data path associated to one UE

The interworking functions of the SC-VNF include now a new protocol stack associated to the connectivity to one of the Service VNFs associated to the specific VSCNO, which may run in the same or other CESC in the cluster.

Generally speaking, the whole user data packet (from the application layer to the IP level) is encapsulated in a new protocol stack compatible with the SDN paradigm adopted for the forwarding functions within the SESAME Light DC. Depending on the traffic steering policies provided by the SDN controller to the SDN-capable switch of the underlying infrastructure, the new protocol stack may require the inclusion the configuration of different header fields at different layers (e.g., destination IP addresses of different Service VNFs, IP-level type of service, etc.) or the inclusion of new header tags between different protocol layers (e.g. NSH). Alternatively, the SDN traffic steering policies may be applied over the original user IP packet based on specific header fields and, in this case, an additional IP header would not be required.

2.1.13 Optional HeNB GW

Note: The HeNB GW is not part of the PoC implementation. The description in this section demonstrates that the architecture described in this document supports an HeNB GW, should one be required.

In the case that the VSCNO requires the deployment of HeNB-GW functionality (HeNB-GW-VNF for the virtualised network element), all the Small Cells associated with the VSCNO appear as a unique Small Cell to the VSCNO MME.

In this case, the S1AP traffic of the different Small Cells is aggregated in a unique VNF instance for the whole CESC cluster, which could be deployed in one of the CESCs in the simplest case (or alternatively across different CESCs for resiliency features).

Therefore, the HeNB-GW-VNF also needs to tracks the associations between UEs, SC-PNFs and GTP-U information.

Figure 31 illustrates the basic connectivity and protocol stack related to two different UEs belonging to the same VSCNO but connected to two different SC-PNFs.

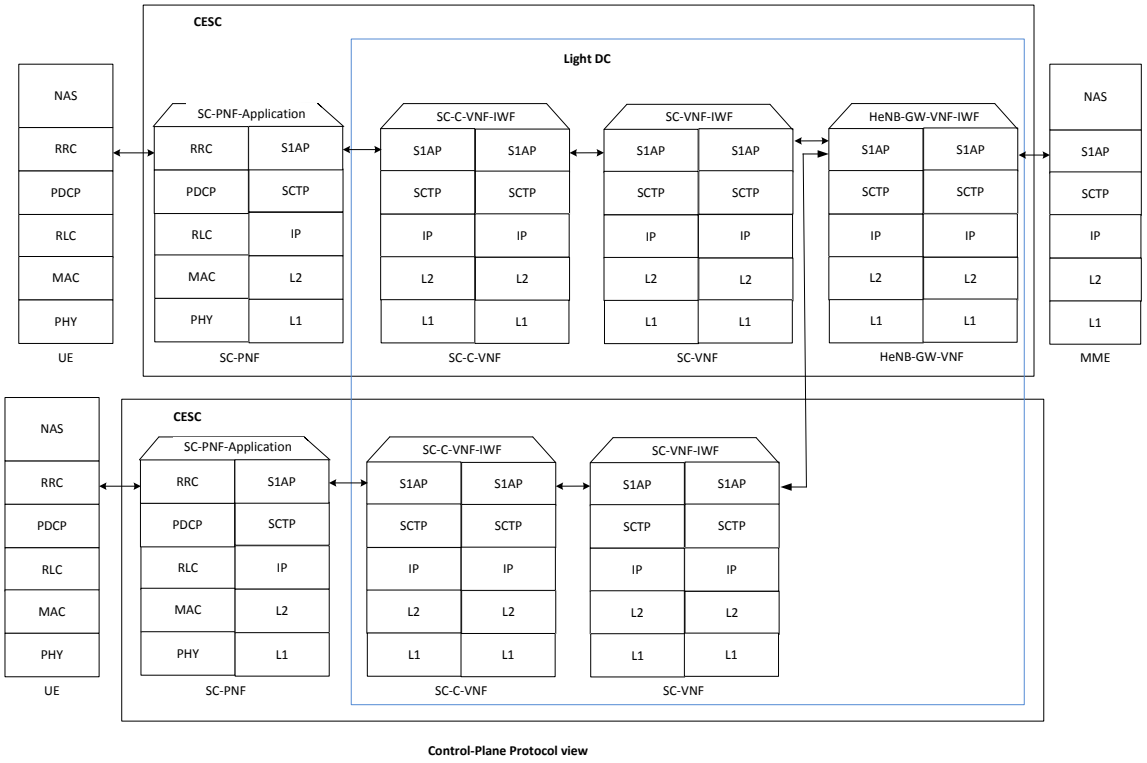


Figure 31: Use of virtualised HeNB-GW network element in the Control Plane and associated protocol stack

Concerning the User Plane, one approach would be for the HeNB-GW-VNF of a VSCNO to aggregate the respective user data of all the CESCs associated to that VSCNO and present a single GTP-U endpoint to the VSCNO SGW. This would alleviate the Control Plane load in the EPC and present the simplest User Plane interface, but this option would impose high traffic load on a single CESC.

3GPP allows for the HeNB-GW to only perform control plane aggregation (see [4], section 4.6.1), so to alleviate concerns about high loading on a single CESC, one simple option is to leave the User Plane running directly between each SC-VNF and the EPC of each VSCNO.

There is also an intermediate option: a HeNB-GW handling the User Plane is not restricted to presenting a single GTP-U endpoint to the VSCNO SGW. For each E-RAB establishment, the HeNB (or HeNB-GW) specifies the HeNB-side Transport Layer Address in the E-RAB Setup Response to the MME. So typically, the use of a HeNB-GW on the User Plane will reduce the number of GTP-U endpoints (transport layer addresses in use) but it does not have to be a single address. So the User Plane handling could be distributed across a number of CESCs. This could perhaps be advantageous if some CESCs had more processing capacity than others and would be nominated to perform the higher-capacity User Plane processing. Another place to perform aggregation might be the CESCs performing the Gateway role for a wireless backhaul solution, as the User Plane must already traverse these nodes to reach the Core Network.

If the optional HeNB-GW is included it will only perform C-Plane aggregation.

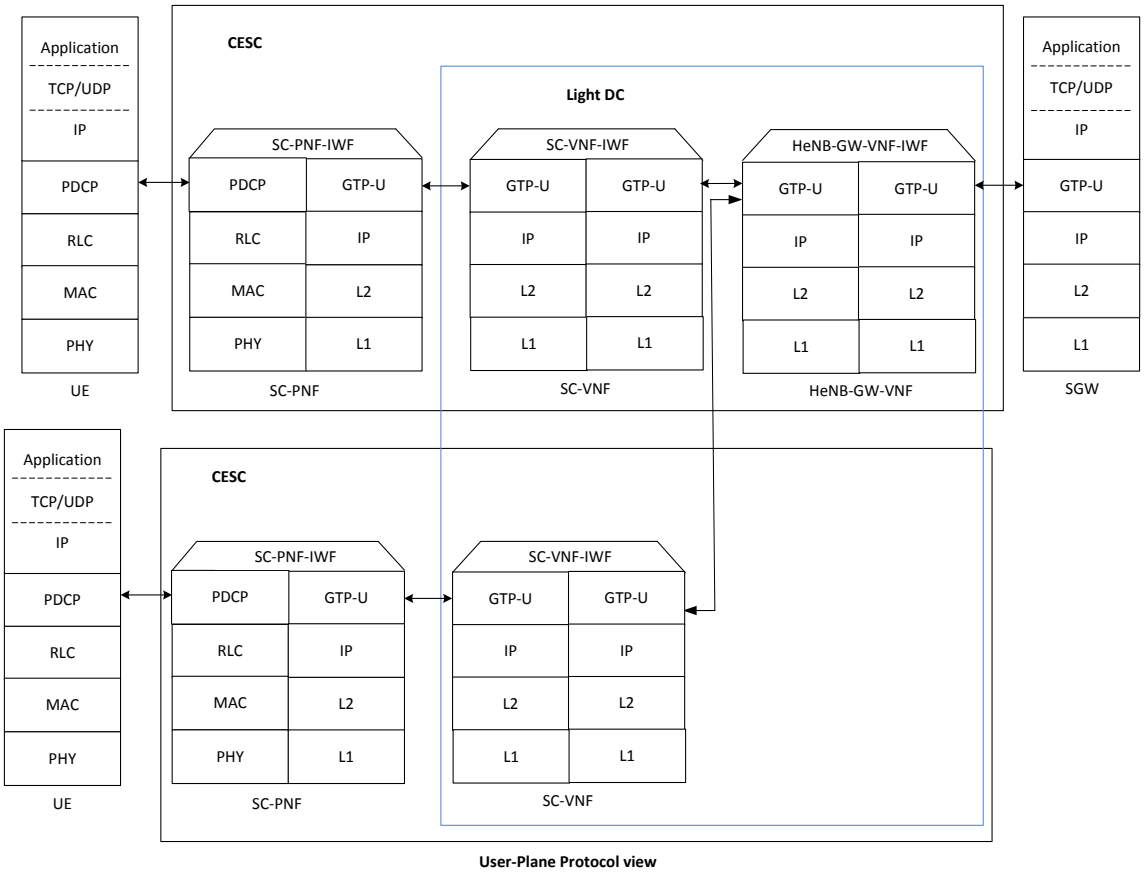


Figure 32: Use of virtualised HeNB-GW network element in the user plane and associated Protocol Stack

2.1.14 X2 Interface

Note: X2 functionality is not provided in the PoC implementation. The detail in this section simply illustrates that the current architecture supports X2 functionality.

The X2 interface is introduced in 4G LTE as a way to create direct connectivity between different neighbouring eNBs. The purpose of this interface is two-fold: to interchange load and interference information for SON features, and to alleviate the impact of horizontal handovers by forwarding downlink user plane data before the handover completes.

In the case of SESAME, the X2 interface applies to the connection between two CESCs and the interface could be implemented at SC-PNF level, at SCVNF level or split based on the different functionalities.

The X2-C interface uses the X2AP protocol for the communication between two Small Cells. Since most of this functionality is related to radio configuration and monitoring parameters, this interface needs to be implemented at the SC-PNF and the SC-C-VNF.

2.1.15 Management Plane Protocol View

Management in SESAME is achieved by one of three mechanisms depending upon the nature of the element being managed:

- Elements that are internal to the EMS make use solely of internal data structures and messages. The details of this mechanism are an implementation detail that is not addressed in this document.
- Physical Small Cells providing the SC-PNF function. These are managed using the industry standard TR-069 CPE WAN-LAN Management Protocol [79] and associated TR-196 Data Model [80].
- CESC Components running in the Light DC such as the SC-Common VNF, SC-VNF and Service VNFs. Each of these is relatively lightweight in nature and, thus, makes use of a correspondingly lightweight management plane.

The CM, FM and PM aspects of the management plane are as follows:

2.1.15.1 TR-069 and TR-196

2.1.15.1.1 Configuration and Fault Management

TR-069 [79] and the TR-196 [80] data model provide a combination of configuration and fault management. The TR-069 and TR-196 standards currently do not provide for performance management.

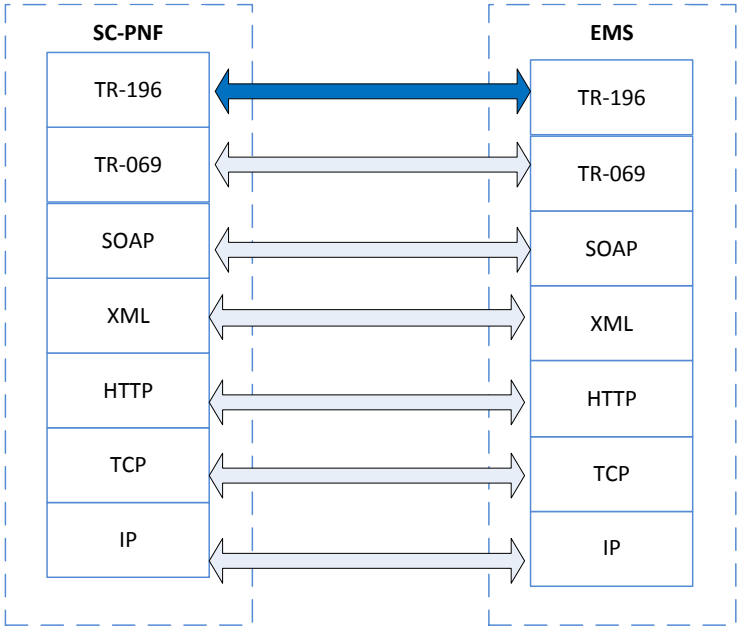


Figure 33: TR-069 Protocol Stack

2.1.15.1.2 Performance Management

Performance management reports generated by the physical small cell are XML files conforming to the format defined in 3GPP 32.435 [70]. These are uploaded periodically to the EMS using HTTP (or optionally HTTPS) according to a configured schedule. The EMS post-processes these reports into individual per-VSCNO reports, which, if configured, are then uploaded to the VSCNO's server, by using an identical mechanism.

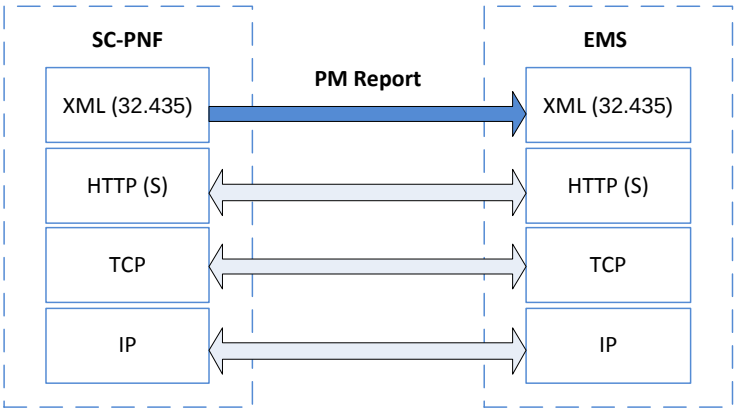


Figure 34: PM Report Protocol Stack

2.1.15.2 Lightweight Component Management

2.1.15.2.1 Configuration Management

The TR-069 protocol [79] is too complicated for the configuration and management of CESC components that require minimal configuration. Thus, for these "lightweight" CESC components a much simpler mechanism is required. This mechanism will be chosen during the implementation phase. Options that will be considered include but are not limited to:

- Transfer of a simple configuration file using a protocol such as HTTP, SCTP or FTP.
- SNMP. Any proprietary protocol that is lightweight with respect to its resource demands and easy to implement on the microserver platform.

2.1.15.2.2 Fault Management

"Lightweight" components report faults by using HTTP or HTTPS to upload an XML file of the form described in Appendix E to the configured EMS address.

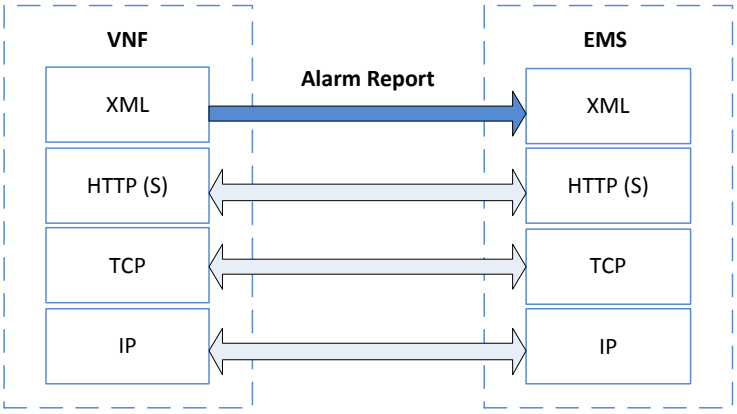


Figure 35: Lightweight Fault Management Protocol Stack

2.1.15.2.3 Performance management

Most "lightweight" components do not support performance management reporting. However if such a requirement arises, they will implement the standard XML file upload mechanism described in *section 2.1.15.1.2*.

2.1.16 Comparison with standard 3GPP MOCN

The following table details how SESAME differs from the standard Multi-Operator Core Network functionality defined by 3GPP:

Standard 3GPP (LTE) MOCN	**SESAME**
Provides for the interconnection of multiple core networks.	SESAME is identical to standard MOCN in these two respects as it uses standard interfaces between the UE and the RAN and between the RAN and the Core Network.
Allows each cell to serve up to a maximum of six PLMNs.	
Implied assumption that infrastructure owner configures each physical cell on behalf of the other PLMN owners participating in the scheme.	In contrast, SESAME provides each VSCNO with a view of their network slice. It allows a VSCNO to provision virtual small cells in a manner similar to how they would provision a physical cell. The assignment of virtual cells to physical cells is automatic allowing the VSCO to provision virtual cells without the need to involve the SCNO in each individual case.
Implied assumption that the network is shared on an equal basis amongst the participants. Standard 3GPP MOCN makes no provision for the ability to provide different PLMNs with different qualities of service.	In contrast, SESAME allows the definition of multiple SLAs allowing different amounts of network resources to be assigned to different virtual cells. These SLAs can operate on a combination of individual cell basis, across a set of cells is a defined area or on a temporal basis.
No consideration of management aspects or of how well the network is performing with respect to an individual participating operator.	In addition to being able to configure their network slice, SESAME provides VSCNOs with tailored fault and performance management features that enable them to verify that their network slice is operating correctly. The SLA Monitoring function determines whether or not each SLA is being met and generates alerts when an SLA breach is detected.
No equivalent in standard MOCN.	SESAME provides mobile edge services that allow VSCNO specific user plane processing by means of service chains. These provide lower backhaul and core network loading by performing user plane processing close to the user.

Table 31: SESAME – MOCN Comparison

2.2 Internal Interfaces

2.2.1 SC-PNF -> SC-C-VNF

For the PoC implementation, based on the S1 functional split, the control plane interface between the SC-PNF and SC-C-VNF is as per standard S1-MME interface [4].

For the S1 functional split, user plane traffic is exchanged directly between SC-PNF and SC-VNF, bypassing the SC-C-VNF.

2.2.2 SC-C-VNF -> SC-VNF

For the PoC implementation, based on the S1 functional split, the control plane interface between the SC-C-VNF and SC-VNF is as per standard S1-MME interface [4].

In additional to the standard S1AP interface and control procedures, there is a requirement for an additional pair of SCTP stream identifiers for SC-C-VNF and SC-VNF specific communication. These stream identifiers are SESAME specific and used by the SC-C-VNF and SC-VNF to exchange messages for front-haul bandwidth management. Going forward new SC-C-VNF and SC-VNF specific messages can be added as and when required.

Table 32 details the list of SC-C-VNF and SC-VNF specific communication messages:

Message Name	Direction	Usage
Status Indication	SC-VNF->SC-C-VNF	This message is sent from the SC-VNF to SC-C-VNF to indicate the status and resource configuration active at the SC-VNF.
Config Request	SC-C-VNF->SC-C-VNF	This message is sent from the SC-C-VNF to SC-VNF to configure the SC-VNF to generate periodic resource utilisation reports at the granularity specified in the message.
Config Response	SC-VNF->SC-C-VNF	This message is sent from the SC-VNF to the SC-C-VNF as a response to the "Config Request" message.
Set BW Request	SC-C-VNF->SC-VNF	This message is sent from the SC-C-VNF to the SC-VNF to configure the split of front-haul bandwidth allocated to SC-VNF.
BW Utilisation Report	SC-VNF->SC-C-VNF	This message is sent from the SC-VNF to the SC-C-VNF to report the current front-haul bandwidth utilized by SC-VNF.
Relieve User Congestion	SC-C-VNF->SC-C-VNF	This message is sent from the SC-C-VNF to the SC-VNF to request the release of user when the SC-C-VNF detects user congestion.
X2 Handover Request	SC-C-VNF->SC-C-VNF	This message is sent from the SC-C-VNF to the SC-VNF to allocate resources for a user during the X2 handover prepara-

Message Name	Direction	Usage
		tion phase.
X2 Handover Request Acknowledge	SC-VNF->SC-C-VNF	This message is sent from the SC-VNF to the SC-C-VNF as a response to the "X2 Handover Request" message.
UE Context Release	SC-C-VNF->SC-C-VNF	This message is sent from the SC-VNF to the SC-C-VNF to release the context associated with a UE in SC-VNF

Table 32: SC-C-VNF and SC-VNF specific interface messages

2.2.3 SC-VNF ->SC-PNF

The user plane interface between SC-VNF and SC-PNF is as per the standard S1-U interface defined in [4].

2.2.4 SC-VNF Management

The management interface between the SC-EMS and SC-VNF is provided by the lightweight management plane described in *section 2.1.15.2*.

2.2.5 SC-C-VNF Management

The management interface between the SC-EMS and SC-C-VNF is provided by the lightweight management plane described in *section 2.1.15.2*.

2.2.6 SC-PNF Management

The management interface between the SC-EMS and SC-PNF is provided by TR-069 management plane described *in section 2.1.15.1*.

2.2.7 SC-VNF to Service VNF(s)

The Service Chain is implemented by a Software Defined Networking (SDN) Forwarding Graph. The details of this process are part of Work Packages 5 and 6 and are not addressed in detail in this document.

2.2.8 Service Chain Provisioning

The bulk of service chain provisioning is performed outside of the EMS and is not addressed in this document. Once a service chain has been made available in the Local Catalogue the SCNO makes it available VSCNOs by creating a Service Chain managed object. The key property of this object is an identifier for the service chain in the Local Catalogue that the EMS provides to the NFVO as part of the provisioning process.

The full details of this scheme will be finalised as part of WP 6.2.

2.3 Wireless Backhauling

A crucial component of the SESAME architecture is the backhauling. The backhauling infrastructure enables communications between the CESCs and the core network, but also interconnects CESCs with each other and with managing system components, i.e. the CESCM. Both signalling traffic, to control and monitor the PNFs and VNFs of CESCs, as well as the access traffic exchanged between UEs and the core network, are carried over the backhaul. Further, the microservers forming the Light DC are interconnected via the backhaul. In SESAME, this is necessary to offer features like SFC, which enables chaining of different edge services running on separate machines.

An important candidate for 5G backhaul technologies are wireless radio communications. In contrast to traditional wired backhauls relying on technologies like Ethernet over copper or fibre optics, wireless backhauls do not require the deployment of additional infrastructure. Instead, each CESC is equipped with at least one wireless radio transceiver dedicated to backhauling.

In the following subsections, we present the main concept of how wireless backhauling is applied in SESAME in the context of backhaul network virtualisation and Self-X backhauling features. Before that, we give a short introduction into which are the wireless radio backhaul technologies that are considered for SESAME deployments.

2.3.1 Wireless backhaul technologies

In 5G deployments, the use of a large variety of wireless technologies is envisaged. One of the main paradigms of 5G is to include above-6 GHz technologies in the wireless access spectrum, which also can be used for backhauling.

Among the wireless technologies, radio communications in the 60 GHz band are predestined for the use in 5G deployments: they offer a high bandwidth and short to medium communication range, which makes them well suited for small cell deployments as they are targeted in SESAME. As an alternative to 60 GHz communications, sub 6 GHz technologies can be integrated in 5G deployments. More specific, wireless devices supporting the well-known IEEE 802.11 standard[30] can be used to implement the backhaul.

In 5G networks, the access traffic is expected to show significant changes over time in terms of quantity and quality, depending on the number of users and the type of services that they request, requiring a high degree of flexibility from the backhaul. In the following subsection, the investigation framework used to perform evaluations of wireless backhauling solutions, which satisfy these requirements and take into account SESAME-*specific* concepts like multi-tenancy, is presented.

2.3.2 Backhauling framework

Classic wired backhauling solutions often make use of star-topologies, where a central switch, common to all network devices, is used to interconnect the network devices (in SESAME that would be the CESCs) with each other and to connect to the external networks, such as the Internet. This network topology is often chosen due to its simplicity and because it is capable of satisfying the needs of simple point-to-point (P2P) connectivity. In wireless backhauls, *however*, we are able to implement different sorts of topologies without any further costs: a device equipped with wireless transceivers may be able to communicate not only with one, but with several other devices over a wireless link. By enabling each CESC in the network to act as a relay (router) node, a mesh network can be formed.

[30] For more details, see: http://standards.ieee.org/about/get/802/802.11.html

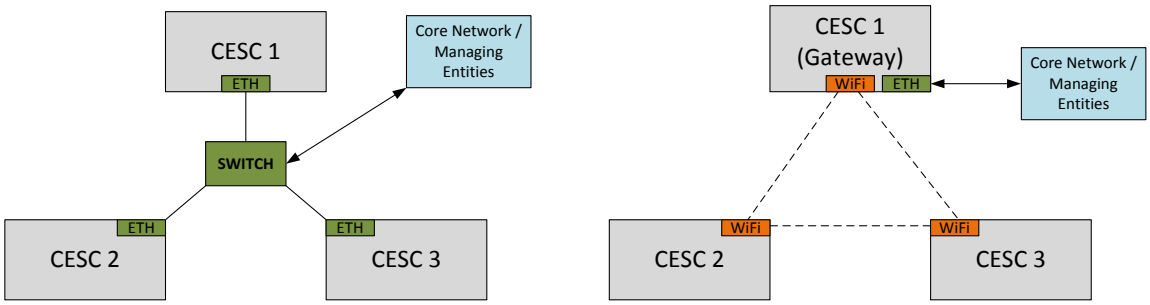


Figure 36: Wired Star Topology and Wireless Mesh Topology Comparison

In addition to the CESCs acting as relay devices in the mesh network, one or several CESCs need to act as egress points (gateway nodes) for the wireless backhaul. As depicted in Figure 36, they are responsible for enabling communications between the cluster of CESCs and the core network or any SESAME managing entities of the CESCM, like the EMS or the VNFM. Gateways are an essential part of the backhaul, because they are the only devices of the mesh that have access to the management/Internet backbone.

An implication of mesh topologies is that routing becomes necessary to interconnect CESCs with each other or with a gateway. Further, to compose routes with high stability and link qualities, it is necessary to know about the state of the wireless backhaul resources. This includes link qualities and availability of links, since they can change over time. In SESAME, we use SDN technologies, which provide solutions for both of these requirements.

As such, an SDN controller is dedicated to manage the backhauling network, it is responsible for keeping track of the network topology and for performing centralised routing decisions in the data and control plane. Further, in order to apply SON features as they are applied at the access radio level, the SDN controller is responsible for collecting and evaluating status information of the network (link qualities, status of wireless interfaces, ongoing traffic). The gathered information is used to enhance the routing algorithms of the SDN controller, by aggregating link state information and information about existent traffic flows in order to assign high performance routes for access and control traffic.

There are two options for the placement of the SDN controller: it runs either internally in the Light DC or on an external machine that manages the virtualized backhauls of all tenants. It is even possible that one particular SDN controller is assigned to each tenant, a decision on the design of the backhaul network that will be taken as its investigations progress over the course of the project.

As in wired networks that rely on SDN technology, applying SDN requires a certain degree of abstraction of the network: the wireless interfaces used by every CESC are represented as virtualized ports of a wireless virtual switch, the wireless radio connections between two network nodes are abstracted to be treated like traditional, wired links. The virtualization of the network nodes and the wireless links allow the SDN controller to perform network slicing, where the wireless backhaul resources are shared and assigned on a per-tenant basis. This aspect of the wireless backhaul is discussed in the next subsection.

2.3.3 Per-tenant virtualization of the backhaul

Each tenant (VSCNO) signs an SLA that determines the type and quality of service they will obtain in the SESAME deployment. In an SLA, there is an agreement about which types and qualities of service each tenant will obtain from a SESAME deployment. Mainly, this translates into

the assignment of radio (access) resources and a minimum service performance to each tenant, according to the signed SLAs.

The SLAs have also an impact on the wireless backhaul, which needs to be taken into account. Since wireless backhauling resources are limited in terms of bandwidth and minimum per-hop delays, it is necessary to assign backhaul resources efficiently to each tenant, so that the KPIs from the SLAs can be met. The SDN controller uses an API to exchange information about the active SLAs and other relevant data with the CESCM. Based on the overall view of the network and the knowledge about SLAs, the SDN controller performs network slicing.

In this slicing, the available backhaul resources are distributed among the tenants to assure that each tenant's performance requirements are met. On one hand, the SDN controller virtualises the available (physical) topology, which includes the wireless links, the relays and the gateways of the backhaul network, as shown in the example topology (Figure 37).

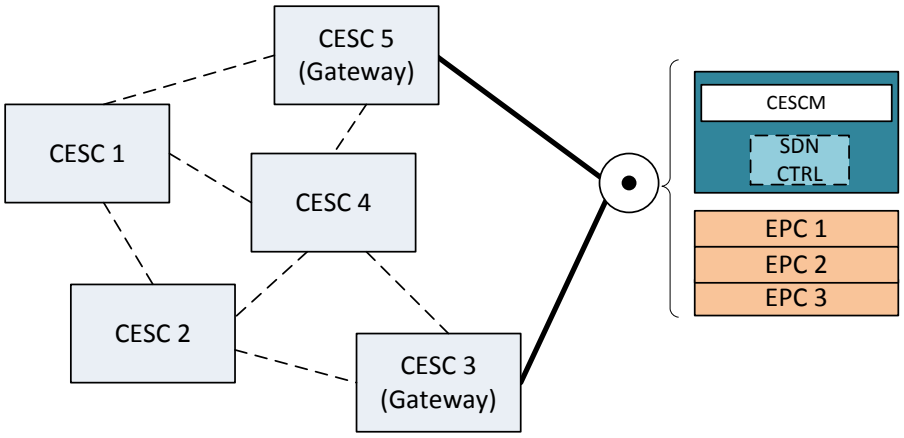


Figure 37: Example topology for a SESAME deployment with several CESCs that act as relay nodes or gateways, respectively

As a result, a subset of all the available virtualized networks elements is assigned to every tenant. Thus, every tenant may have a different share of the backhaul network, yet being agnostic about it. An example of such differing per-tenant virtualisations applied to the example topology is shown in Fig.38. Each of the tenants is assigned a different set of wireless links (except the link marked in red) and a different gateway to reach the tenant's EPC. In spite of the obvious differences in the topology, all CESCs are included in each topology and each topology includes a gateway, required for the connection between the CESC cluster and the tenant's correspondent EPCs.

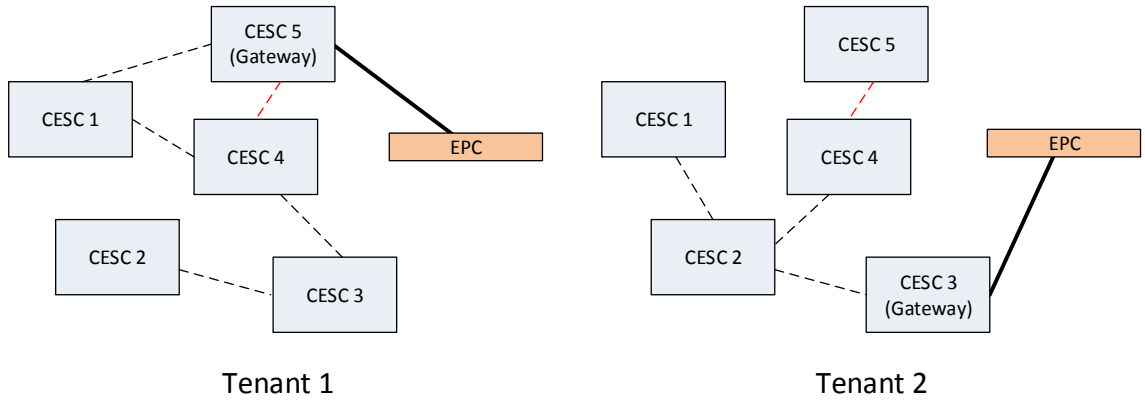


Figure 38: Possible Per-Tenant virtualisation of the Example Topology

On the other hand, the wireless radio resources are virtualized at the physical (radio) level. This includes the virtualization of the bandwidth of wireless links in terms of available data rates and the virtualization of frequency spectrum of the links. The SDN controller monitors the state of these physical resources and performs slicing to assign them to different tenants. For example, a wireless link may offer a transmission ratio of up to 1 Gbit/s. If the link is used by multiple tenants, the available data rate needs to be shared among them in such a way that the requirements of ongoing transmissions (e.g., video streams) and also the SLAs signed with each tenant can be met. To achieve this, the SDN controller uses the knowledge about the state of the wireless backhaul resources. This knowledge is not only necessary for the on-the-fly assignment of wireless resources, in a more general view, but it is also required to apply Self-X features, as discussed in the following subsection.

2.3.4 Self-X backhauling features

In order to achieve a high and stable performance in the wireless backhaul, a series of Self-X features can be applied. Whereas at the access radio SON features are applied by the NMS or EMS, as detailed later on in *Section 3*, in the wireless backhaul the SDN controller is the responsible for applying self-planning, self-optimization and self-healing. In the backhaul, centralized and decentralized Self-X approaches as introduced by cSON and dSON solutions[31], as well as hybrid solutions are possible. The centralized part of the self-x functions is executed in the SDN controller. The decentralized aspects of SON are executed in the CESCs. In the following, a set of basic Self-X features applying to the wireless backhaul are discussed.

2.3.4.1 Self-Planning

Aware of the available wireless interfaces and the links of the mesh network formed by the CESCSs, the SDN controller can assign wireless channels to the interfaces to assure an efficient spectrum sharing and reuse of the backhaul during the planning phase. In the same process, each active tenant is assigned a share (slice) of the backhaul network, as detailed in *Section 2.4.3*. However, due to the high dynamicity of the traffic occasioned by fluctuations in the number of users and varying traffic patterns, it is likely that the slice for a tenant is modified during network operation, which is part of the self-optimization discussed below. The self-planning applied in the wireless backhaul is mainly performed by the SDN controller that has a general network overview. Thus, we can put declare the self-planning to be a cSON feature.

2.3.4.2 Self-Optimization

During runtime of the SESAME deployment, the SDN controller gathers detailed information about the state of the backhaul links and the ongoing control and data transmissions. As a result, the controller is capable of sharing the available wireless backhaul resources among the tenants so that the requirements (SLAs) can be satisfied.

The actions of the SDN controller not only limit to assigning different shares of the network topology and the physical bandwidth (in terms of wireless link data rates) to the tenants, but also on rerouting traffic throughout the network in such a way that other network policies, like congestion avoidance or energy saving, can be satisfied. This cSON approach is similar to the one followed in MLB algorithms for access traffic, yet applied to the wireless backhaul resources.

2.3.4.3 Self-Healing

During network operation, it may be necessary to apply self-healing features in the wireless backhaul. While wireless links are stable in general, under certain meteorological conditions or upon obstruction of the line of sight between wireless transceivers, one or several links may

[31] For more details see, *for example*: https://en.wikipedia.org/wiki/Self-organizing_network

disappear (temporarily). Another incident could be the physical failure of a wireless interface, which would cause the loss of all wireless links associated to the interface.

In any of these cases, the controller needs to be able to react by redirecting traffic over alternative routes, taking into account how the new assignment of wireless backhaul resources will affect the overall network performance and whether the new state would be acceptable in terms of meeting SLA requirements. For this purpose, algorithms are designed to assure robustness of the network by, for example, calculating backup paths that are as disjoint as possible from main data paths in order to be able to provide alternative routes in case of failures. Self-Healing can be applied both as cSON and dSON features. In a cSON approach, failures detection and correction are responsibility of the SDN controller, whereas in a dSON approach CESCs can react to failures on their own. Such features can include fast-rerouting without waiting for new instructions from the SDN controller.

2.4 Northbound Traffic Interface to EPC

In a MOCN configuration, core network operators A, B, and C might share the LTE radio access network. Each operator provisions standard S1 interfaces towards each CESC, optionally using S1-Flex so that multiple MMEs are used to provide load balancing and resilience for signalling. The operators may also spread user plane load across multiple SGWs. A core network operator is identified by a PLMN ID (MCC + MNC[32]), such that the traffic is routed to the correct EPC via the PLMN ID. The available core network operators in the shared radio are broadcast to the UE in the system information (SI) at E-UTRAN. For UE-initiated requests, the traffic routing works as follows:

1. The UE reads the PLMNs from broadcast SI and performs network selection among available PLMNs.
2. The UE indicates (in the RRC connection request) the chosen CN operator and sends an TA Update request to the network (SC-PNF).
3. The SC-Common VNF relays the request towards the SC-VNF associated with the PLMN-ID.
4. The SC-VNF performs NAS Node Selection Function (NNSF) to determine which MME to send the request to; and sends TA Update request to the MME.

The MME determines whether the UE is allowed to attach to the network and sends the accept/reject message back to the UE. In the case of an TA Update accept message, the MME assigns the UE with a GUTI which can be used by for routing further transactions of the subscriber to the same CN.

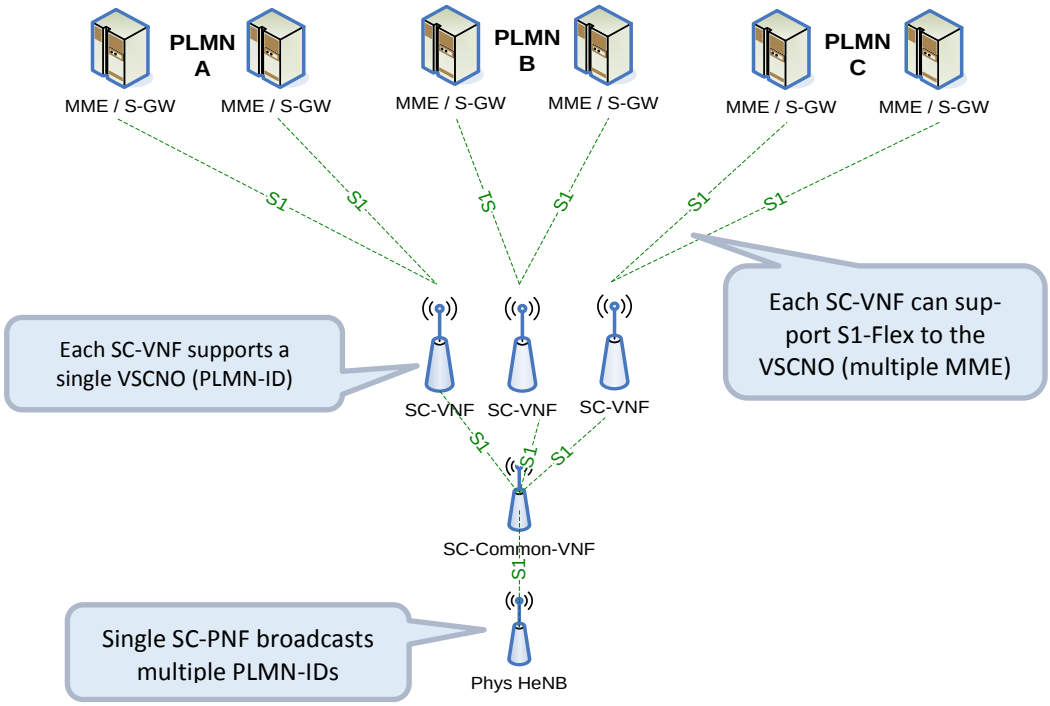


Figure 39: Northbound Traffic Interface for multiple PLMNs

For details of the control plane and user plane protocol stacks, refer to Figure 28 and Figure 29 in *sections 2.1.11* and *2.1.12,* respectively.

[32] See, for example: https://en.wikipedia.org/wiki/Mobile_country_code

2.5 Southbound Traffic Interface to UEs

The southbound traffic interface to UEs is the 3GPP standard LTE Uu interface. This a fundamental constraint of the SESAME architecture which must operate with the very large number of deployed handsets.

2.6 Northbound Management Interfaces

2.6.1 Configuration Management

As described in Deliverable 2.4, the northbound configuration management interface is based on the SOAP Solution Set of the 3GPP Configuration Management Integration Reference Point (IRP). It uses the following protocol stack:

Protocol component	Version / Notes
SOAP	SOAP version 1.1 as defined by: http://www.w3.org/TR/soap/
XPATH	XML Path Language (XPath) Version 1.0 http://www.w3.org/TR/1999/REC-xpath-19991116
XML	XML version 1.0
HTTP	HTTP version 1.1 exposed on port 8080 of the client interface of the SESAME EMS.
WSDL	WSDL version 1.1. http://www.w3.org/TR/2001/NOTE-wsdl-20010315 The services available are described by WSDL documents on the following addresses:
GenericIRP	http://<EMS-HOST>:8080/3gpp/GenericIRP?wsdl
NotificationIRP	http://<EMS-HOST>:8080/3gpp/NotificationIRP?wsdl
KernelCMIRPEndpoint	http://<EMS-HOST>:8080/3gpp/KernelCMIRP?wsdl
BasicCMIRPEndPoint	http://<EMS-HOST>:8080/3gpp/BasicCMIRP?wsdl
SESAMEEndpoint	See Appendix B / Section 8.2 of the present deliverable

Table 33: Protocol stack used for the configuration management

Details of the supported IRPs and methods are provided in Appendix B.

2.6.2 Fault Management

As described in Deliverable 2.4, the northbound fault management interface is based on the SOAP Solution Set of the 3GPP Fault Management Integration Reference Point (IRP) described in [78]. It meets the Fault Management requirements set out in [73] and allows a northbound system, operated for example by a VSCNO, to retrieve both current and historical alarm information.

Current alarms are the list of “active” alarms; those that have not been cleared and which are active until the fault that caused the alarm is corrected and a "clear alarm" is generated.

Historical alarms are those alarm events that occurred in the past, regardless of whether or not a "clear alarm" was generated.

Each VSCNO only receives only alarms that are relevant to them. Specifically:

- Alarms raised on any of the managed objects in their or managed object sub-tree. For example an SC-VNF instance that is unable to establish an S1 connection to an MME will raise an appropriate Link Failure alarm.
- Alarms raised on any SC-PNF that hosts an SC-VNF belonging to the VSCNO.
- The VSCNO is only able to perform alarm management operations, such as a manual clear, on alarms raised on their own sub-tree. Alarms raised on common SC-PNF objects are effectively read-only.

The Fault Management IRP uses a similar protocol stack to the Configuration Management IRP described in *section 2.6.1* above. The EMS exposes the corresponding WSDL document (see [78]) on the following address:

http://<EMS-HOST>:8080/3gpp/AlarmIRP?wsdl

Details of the supported FM IRPs and methods are provided in Appendix C.

2.6.3 Performance Management

As discussed in *section 2.1.15.1.2*, Performance management reports generated by the physical small cell are XML files conforming to the format specified by 3GPP 32.435 [70]. These are uploaded periodically to the EMS and post-processed into individual per-VSCNO reports, which are then made available via the northbound PM interface.

The SESAME prototype implementation meets a sub-set of the PM IRP requirements set out in 3GPP 32.411 [67]:

- It allows users to query and retrieve PM reports using the File Transfer IRP defined in 3GPP 32.342 [69].
- It **does not** support the creation, scheduling and management of measurement jobs. PM reports are generated at a fixed rate controlled by the configuration of the SC-PNF and there is no concept of a measurement job.

A northbound client of the PM interface (typically a VSCNO) uses the *listAvailableFiles* IRP (see Appendix B) to determine the set of files available on the EMS that were received within a specified time window. The returned *fileInfoList* result details the available files and, for each file, a location on the EMS from which the file can be obtained using either FTP (deprecated) or SFTP (preferred).

The contents of the returned *fileInfoList* are filtered according to the credentials of the requesting user such that a VSCNO may only obtain PM reports that relate to their individual network slice.

The File Transfer IRP uses the SOAP solution set described in 3GPP 32.346 [71] and, as such, uses a similar protocol stack to the Configuration Management IRP described in *section 2.6.1* above. The EMS exposes the corresponding WSDL document on the following address:

http://<EMS-HOST>:8080/3gpp/FTRPSystem?wsdl

Details of the supported File Transfer IRP methods are provided in Appendix D.

2.6.4 Northbound Interface Authentication and Security

The northbound interface is an abstraction layer to the internal modules of the NFVO for the requests coming from the CESCM Northbound interface. It has been explained in WP4; Northbound interface is popularly explained in the SDN terminology to provide an abstraction layer that serves an entry point to the internal NFVO components and it is placed in the upper part of the NFVO architecture diagram. The interface connects SDN applications to the controller. An application can request information, such as statistics and incoming connections from the controller. It can also send commands to the controller, in order to control the network, such as adding or removal of flow rules.

In SESAME project, in order to prevent security issues in the Northbound interface, every call to the NFVO's interface need to be secured and done by authorised users and modules. Although some security mechanisms can be proposed in the initial phase of the project development, a thorough threat analysis can reveal more necessary security amendments. SESAME will identify potential threats that facing northbound interfaces by using STRIDE threat modelling [44]. Table 34 illustrates STRIDE categories and security property for each category.

STRIDE threat categories	Threat category Security property
Spoofing	Authentication
Tampering	Integrity
Repudiation	Non-repudiation
Information Disclosure	Confidentiality
Denial of Service	Availability
Elevation of Privilege	Authorization

Table 34: STRIDE threat categories

There are number of threat vectors related to the SDN Northbound interface [45], which are categorised into two, as follows: (i) Attacks on and vulnerabilities in controllers (including applications) and; (ii) Lack of mechanisms to ensure trust between the controller and management applications. The first category is further subdivided into multiple vulnerabilities related to the Northbound interface. These vulnerabilities are as follows:

1. Service Chain Interference: When a message is forwarded from application to application, a malicious application may not send the message on to the next application, causing that message to be lost. This results in Control Message Drop or Infinite Loops.
2. Control Message Abuse: A malicious application can arbitrarily insert new flows in the switches' flow table. This results in Flow Rule Modification or Flow Table Clearance.
3. Northbound API Abuse: A malicious application may request the controller to disconnect other applications. This results in Event Listener Un-subscription or Application Eviction.
4. Resource Exhaustion: A malicious application can continuously send requests to the controller, consuming all its resources. This results in Memory Exhaustion or CPU Exhaustion.

STRIDE threat model

The first step in STRIDE method is to create a data flow diagram and the second step is to identify threats category using STRIDE method, the categories are explained below:

Spoofing: Illegally accessing and using another user's authentication information, such as username and password. Only authenticated actors should be able to use the northbound interface. Authentication credentials should not be guessed, listened or obtained.

Tampering: Data tampering involves the malicious modification of data. Data sent over the northbound interface should not be tampered with, as such high level encryption mechanism should be used.

Repudiation: Repudiation threats are associated with users who deny performing an action without other parties having any way to prove otherwise. When there is no secure log of commands sent over the northbound interface, it is relatively easy for a malicious actor to perform actions anonymously. A Nonrepudiation mechanism should be put in place in all the northbound interfaces.

Information disclosure: The northbound interface handles information about the network state and its configuration. While not extremely sensitive, this data can be valuable to some parties. Therefore, this information should not be disclosed to unintended parties. Encrypting network traffic and requiring authentication will aid in preventing information from being disclosed.

Denial of Service (DoS): The northbound interface is important for the SDN infrastructure. When it is unavailable, applications cannot do their work. Denial of service is a threat to this interface. A malicious user could either send a large amount of traffic to the northbound interface, or he could send resource-intensive requests to the controller, both resulting in the northbound interface becoming unavailable. Solutions for this include making the interface accessible only from a trusted network, or using other traditional DDoS mitigation techniques.

Elevation of Privilege: Applications should only have the least amount of privileges needed for their operation. For example, a monitoring application should not have the right to write to the controller, only reading will suffice. In addition, there will be very few applications which need the right to alter the device configuration. The controller should enforce some kind of access control on its API functions to prevent an application from accessing too much of the API.

The prototype implementation makes use of standard SOAP *username* and *password* credentials to identify the client originating an IRP request and to filter the results accordingly.

The use of Web Services Security (see http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wss) to digitally sign or encrypt SOAP messages in relation to the STRIDE threat categories described above is FFS.

2.7 Outstanding Design Question and Design Issues

The following outstanding issues need to be addressed before implementation of the prototype is complete.

2.7.1 VNF to VM / Core Allocation

The ARM processor planned for the Light DC has eight cores total. If each VNF runs in its own Virtual Machine then this number is too small. For example, if there needs to be one VM for the OS, one for the SC-Common VNF and one for each of the six SC-VNFs. This has already used all eight cores. One suggestion is to use containment. Another is that the SC-VNF is multiplexed. These options require assessment.

2.7.2 Traffic Model

Without a traffic model, it is not possible to accurately estimate the CPU and Networking requirements of VNFs for inclusion in resource templates. Such a model needs to cover aspects such as:

Control Plane (CP)

- How many UE registrations there are in the busy hour.
- How many E-RABs are set up.
- How many paging messages are received.

User Plane (UP)

- How much traffic per E-RAB is generated.
- What processing is performed by each VNF in the service chain.

The traffic model can be broken into 2 primary parts:

- The C-Plane (Signalling) activity necessary for mobility management of UEs and the establishment/clearing of calls (E-RABs).
- The U-Plane handling: routing of the E-RAB GTP-U packets between Core and UE; and the User Plane Service Chaining functions within the Light DC.

Each of these can be broken into component parts; and separated per-tenant.

2.7.2.1 Signalling

It is proposed that the primary measure of Signalling Load (or conversely the Signalling Capacity) is described in terms of Weighted Transactions Per Second (TPS). Each UE-*related* high-level procedure available on the S1 interface is assigned a transaction "weight" representing the relative processing load of the procedure. As a first estimation, the weight is deemed equivalent to the number of messages.

S1 procedure	**Transaction Weight**
Initial Attach	11
UE/MME Initiated detach	3
Active to Idle	6
Network-initiated Idle to Active (paging)	1

UE-initiated Service request	7
S1-based Inter eNB Handout	5
S1-based Inter eNB HandIn	4
TAU Request (including Auth)	6
Bearer Setup	2

Table 35: Example Procedure Weights

As the implementation is proven, the CPU load of these procedures can be measured or estimated; and the signalling capacity of a component declared as a number of Weighted TPS assuming an assigned vCPU capacity.

As the real-world signalling patterns become better understood (the number and/or relative mix of the S1 procedures) it will be possible to estimate TPS requirements in terms of number of users (and perhaps types of users); and thus, the overall signalling requirements of the CESC per user, and per tenant, can be estimated for capacity planning and resource sharing.

2.7.2.2 User Plane

As a broad estimation, the generic processing load required to forward the GTP-U packets associated with an E-RAB is a function of bandwidth and packet size. The GTP-U forwarding capacity of a component (e.g. SC-VNF) can be expressed in terms of an overall bandwidth (Mbps) assuming a specified mix of packet sizes (and IMIX[33]); and an assigned vCPU capacity.

For a generic IPv4-based traffic model, a typical "IMIX" specifies a ratio for small, medium and large frames appropriate to the technology.

Packet category	**Size (B)**	**Ratio**
Small	64	**7**
Medium	576	4
Large	1500	1

Table 36: Typical IMIX

This approach can be used as a starting point for describing capacity of CESC components:

Throughput (Mbps) assuming a stated IMIX (sizes & ratio) for an assigned vCPU capacity.

As real-world traffic models are observed, the IMIX can be refined as needed (packet size and ratios).

The specific processing requirements of the Service Chain elements will be specified as part of task 6.2. If the services have predefined characteristics in terms of packet sizes, this could be fed back into the overall IMIX model for generic forwarding components.

[33] For more information about IMIX (Internet Mix) see, for example: https://en.wikipedia.org/wiki/Internet_Mix

3 Initial studies on Self-X Functions

Self-Organising Networks (SON) refers to a set of features and capabilities for automating the operation of a network so that operating costs can be reduced and human errors can be minimised [1]. With the introduction of SON features, classical manual planning, deployment, optimization and maintenance activities of the network can be replaced -and/or supported- by more autonomous and automated processes, thus making network operations simpler and faster. SON functions, denoted in the context of SESAME as "Self-x" functions, are organised around the following main categories, which are based on previous references (such as [1]and [4]):

- ***Self-planning:*** Automatization of the process of deciding the need to roll out new network nodes in specific areas, identifying the adequate configurations and settings of these nodes, as well as proposing capacity extensions for already deployed nodes (e.g. by increasing channel bandwidths and/or adding new component carriers). Specific functions belonging to this category include the planning of a new cell and the spectrum planning.

- ***Self-optimization:*** Once the network is in operational state, the self-optimization includes the set of processes intended to improve -or maintain- the network performance in terms of coverage, capacity and service quality by tuning the different network settings. Examples of functions include Mobility Load Balancing (MLB), Mobility Robustness Optimisation (MRO), Automated Neighbour Relation (ANR), Coverage and Capacity Optimization, optimization of admission control, optimization of packet scheduling, inter-cell interference coordination and energy saving.

- ***Self-healing:*** Automation of the processes related to fault management (i.e., fault detection, diagnosis, compensation and correction), usually associated to hardware and/or software problems, in order to keep the network operational, while awaiting a more permanent solution to fix it and/or prevent disruptive problems from arising. Examples of self-healing functions include Cell Outage Detection and Cell Outage Compensation.

Self-x functions might automatically tune global operational settings of a small cell (e.g., maximum transmit power, channel bandwidth, electrical antenna tilt) as well as specific parameters corresponding to Radio Resource Management (RRM) functions (e.g., admission control threshold, handover offsets, etc.).

Regarding the architectural models for implementing the self-x functions, the following possibilities are distinguished ([2], [3]):

- Centralized SON (cSON): Solution where the self-x algorithms are executed at the NMS or at the EMS.
- Distributed SON (dSON): Solution where the self-x algorithms are executed at the Network Element level (i.e. autonomously within a single SC or in a distributed manner among several SCs).
- Hybrid SON: It combines cSON and dSON, in such a way that part of the self-x functionalities are distributed and reside at the SC while others are centralized and reside at the EMS and/or the NMS. This is illustrated in Figure 39. In this case, the cSON functions can be used to provide guidelines and parameters to the distributed SON functions based on information retrieved from them in terms of, e.g., performance measurements.

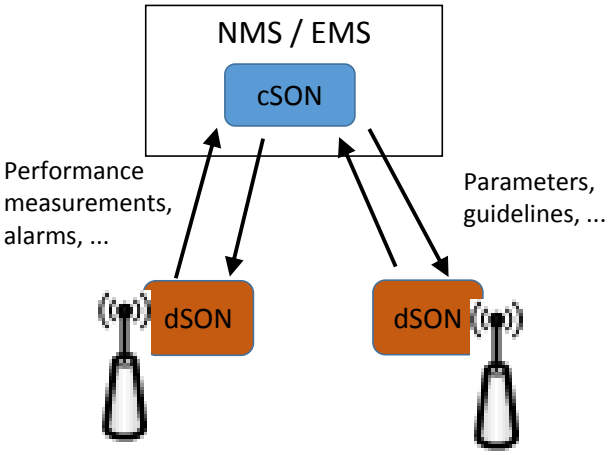


Figure 40: Hybrid SON Model

Based on the above model, Figure 40 depicts a simplified view of the SESAME architecture focusing on the relationship with Self-X functionalities.

As shown in the figure, the PNF EMS and SC EMS include the cSON functions and the centralised components of the hybrid functions. In turn, the dSON functions -or the decentralised components of the hybrid functions- reside at the CESCs. It is worth mentioning that, depending on the implementation, and assuming that the SLA monitoring function is effectively part of the EMS, as mentioned in *section 2.1.8.2.5*, it could also be possible to associate cSON to the SLA monitoring, which also has visibility of PM counters, KPIs and SLAs. This could relieve the SC EMS from the cSON decisions.

Furthermore, whatever self-x function is considered of interest to be deployed, it can be implemented as a PNF or, if proper open control interfaces with the element controlled by the self-x function are established, it can also be implemented as a VNF.

The implementation as VNFs provides:

(i) An inherent flexibility through easy instantiation, modification and termination procedures;

(ii) An inherent efficiency in hardware utilisation, since VNFs are executed on a pool of shared NFVI resources, and;

(iii) An inherent capability to “add” new functionalities and/or extend/upgrade/evolve existing VNFs. In SESAME, this can be applied to distributed self-x functions that would run as SC-VNFs in the light DC.

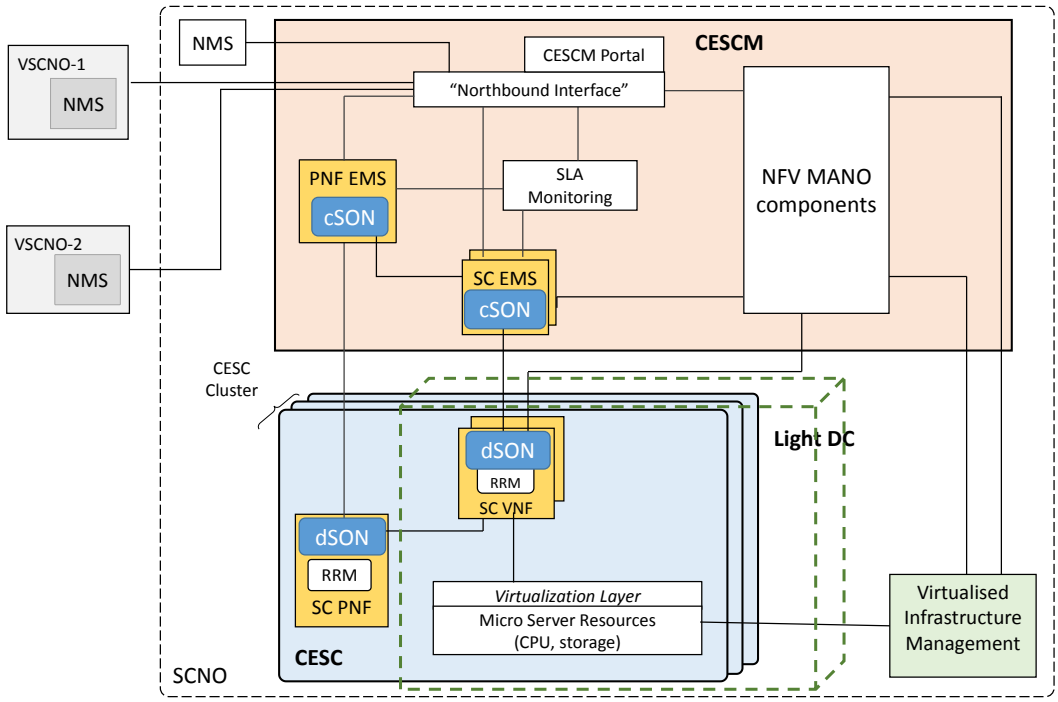


Figure 41: SESAME Architecture (simplified view) in relation to Self-X

Under the above framework, this section provides different initial studies regarding the use of self-x functions in the context of SESAME. Particular emphasis is put on the implications of multi-tenancy, because in multi-tenant scenarios -like those considered in SESAME- it should be distinguished between those self-x functions that are tenant-*specific* (i.e., the configuration of parameters can differ from tenant to tenant) and those that are common to all tenants. In this respect, *section 3.1* analyses the implications of multi-tenancy over the self-x functions related with mobility control and *section 3.2* performs an analysis considering the implications of the slicing process. Besides, *section 3.3* presents an artificial intelligence-*based* framework for extracting knowledge from the network and exploit in the development of self-x functions. This establishes a basis for the algorithmic development of self-x functions. The presented framework is particularized in different application use cases in *section 3.4*. Finally, *section 3.5* focuses on the use of self-x for content caching, reflecting that self-x in SESAME encompasses also additional views beyond the radio-*related* aspects.

3.1 Analysis of multi-tenancy support in self-x/RRM functions for mobility control

This section intends to analyse the implications of multi-tenancy on the RRM and Self-X functions that are related with mobility control. The main motivation behind this analysis is that, despite mobility control is a fundamental functionality to ensure a seamless experience to the User Equipments (UEs) of the different operators when moving across the cells of a shared RAN and when entering and leaving the shared infrastructure, no previous works in the literature have addressed this issue yet.

Let us consider the scenario depicted in Figure 42. The RAN of the SCNO is composed by different CESCs denoted in the following as small cells (e.g. small cells SA, SB, SC) and provides service to a Tenant (i.e. the VSCNO) e.g., within a stadium. The VSCNO is an MNO, with its own RAN around the stadium (e.g. cells TA, TB, TC).

Figure 42 includes the relevant elements of the SESAME architecture for the analysis considered here. In particular, the interconnection of the small cells (SCs) of the SCNO to the Evolved Packet Core (EPC) of the VSCNO is done through the S1 interface, delivering both data (e.g., transfer of end-users traffic) and control (e.g., activation of radio bearers) plane functions. Using current 3GPP principles [4], the support of MOCN at the small cell is provided by using the S1-flex mechanism that allows connecting one small cell to multiple EPC nodes (e.g. belonging to different VSCNOs). The SCNO's SCs are connected with the RAN of the VSCNO through the X2 interface. X2 connectivity can be provided through the X2 GW [4] in case it is used. The X2 interface allows neighbour cells to exchange different types of information (e.g. load, interference, handover information, trace information, information to support self-optimisation, etc.) for coordination purposes and supports procedures/messages for parameter negotiation (e.g. to request handover parameter changes, etc.) [5].

As shown in Figure 42, there is partial overlapping between the coverage of the SCNO's RAN and the VSCNO's RAN. Mobility of UEs between cells of both RANs is supported in both ways (i.e. referred to as incoming traffic when going from the VSCNO's RAN to the SCNO's RAN and outgoing traffic in the opposite direction).

Mobility control of connected terminals is realised through the handover (HO) function, which is one of the central RRM functionalities. The RRM-HO function, executed at each small cell (SC), is used to determine the cell to which a given UE is connected to. Decisions made by the RRM-HO function are based on measurements that are compared through a set of parameters (e.g., thresholds, offsets). The RRM-HO function at a specific SC only considers as candidate cells those that are listed in the so-called Neighbour Relation Table (NRT) associated to that SC. Some parameters could be statically configured from the NMS/EMS (where EMS encompasses here both the PNF EMS and the SC EMS) or dynamically adjusted at runtime by self-x functions such as ANR, MRO and MLB. In a general case, the RRM-HO and self-x functions implemented in the VSCNO's RAN and in the SCNO's RAN are likely to differ (e.g., different vendor's equipment in each RAN with vendor-specific implementations of RRM/self-x functions) or be differently parameterised.

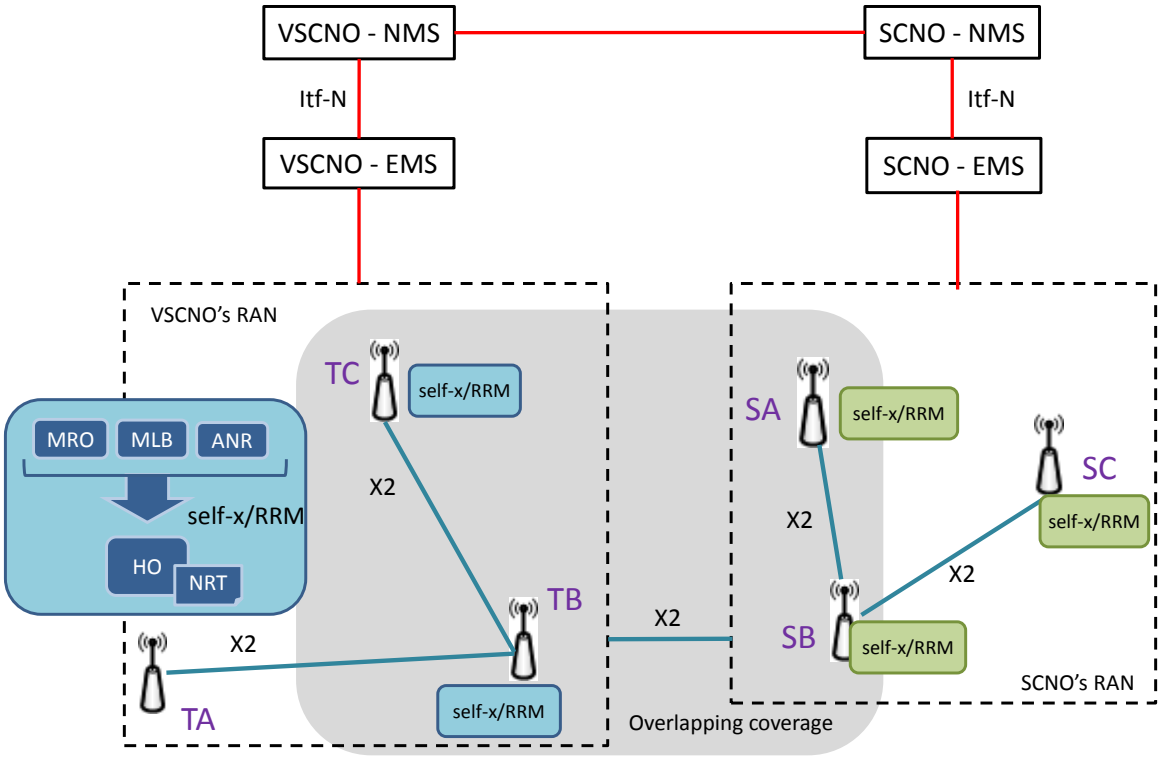


Figure 42: Considered Scenario for the analysis of Multi-Tenancy

3.1.1 RRM-HO function

The HO function commonly considers the measurement reports provided by the UE including e.g., Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ) values for the serving and neighbour cells. Furthermore, the HO algorithm can also consider as an input the load level at neighbour cells. In this case, load information is provided by neighbour cells via X2 interface. A detailed list of HO parameters is found in [6]. They are associated to the detection of certain events used by the HO algorithm to trigger the execution of an HO (e.g. detecting that a neighbour cell becomes offset better than the serving cell, detecting that a neighbour cell becomes better than a threshold, etc.). For each event, tuneable parameters include offset values, hysteresis values, time to trigger, thresholds, etc.

Typically, within a RAN, all cells (from the same vendor) will implement the same HO function. However, by setting the HO parameters on a cell-by-cell basis, the behaviour of the HO function (e.g. the precise time that a HO decision is made) can be different in each cell.

In order to properly steer the connected UEs across the VSCNO's RAN and the SCNO's RAN, neighbourhood relationships shall be properly captured in the corresponding cells. For the example of Figure 42, the NRT at TB should include SA and SB to enable an incoming handover to the SCNO's RAN. Similarly, the NRT at SB should include TB and TC to enable handovers to the VSCNO's RAN.

As long as coverage overlapping and traffic steering strategies between the SCNO's RAN and each of the VSCNOs are likely to differ, the RRM-HO function shall be VSCNO-aware (i.e. the RRM-HO shall be able to associate E-RABs with VSCNOs and enforce the VSCNO-*specific* policies). Furthermore, some of the parameters used by the RRM-HO function (e.g. offset values, hysteresis values) shall also be parametrised per VSCNO when pursing an optimised operation of the HO function. This latter aspect is analysed in the following from the perspective of the self-x functions that impact on the adjustment and optimisation of the different parameters used by the RRM-HO function.

3.1.2 Self-x -Automated Neighbour Relation

The configuration of the NRT in each of the cells can be realised through the Automated Neighbour Relation (ANR) function, avoiding the burden of human interactions between the VSCNO and SCNO to exchange information about the cells in close vicinity.

From the SCNO perspective, each SC will maintain a single NRT list that includes neighbour SCs from the SCNO and, in overlapping coverage areas, the cells of the different VSCNOs.

The ANR function relies on different procedures to find new NRs (Neighbour Relations), such as UE-*assisted* neighbour discovery that uses UE measurements to identify new NRs, network listen measurements done by the eNodeB (eNB), and X2 assisted network discovery (e.g., when a neighbour eNB attempts an X2 connection setup with another cell, it is automatically added in the NRT of this cell) [7].

When a new neighbour is detected, the procedures explained in *sections 2.1.6.6* or *2.1.6.8* will be used to setup the X2 with the new cell.

From the MME point of view, the eNB sends an eNB Configuration Transfer message to the MME. If the MME receives the SON Configuration Transfer IE, it shall transparently transfer the SON Configuration TransferIE towards the eNB indicated in the Target eNB-ID IE which is included in the SON Configuration Transfer IE. On the way back, the MME will send an MME Configuration Transfer. The purpose of the MME Configuration Transfer procedure is to transfer RAN configuration information from the MME to the eNB in unacknowledged mode. For instance, the eNB retrieves the IP address from MME to setup the X2 interface. Thus, in this case, the serving eNB gets in touch with the MME (if an X2 connection does not exist) to assist itself for creating a X2 tunnel with the target eNB. Once this tunnel is established, the serving cell eNB forwards the CGI[34]-info to the target eNB. Thus, both the eNB's updates its own respective NRT (Neighbor Relation Table).

Regarding UE measurements, a UE receives instructions about how to configure the measurement process. The UE will be indicated what frequencies to measure, possibly specifying as well the list of cells to measure in a given frequency (i.e., the cells that are defined in the NRT). This is indicated through dedicated RRC signalling for UEs in connected mode.

Similarly, the UE will be instructed about the reporting criterion (i.e. the event that triggers the UE to send a report). Example events can be detecting that a neighbour cell becomes offset better than the serving cell, detecting that a neighbour cell becomes better than a threshold, etc.

Given that the SCNO and a VSCNO will usually operate at different frequencies, the automated detection of neighbour cells from different RANs requires inter-frequency measurements, for which the UEs must be properly instructed while staying in the overlapping coverage area between the VSCNO and the SCNO's RAN. Then, on the one hand, UEs perform measurements at the frequency of the VSCNO while they are connected to the SCNO's RAN. On the other hand, UEs perform measurements at the frequency of the SCNO's RAN while they are connected to the VSCNO's RAN. For the example of Figure 42, TB should configure the UEs to measure on the frequency that the SCNO is operating. In this way, UEs will be able to detect SA and SB, report measurements from these cells and the ANR at TB will update the NRT at TB accordingly. Once this is accomplished, HO from TB to SA or SB will be possible. Equivalently, SB should configure the UEs depending on the VSCNO that they belong to, so that each UE will measure on the frequency of its VSCNO. In this way, the UEs of the VSCNO illustrated in Figure 42 will be able to detect TB and TC, so that the NRT at SB is updated accordingly. Therefore, measurement configuration of the UEs will be different depending on the VSCNO that they belong to.

[34] For more information see, *for example*: https://en.wikipedia.org/wiki/Common_Gateway_Interface

3.1.3 Self-X -Mobility Robustness Optimisation

Mobility Robustness Optimisation (MRO) function will automatically set HO parameters, aiming at avoiding different HO problems, such as connection failures due to mobility (too late handover, too early handover, handover to wrong cell), unnecessary HOs and ping-pongs.

In general, optimised HO parameters will be different at each SC because each SC exhibits a particular situation with respect to its neighbours. This will be particularly relevant to be considered for SCs deployed in the overlapping coverage areas between the SCNO and VSCNO's RAN. In this case, different coverage footprints from different VSCNOs will lead to different types of HO problems experienced by the UEs of each VSCNO. Therefore, the MRO function at the SCNO's RAN shall exploit HO-*related* performance measurements per VSCNO which can be useful in detecting the likely different HO-*related* issues arisen between the SCNO's cells and the cells of different VSCNOs. Consequently, VSCNO-*specific* HO parameter settings shall be supported to achieve optimised HO operation for all VSCNOs.

For example, let us consider the situation illustrated in Figure 43, for a UE heading from the SC to a VSCNO's cell. The overlapping area between the cell of VSCNO B and the SC is very small, while there exists a large overlapping between the cell of VSCNO A and the SC. In such situation, one could expect high call dropping rate (CDR) for UEs from VSCNO B due to too late handovers. The degradation of the CDR for VSCNO B in that specific cell should lead the MRO at the SC to conclude that the HO offsets should be reduced (so that, as soon as the VSCNO B's cell is detected, the HO is quickly executed). Regarding VSCNO A, if a low HO offset were defined for its UEs, these UEs would be handed-over to the VSCNO A's RAN at a very early stage. Then, depending on the HO algorithm and its parametrisation at the VSCNO A's RAN, a ping-pong effect might arise and UEs could be handed-over again to the SCNO's RAN. In such case, the observation of a relevant ping-pong effect should lead the MRO to conclude that the HO offset should be increased, so that UEs from VSCNO A would not be handed-over until a much stronger signal from VSCNO A's cell was received.

Regarding the HOs between two SCs of the SCNO's RAN, if the HO parameters were not sufficiently optimised and HO problems occurred, this would affect in the same way to all UEs, regardless of the VSCNO they are associated to. Therefore, the MRO function would be in charge of tuning the setting of HO parameters in the SC, without making distinctions among VSCNOs.

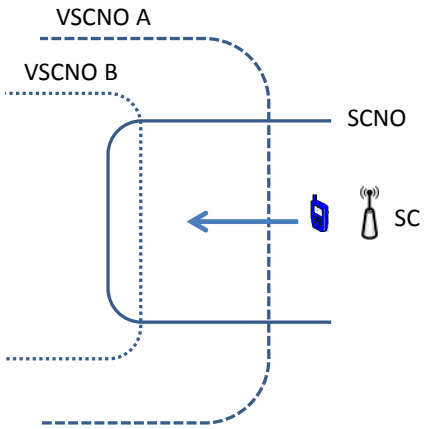


Figure 43: Impact of different overlapping between the VSCNOs' cells and the small cells on the MRO function

3.1.4 Self-X -Mobility Load Balancing

The objective of Mobility Load Balancing (MLB) is to distribute cell load evenly among cells or to transfer part of the traffic from congested cells to less loaded cells. MLB relies on exchanging cell specific load information between neighbour cells over the X2 interface (e.g. resource block usage separately for Guaranteed Bit Rate and non-Guaranteed Bit Rate E-RABs, available capacity that a cell can accept, etc.).

MLB function is a hybrid self-x function in which the MLB decisions are made at each SC according to policies controlled by the EMS and/or NMS [6]. MLB is supported by different procedures for transferring load between cells. For UEs in connected mode, MLB relies on the handover process, either by forcing HOs of specific UEs to a neighbour cell or by adjusting the HO parameters of a neighbour cell to facilitate that more UEs make HOs towards this cell. In this respect, the X2 interface includes a procedure to negotiate the changes in HO parameters between two neighbour cells, in order to facilitate a coordinated operation between them.

For UEs in idle mode, MLB relies on the modification of the cell reselection parameters (i.e. cell reselection offsets and cell reselection priorities) of each neighbour cell that are broadcast in the System Information Block (SIB) messages (e.g. SIB Type 4 for intra-frequency neighbours and SIB Type 5 for inter-frequency neighbours). When an idle mode UE detects a neighbour cell, it will use these parameters together with the received power to decide if it camps on this cell. Therefore, by adjusting these parameters, the MLB function can favour that more or less idle UEs camp in the different cells. However, in this case it is not possible to broadcast multiple parameters on a per VSCNO basis for a neighbour cell.

In a typical multi-tenant scenario, such as a stadium, high correlation among the traffic profiles (in time and space) associated to each VSCNO can be expected. Clearly, high load levels will be observed during e.g., a football match all over the stadium and for all the different VSCNOs simultaneously. However, some cases and situations (e.g., supporters from the visitor team are grouped in a certain area of the stadium, youth local supporters are usually grouped right behind the goalkeeper) as well as different market segments associated to the different competing operators acting as VSCNOs in the stadium (e.g., a low cost MNO will usually have youth customers, who in turn may stay in the areas of the stadium where attendees are standing) may lead to differences in the load levels associated to the different VSCNOs in the different cells. Therefore, the analysis of MLB strategies in multi-tenant scenarios requires further attention, since the load levels from the different VSCNOs in the different cells needs to be considered. At this point, it is worth remarking that X2 interfaces should be able to exchange load information on a per VSCNO basis.

To illustrate how the MLB actions can vary depending on the VSCNOs' load in different SCs, let us consider the example shown in Figure 44 with VSCNO A and B. Assuming the planned load level for each VSCNO, as shown in the left side of the figure, let us consider three different cases for the actual load distribution in small cell SC1, denoted as I, II, III.

Case I corresponds to an overload situation in which the aggregate load of both VSCNOs in SC1 exceeds the maximum acceptable level in the cell (i.e., the overload situation is causing performance degradation). This overload situation is due to VSCNO A, whose load substantially exceeds its planned level. In order to handle this situation, RRM Congestion Control techniques can reduce the load by e.g. reducing the bit rate of best-effort traffic. Additionally, assuming that a neighbour small cell SC2 has some capacity available, the MLB can transfer part of the load of SC1 to SC2 through the adjustment of HO parameters. In this case, only the load of UEs located in the overlapping coverage area between the two cells can be transferred to SC2.

Given that the transfer of a UE from SC1 to SC2 may cause some performance degradation (e.g., the UE is transferred to SC2 even if the received signal from SC2 is worse than the signal from SC1), the adjusted parameters favouring the HO should initially be applied to UEs from VSCNO A, since it is the VSCNO originating congestion.

The fraction of the load that can potentially be transferred to a neighbour cell depends on the UEs spatial distribution within the cell. In principle, UEs in SC1 that are in close vicinity to SC2

will be handed-over. By increasing the offset, UEs that are further away from SC2 can also be transferred, at the expense of a certain degradation in performance (e.g., lower peak bit rate), as Figure 44 illustrates.

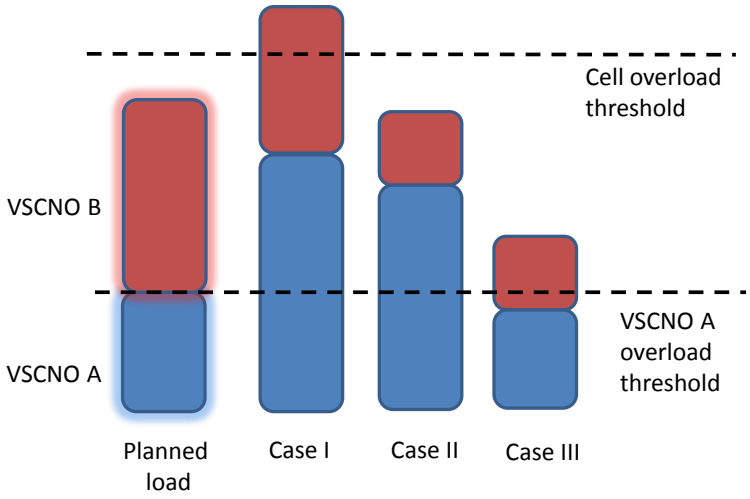


Figure 44: Illustration of MLB with different loads per VSCNO

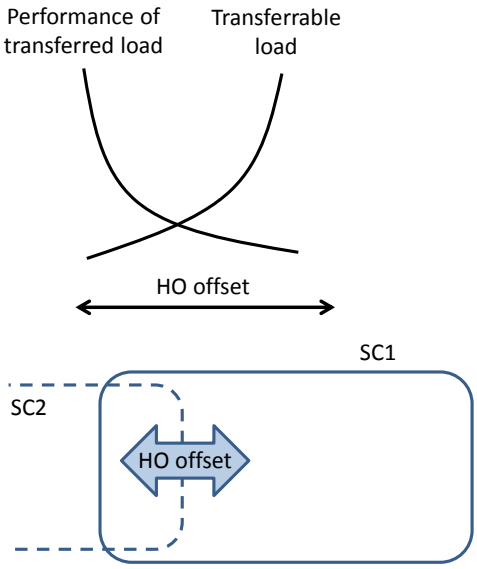


Figure 45: Illustration of trade-off between transferrable load and performance

In Case II of Figure 44, the load of VSCNO A in SC1 exceeds its planned level, but this does not generate overload in the cell. In this case, MLB actions to transfer some of the excess load of VSCNO A to SC2 would be required if the excess of load of VSCNO A causes some degradation in the performance of VSCNO B.

Finally, in Case III, neither the load of VSCNO A nor the load of VSCNO B exceeds the planned level in SC1. Therefore, MLB actions are not strictly needed in this situation. Still, in case that the load in SC2 were very low, transferring part of the load to this cell could be of interest if, by doing so, the performance observed by the UEs in SC1 could be improved.

In the case that load balancing involves two neighbour cells from different RANs (e.g., SB and TB in Figure 42), the availability of X2 interface between the SCs of the SCNO and the cells of the VSCNO facilitates the coordination for MLB purposes. The X2 enables that an SC receives information about the available capacity that each neighbour cell can accept. Computation of the

load per VSCNO should be supported and exchanged accordingly. Then, the MLB function at the SC can make decisions accordingly, thus minimising the risk that a HO is not accepted at a neighbour cell, or that a ping-pong occurs if the target cell of the VSCNO decides to make a HO back to the SC. Otherwise, in the absence of such load information, the SC can only initiate blind MLB actions to arbitrary neighbour cells, which would lead to ping-pong behaviour. Coordination of MLB at different SCs through the use of cSON will also avoid this undesired effect. Like in the previous example of Figure 45, MLB decisions will depend on how the UEs of each VSCNO are spatially distributed and on the overlapping coverage areas between the SC and cells of the different VSCNOs.

3.1.5 New Approaches to Mobility Load Balancing (MLB) and the User Association Problem

As discussed in the previous section, MLB is a self-optimisation functionality that intelligently spread users across system resources to ensure QoS and improve edge users throughput. MLB is typically triggered in response to local instances of overload. This reactive approach enables overloaded cells to redirect a percentage of their load to neighbouring less loaded cells hence alleviating congestion problems.

Traditionally, all users use the same set of handover parameters (e.g. hysteresis margin and time to trigger). Moreover, mobility and interference are normally treated separately. Ideally, a proactive approach to MLB is needed for MLB offloading taking into account interrelated factors such as interference, load, speed and including an enhancement to small cell discovery and user association.

Standard MLB makes use of Cell Range Expansion (CRE), which is achieved by either cell coverage or mobility parameters adjustments. CRE increases the downlink coverage footprint of a low power cell by adding a positive bias value. Offloaded users may experience unfavourable channel from biased cells and strong interference from unbiased higher power cells. CRE forces alternate cell selection without considering loading or resource allocation in the corresponding cell. Re-association of a user to a cell other than the one offering the largest signal strength as is sometimes implemented by traditional MLB approaches described above, often leads to reduced desired signal level and an increase in interference level which results in an overall network performance degradation. Advanced MLB makes use of CRE together with the 3GPP release 10 Almost Blank Subframes (ABS) and reduce-power ABS (RP-ABS) features, designed for macrocell interference mitigation, in HetNets. ABS is a time domain interference avoidance technique, which improves the overall throughput of the off-loaded users by sacrificing the throughput of unbiased cells. Given an ABS ratio (i.e. a ratio of blank over total sub-frames), a user may select a cell with maximum ABS ratio. CRE together with ABS is classified as distributed cell association scheme.

Advanced approaches such as multi cell load balancing in dense small cell deployments can help reduce blocking probability and improve network performance. Such action makes use of clustering of cells which in turn ensures that resources are appropriately allocated to groups of similar cells and the frequency of invocation of other SON algorithms is reduced, thereby minimising conflicts.

Examples of MLB clustering and cell selection approaches are illustrated in Figure 46:

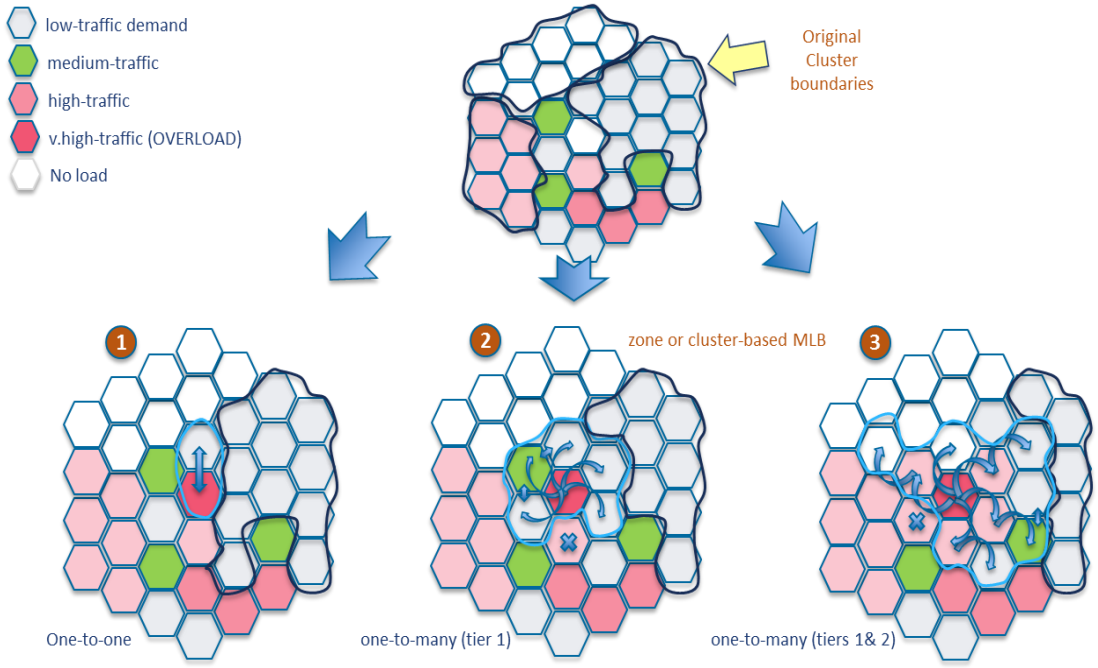


Figure 46: Classification of existing clustering and cell selection approaches from literature

The main modelling approaches for user association rely on utility modelling [46-50]. Examples of utility functions include spectrum efficiency, energy efficiency, QoS, outage and fairness. These approaches include game theory, combinatorial optimisation and stochastic geometry. The MLB user association problem is inherently a combinatorial optimisation problem. A new solution to this problem is based on the Knapsack Optimisation[35] algorithm which is well suited as a centralised optimisation approach to be implemented in the light DC and also addresses multi-tenancy. The future work will concentrate on the application of Knapsack Optimisation to different SESAME use cases and scenarios and comparing its performance to traditional approaches relying on CRE and ABS.

3.1.6 Conclusions

This section has analysed the implications of multi-tenancy on the RRM and self-x functions that support mobility control. In particular, the key role of ANR function has been analysed and the requirement to configure the measurements of each UE depending on the tenant that they belong to has been addressed so that the neighbourhood relations are properly captured at both the VSCNO's RAN and SCNO's RAN.

Regarding the MRO function, a distinction has to be made depending on the involved neighbour cells. For HOs involving the VSCNO's RAN and SCNO's RAN, the different coverage footprints of different VSCNOs will lead to different HO problems experienced by each VSCNO and, *consequently*, HO parameters have to be set differently for each VSCNO. For HOs between the small cells of the SCNO, the MRO function should not make distinctions among VSCNOs.

Regarding the MLB function, this section has shown that MLB actions in a cell should be different for each VSCNO, taking into consideration the different load distribution of each VSCNO in each cell and the spatial distribution of the UEs within the cell. Besides, it has been highlighted

[35] For the so-called Knapsack problem see, for example: https://en.wikipedia.org/wiki/Knapsack_problem

that different criteria for MLB could be considered in a multi-tenant context. A new centralised approach to MLB and user association in dense multi cell scenarios will be also studied.

3.2 Analysis of RAN slicing in relation to RRM and self-x functionalities

The focus of this section is placed on elaborating the network slicing concept applied to a multi-cell RAN, which is deployed by an infrastructure provider (i.e. the SCNO) over a certain geographical area and is shared among several tenants (i.e. VSCNOs) that offer mobile communications services to their own customers. The sharing is sustained in network slicing principles, so that each VSCNO is provided with a RAN slice, i.e. an instantiated logical RAN able to provision a specific capacity with potentially specific capabilities (ranging from just different network policy settings to the support of different protocols and features), realised together with other slices on the common physical infrastructure. As stressed in [35], isolation among slices is a fundamental requirement to ensure that the traffic of one slice does not negatively impact on other slices. Besides, it is essential that this isolation is implemented in a way that leads to an efficient use of radio resources.

Mechanisms to ensure isolation are well established in multi-tenant data centres or wired network domains [36]. However, when considering a RAN, the requirement of isolation is more challenging and deserves a deeper analysis due to the fact that the radio channel is an inherently shared medium. In this respect, isolation is identified in [37] as one of the main open research challenges in wireless network virtualization. In a single cell scenario, as considered in [38] and [39], isolation can be achieved by assigning an orthogonal set of physical radio resources for a certain period of time to each tenant in accordance with its requirements. Then, flexibility can be left to the tenant in the way that the radio resources are assigned to its customers during that period of time. Nevertheless, when considering the slicing of a multi-cell RAN, the isolation needs to consider also the potential interferences between transmissions of different tenants at different cells. Therefore, in addition to "traffic isolation", which refers to avoiding that variations in the traffic load level generated by one tenant affect the performance experienced by another tenant, also "radio-electrical isolation", which refers to avoiding mutual interferences on the air interface among the transmissions of different tenants, needs to be considered. In this respect, this section intends to analyse the multi-cell RAN slicing problem from a comprehensive perspective including these two aspects.

More in details, the analysis of RAN slicing is conducted in next section in relation to the RRM and self-x functionalities that can support the split of radio resources among slices. Based on this, four different RAN slicing approaches are presented. The analysis is completed with a comparison among the different alternatives considering different perspectives, such as the granularity in the assignment of radio resources to the different tenants or the degree of isolation and customisation.

3.2.1 RAN slicing

From a radio resource perspective, let us consider a given amount of spectrum resources and a RAN composed by N cells deployed over a certain geographical area. Without loss of generality, let us assume that the spectrum resources can be flexibly arranged in a set of carriers (i.e. channels), with the same or different carrier bandwidth, and that the time/frequency dimensions of each carrier are organized in Resource Blocks (RBs), which constitute the basic physical radio resource unit used for the dynamic allocation of the radio capacity. For example, in the context of Long Term Evolution (LTE) radio access, carrier bandwidths range from 1.4 to 20 MHz and a carrier of e.g. 5 MHz is organized in 25 RBs of 180 kHz/0.5 ms each.

From a service perspective, the RAN provides Radio Access Bearers (RAB), which are the data delivery services offered for information exchange between the User Equipment (UE) and the mobile core network. In the context of LTE, a RAB is denoted as Evolved-RAB (E-RAB), it is designed to transfer IP packets over the air interface and its expected behaviour is parameterized with a set of Quality of Service (QoS) attributes (e.g. guaranteed bit rate, QoS class identifier, Allocation and Retention Priority) associated with the particular RAB and/or with the corresponding UE (e.g. maximum aggregated bit rate for all RABs of a UE).

On this basis, the deployment and exploitation of the RAN resources (cells and spectrum) to fulfil a given traffic demand (number and characteristics of the RABs) involve the following functionalities:

- Spectrum planning: This function decides how the spectrum resources are arranged in carriers and how these carriers are assigned to the different cells.
- Inter-Cell Interference Coordination (ICIC): This function intends to mitigate the inter-cell interference that appears when neighbour cells are using the same carrier. For this purpose, ICIC establishes a set of limitations in the usage of the different RBs such as limiting the transmit power of a cell in certain RBs or forbidding the data transmission of a cell in certain RBs. Then, from a general perspective, the ICIC strategy determines the set of RBs that each cell is allowed to use in each of its allocated carriers as well as the maximum transmit power for each RB.
- Packet Scheduling (PS): This function decides, for each carrier assigned to a cell, how the set of available RBs (i.e. those allowed by the ICIC strategy) are used to transfer the data traffic of the established RABs. The scheduling process operates at a resolution given by the so called Transmission Time Interval (TTI), which currently is 1 ms in LTE and is expected to be even lowered in 5G to better support ultra-low latency applications. The PS is also responsible of selecting the physical layer parameters used in the RB transmissions (e.g. modulation and coding scheme, antenna mapping in case of multi-antenna transmission). This determines the amount of bits of each RAB that can be served in a TTI, thus controlling the bit rate (b/s) delivered to each RAB.
- Admission Control (AC): This function makes the decision on whether the establishment request of a new RAB is accepted or rejected in a given cell. The AC should account for the overall resource utilisation in the cell, the QoS requirements of already active RABs and the requirements of the new RAB request.

It is worth noting that the achievable resource utilisation efficiency on a specific RB (e.g. spectral efficiency in bit/s/Hz) depends on the physical layer techniques applied to the radio link (e.g., adaptive modulation and coding, hybrid automatic repeat request, multiple-input-multiple-output), the propagation conditions and the involved decision making processes (i.e., the aforementioned spectrum planning, ICIC, PS, AC), which impact on the interference level observed in the air interface.

Let us assume that the RAN is to be shared among M tenants, denoted as T_1, ..., T_M, and that each tenant is provided with a RAN slice. According to the above functionalities, different RAN slicing approaches are discussed in the following.

a) RAN slicing at Spectrum Planning level

In this case, which is illustrated in Figure 47, the RAN slicing is implemented by arranging the spectrum resources of the RAN infrastructure provider into a number of carriers and assigning different carrier(s) to each tenant. The amount and characteristics of the carriers assigned to a tenant should be sufficient to provide the required capacity and coverage over the whole scenario. Once particular carriers are assigned to a tenant, a tenant-*specific* spectrum planning function can decide how these carriers are used across the N cells, taking into account the specifics of the provided services and the spatial traffic distribution of their UEs. Moreover, given that the separate carriers ensure both radio-electrical isolation and traffic isolation, not only the spectrum planning function but the rest of mechanisms involved in the resource allocation to the RABs (i.e., ICIC, PS and AC) can be implemented according to tenant-*specific* policies. This is illustrated in Figure 47 by depicting in red the functions that can be particularized for tenant T_1 and in green those that could be specific for tenant T_M.

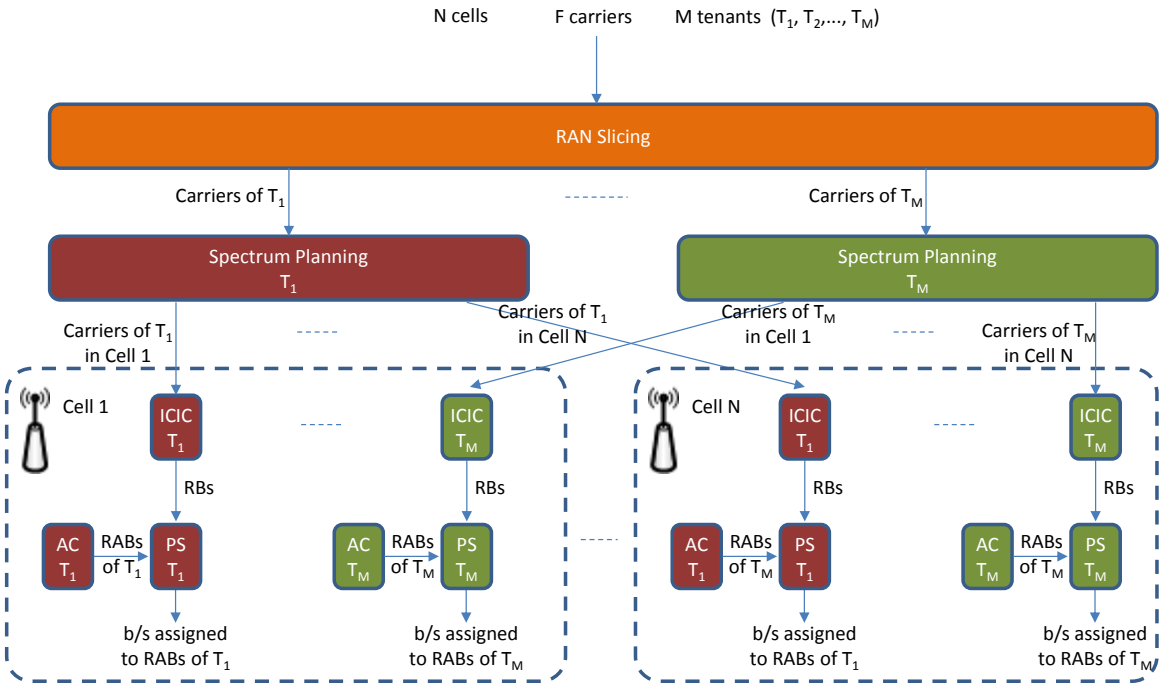


Figure 47: RAN Slicing at Spectrum Planning Level

b) RAN slicing at ICIC level

In this case, which is illustrated in Figure 48, the spectrum planning decides the assignment of carriers to cells by taking into account the expected spatial distribution of the traffic aggregate and without assigning dedicated carriers per tenant. In this way, any specific carrier assigned to a cell will in general be shared among the tenants.

Then, the RAN slicing is implemented at carrier level in each cell by assigning a certain number of RBs to each tenant. The assignment is done in a way that different tenants use different RBs across all the cells that can mutually interfere. This will ensure that no interference exists between transmissions of different tenants in these cells, thus achieving both radio-electrical isolation and traffic isolation between slices. Coordination across cells to carry out the split of RBs consistently is needed.

Once RBs are allocated per tenant, an ICIC strategy can be used to establish the usage of the allocated RBs among the different neighbour cells to mitigate the inter-cell interference within the RAN slice of the tenant, thus obtaining the set of RBs of a carrier that each tenant can use in each cell. Thanks to the isolation between slices, different ICIC policies might be applied in each slice (e.g. one tenant implements a full reuse of its RBs in all the cells while another tenant implements a partial reuse). Similarly, tenant-specific PS and AC algorithms can be implemented. This is illustrated in Figure 48 by depicting in red and green the functions that can be particularized for tenants T_1 and T_M, respectively, while the spectrum planning function, which is common for all the tenants, is depicted in blue.

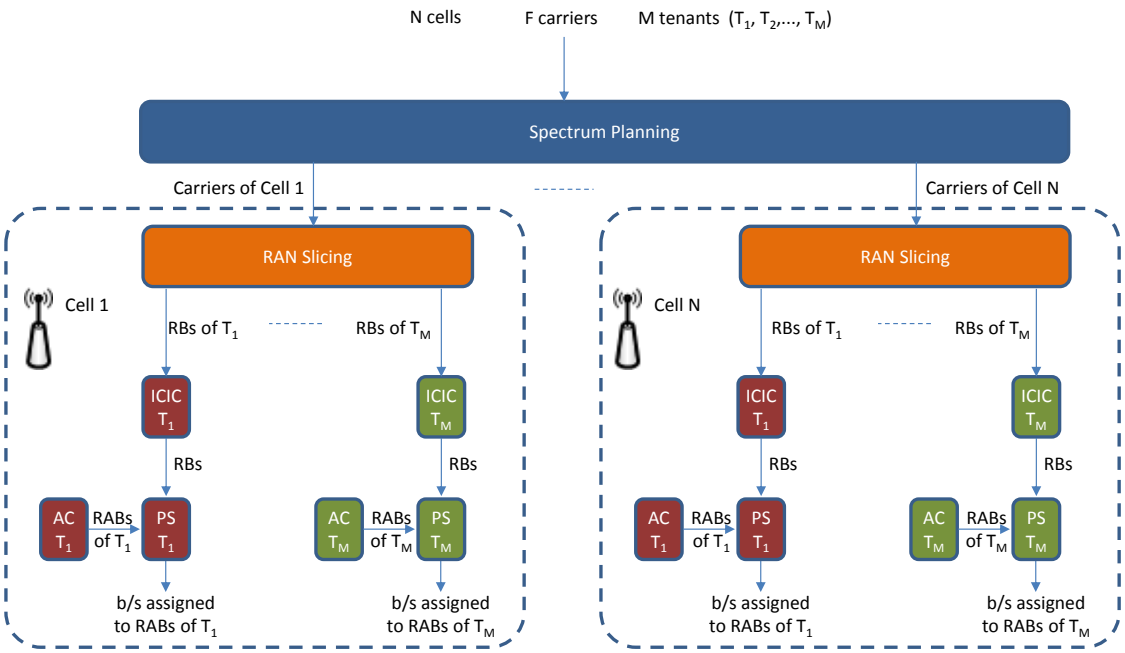


Figure 48: RAN slicing at ICIC Level

c) RAN slicing at Packet Scheduling level

This approach is illustrated in Figure 49. In this case, neither the spectrum planning nor the ICIC strategies assign dedicated radio resources (carriers and/or RBs) to the tenants across multiple cells. Instead, the RAN slicing is implemented at each cell by deciding the distribution of the set of allowed RBs among the tenants. By limiting the amount of RBs that can be used by each tenant, traffic isolation is ensured inside each cell. However, inter-cell interference among tenants may arise because the ICIC function does not make distinctions among tenants when deciding the RBs that can be used by a cell.

Thanks to the provided traffic isolation inside a cell, this approach enables the application of tenant-specific PS algorithms to decide how the capacity provided by the RBs of the slice is distributed among its admitted RABs (e.g. different prioritisation criteria to serve the RABs could be used for each tenant in high load conditions). In turn, spectrum planning and ICIC functions are common to all the tenants, so they are represented in blue in Figure 49.

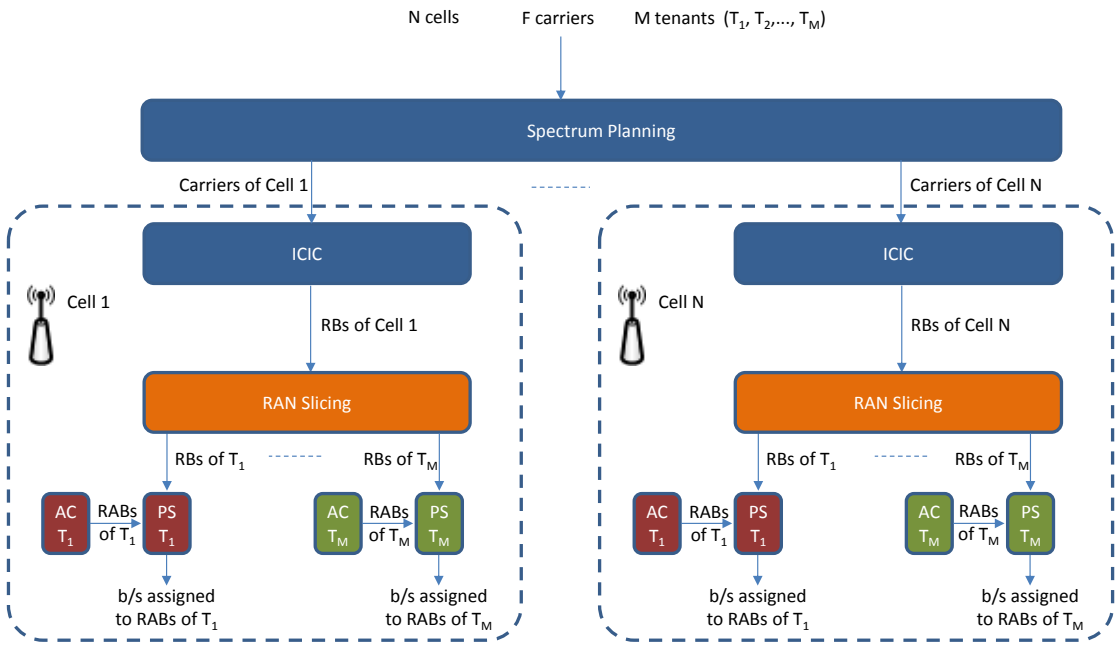


Figure 49: RAN Slicing at PS level

d) RAN slicing at Admission Control level

In this case, which is illustrated in Figure 50, RAN slicing is implemented by acting on the RAB admission process at each cell. A RAB requesting admission firstly goes through a tenant-*specific* AC that decides if it can be admitted or not (e.g., a tenant may implement a different allocation and retention strategy, giving priority or pre-empting some specific types of RABs when capacity limitations per slice are reached). Then, the RAN slicing function makes the final decision of acceptance or rejection considering the overall situation of all the tenants (i.e. requirements of all the RABs admitted, overall resource consumption, etc.) in order to ensure a proper split of capacity.

Like in the RAN slicing at PS level, radio-electrical isolation is not guaranteed because inter-cell interference between tenants may exist. Besides, the PS function in a cell is common to all the tenants (i.e., it is depicted in blue in Figure 50), so it distributes the available capacity among all admitted RABs without enforcing any tenant-specific treatment. For this reason, the AC executed by the RAN slicing is the only control mechanism to avoid mutual effects between the traffic of different tenants and thus achieve a certain degree of traffic isolation.

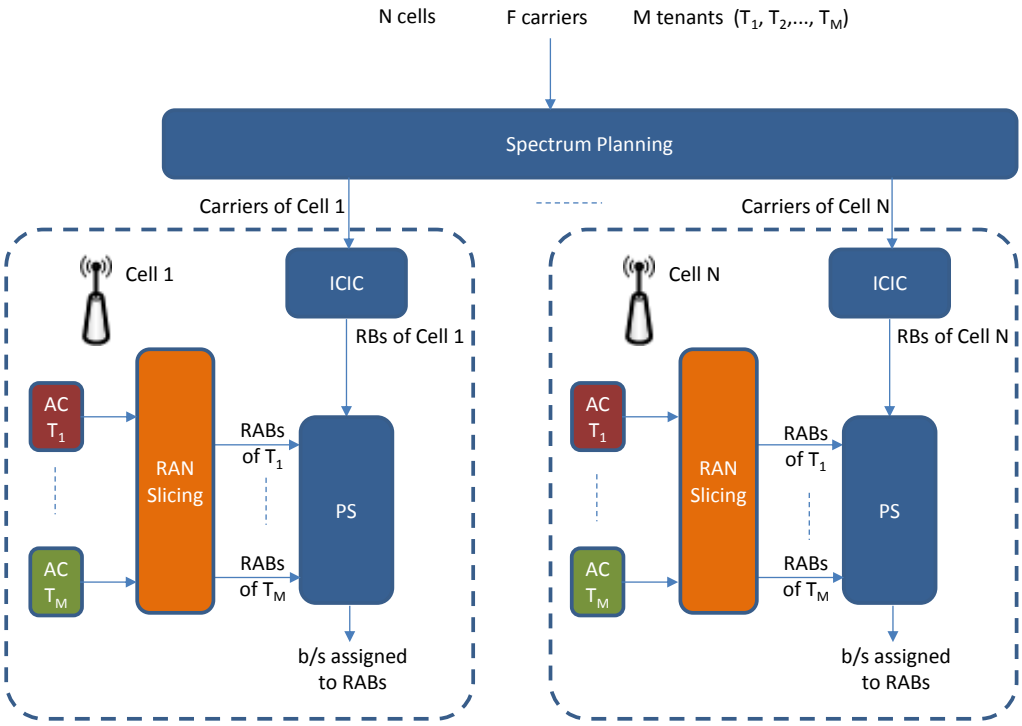


Figure 50: RAN Slicing at AC Level

3.2.2 Analysis of the RAN slicing approaches

The analysis of the proposed RAN slicing approaches should take into account different perspectives, as summarized in Table 37.

A first aspect to consider is the granularity achieved in the assignment of the radio resources to the different tenants, across the frequency dimension (i.e. the minimum amount of bandwidth that can be assigned to a tenant), the time dimension (i.e. how often the assignment of resources to a tenant can be modified) and the space dimension (i.e. the minimum geographical region where a resource assignment to a tenant applies). Granularity directly impacts on the flexibility to modify the resource assignments to adapt to different conditions and thus achieving a better utilisation of these resources.

Focusing on the frequency dimension, the granularity of the "RAN slicing at spectrum planning" is one carrier. Then, in the case of LTE, since there is flexible carrier bandwidth, the frequency granularity can be as low as 1.4 MHz and can scale up to 20 MHz. Instead, for both the "RAN slicing at ICIC" and "RAN slicing at PS" a better granularity is achieved, since the minimum frequency-domain unit that can be assigned to a tenant is a RB (i.e. 180 kHz in LTE). Regarding the "RAN slicing at AC" no frequency granularity exists because the slicing is not applied over the physical radio resources but over the RAB establishment requests.

In the time dimension, the "RAN slicing at spectrum planning" could change the carriers assigned to the tenants, which involves re-planning the network to modify the assignments of carriers to cells by means of dynamic spectrum allocation strategies [40]. Such re-planning processes are expected to occur on a relatively long-term basis. Instead, the "RAN slicing at ICIC" involves modifications not in the carriers assigned to the cells but only in the way how the RBs of these carriers are distributed among the tenants. Therefore, in this case the time granularity is constrained by the time scale of operation of the ICIC, which typically is in the order of some hundreds of ms [1]. In turn, for the "RAN slicing at PS", the time granularity of the RAN slicing will be determined by the PS that operates at TTI level, i.e. 1 ms in LTE. Finally, for the "RAB slicing at AC", the capability of dynamically modifying the split of capacity for each tenant is associ-

ated to time scale of operation of the AC, which is executed each time that a new RAB establishment request arrives to the system.

The analysis of the space domain granularity is related to the geographical scope over which the RAN slicing to tenants can be modified. For the "RAN slicing at PS" and "RAN slicing at AC", since slicing is performed at cell level, different splits of resources can be done at each cell, thus providing high flexibility. Instead, for the "RAN slicing at spectrum planning" or "RAN slicing at ICIC", since both involve the consideration of interference levels among cells, the space granularity will expand over the whole service area (or maybe over smaller portions of this service area if these portions do not mutually interfere).

Another relevant aspect to consider in the analysis is the level of isolation achieved with each approach. The highest radio-electrical isolation is obtained with the "RAN slicing at spectrum planning", because in this case each tenant will use different carriers in the whole scenario. Therefore, interference among tenants can only appear in the form of adjacent channel interference, which typically leads to Carrier to Interference Ratio (CIR) values well above 30 dB. For the "RAN slicing at ICIC", a high isolation is also achieved, because transmissions of different tenants in different cells will use different RBs. Therefore, no inter-cell interference among tenants will exist. Only adjacent channel interference or interference between adjacent subcarriers inside the same carrier due to imperfect synchronization and Doppler effect will be present, which again leads to high CIR values typically above 25 dB.

For the "RAN slicing at PS" and "RAN slicing at AC", inter-cell interference among tenants can potentially appear, depending on how the ICIC and spectrum planning strategies have distributed the carriers and RBs among the cells. For example, if the ICIC decides to reuse an RB in two neighbouring cells, this RB can be used by a different tenant in each cell, leading to inter-cell interference between both tenants. Instead, if the ICIC decides that an RB is not reused in two neighbouring cells, no inter-cell interference between tenants will exist.

Regarding traffic isolation, with RAN slicing at spectrum planning, at ICIC or at PS, it is ensured because the PS of a tenant can only use at maximum the carriers and/or RBs that the slicing process has assigned to this tenant. Therefore, an overload of this tenant will only impact on the performance experienced by the RABs of this tenant, but not on the RABs of the others. Instead, when RAN slicing is done at AC, the PS does not distinguish among the RABs of the different tenants when distributing the available capacity. Therefore, the isolation relies on the capability of AC and RAN slicing to restrict the admitted RABs of each tenant in order to prevent that a high load situation for one tenant can impact on others. However, due to the high variability of data traffic and the fact that AC typically operates based on statistical estimations of the resource consumption of the RABs, overload situations cannot be totally avoided, so this approach achieves the lowest traffic isolation among the considered ones and depends on how restrictive the AC is.

Finally, another aspect to consider is the capability of each technique to customise the different RRM and self-x strategies on a per tenant basis. The highest degree of customisation is provided by "RAN slicing at spectrum planning" that allows implementing tenant-specific policies for spectrum planning, ICIC, PS and AC. In turn, the "RAN slicing at ICIC" allows customising ICIC, PS and AC, while the "RAN slicing at PS" allows customising PS and AC on a per-tenant basis. The lowest customisation is obtained with the "RAN slicing at AC", where only AC can be tenant-specific.

	RAN slicing at Spectrum Planning level	RAN slicing at ICIC level	RAN slicing at PS level	RAN slicing at AC level
Granularity in the frequency domain	1 carrier (starting at 1.4 MHz)	1 RB (180 kHz)	1 RB (180 kHz)	Not applicable
Granularity in the time domain	Relatively long-term	Every ICIC period (typically hundreds of ms)	Every TTI (1ms)	Associated to RAB establishment request rate
Granularity in the spatial domain	Whole scenario	Whole scenario (might be less if sets of non-interfering cells are identified)	One cell	One cell
Radio-electrical isolation	High	High	Medium	Medium
Traffic isolation	High	High	High	Medium
Degree of customisation	Spectrum planning, ICIC, PS and AC can be tenant-specific	ICIC, PS and AC can be tenant-specific	PS and AC can be tenant-specific	AC can be tenant-specific

Table 37: Comparison between RAN slicing strategies

Analysis of the radio-electrical isolation

In the following, a numerical analysis is provided to further illustrate the behaviour of different RAN slicing approaches regarding the provided radio-electrical isolation. The analysis considers an illustrative scenario with 2 cells that provide access services to 2 tenants A and B. The distance between the two cell sites is 400 m.

Each cell has two carriers allocated, f_1 and f_2, each one with 25 RBs. The total transmitted power available per carrier is 10 dBm. For the "RAN slicing at spectrum planning", carrier f_1 is assigned to Tenant A and carrier f_2 is assigned to Tenant B. Instead, for the "RAN slicing at PS" the two carriers are shared indifferently between both tenants.

Two possible options are considered for ICIC, namely a Full Frequency Reuse (Full FR) and a Fractional Frequency Reuse (FFR) ICIC strategy [41]. In Full FR, all the RBs of a carrier are used by the two cells and, the total transmitted power is equally split among the 25 RBs (i.e., the transmitted power per RB is -4 dBm). In FFR, the 25 RBs of a carrier are split between 9 inner RBs and 16 outer RBs. The inner RBs are used at both cells and assigned to inner users (i.e., UEs at distance below 140 m). The outer RBs are further split in two equal sets of 8 RBs, one used at Cell 1 and the other at Cell 2. Outer RBs are assigned to outer users (i.e., UEs at distance above 140 m). For FFR, the total transmitted power in one carrier is equally split among the 17 RBs (9 inner + 8 outer) assigned to the carrier, so the transmitted power per RB is -2.3 dBm.

The analysis assesses the average downlink Signal to Interference and Noise Ratio (SINR) measured in an RB by the two reference UEs of Tenant A located at Cell 1 and depicted in Figure 51. UE1 is at distance 85 m from the cell site and UE2 is located at distance 180 m. Noise power per RB is -115.44 dBm and the propagation losses in dB are 128.1 + 37.6 log (d(km)).

SINR depends on the inter-cell interference, which arises if Cell 2 transmits on the same RB used by Cell 1 to transmit to a reference UE. Therefore, besides the carrier allocation and ICIC strategy, inter-cell interference is influenced by the traffic load at Cell 2, since the load level influences on how often the PS of the neighbour Cell 2 assigns this particular RB. Then, for illustrative purposes, it is assumed that Tenant A requires 6 RBs at Cell 2, while two different load conditions are considered for Tenant B at Cell 2: L1 (Tenant B has no traffic at all) and L2 (Tenant B requires 14 RBs). Besides, when FFR is considered, it is assumed that the number of required RBs in inner and outer areas is the same. As for the PS of Cell 2, it is assumed that, on average terms, it distributes the load of a tenant uniformly among all the RBs that can be used by that tenant.

The analysis also includes the maximum bit rate achievable by each reference UE. It depends on the SINR according to the model presented in section A.1 of [42] and assumes that a UE gets assigned in Cell 1 the maximum possible number of RBs.

Figure 51 shows the considered scenario and illustrates graphically with different colours how the two carriers f_1 and f_2 are used by the two tenants depending on the RAN slicing and ICIC approaches. The figure indicates the SINR (in dB) and the maximum bit rate (in Mb/s) for the two reference UEs with the different approaches and traffic load conditions.

For the Full FR strategy and “RAN slicing at spectrum planning”, the SINR and the maximum bit rate obtained by the two reference UEs of Tenant A are not affected by Tenant B’s load conditions because both tenants always use different carriers. Instead, with “RAN slicing at PS”, the two carriers are shared, and both the SINR and the bit rate of the two reference UEs are reduced when increasing the load of Tenant B in Cell 2 because this leads to a higher utilisation of the RBs in this cell and thus to higher inter-cell interference. The reduction in SINR is more significant for UE2 than for UE1 because it receives more inter-cell interference.

For the FFR strategy and the “RAN slicing at spectrum planning”, again the SINR and bit rate of the UEs are not sensitive to the load of Tenant B. In absolute terms, higher SINR compared to Full FR are observed because of lower interference, at the expense of lower peak bit rates because of the further split of radio resources. For the “RAN slicing at PS” and for UE1, who is an inner user, again a similar effect like with Full FR is observed (i.e., slight reduction of SINR and bit rate when increasing the load of Tenant B in Cell 2). However, more noticeable is the observation that UE2, who is an outer user, is not influenced by the load of Tenant B in Cell 2 in this case, because the FFR ICIC strategy ensures that it will not experience inter-cell interference thanks to the fact that outer RBs are used exclusively by Cell 1. In this way, UE2 achieves radio-electrical isolation from Tenant B through the ICIC technique. Nevertheless, this improvement is at the expense of a reduction in the maximum bit rate, because with FFR the maximum number of RBs that can be assigned to UE2 is limited by the amount of outer RBs in Cell 1.

Figure 51 also shows that the maximum bit rate achievable by a reference UE with “RAN slicing at PS” is higher than with “RAN slicing at spectrum planning”, because in the latter case the maximum number of RBs that can be assigned to a reference UE is limited by the number of RBs of carrier f1, while in the first case all the RBs of the two carriers can potentially be assigned to this UE. Therefore, “RAN slicing at PS” technique offers higher flexibility than “RAN slicing at spectrum planning” to allocate higher maximum bit rates to the UEs and to better distribute the load among the carriers.

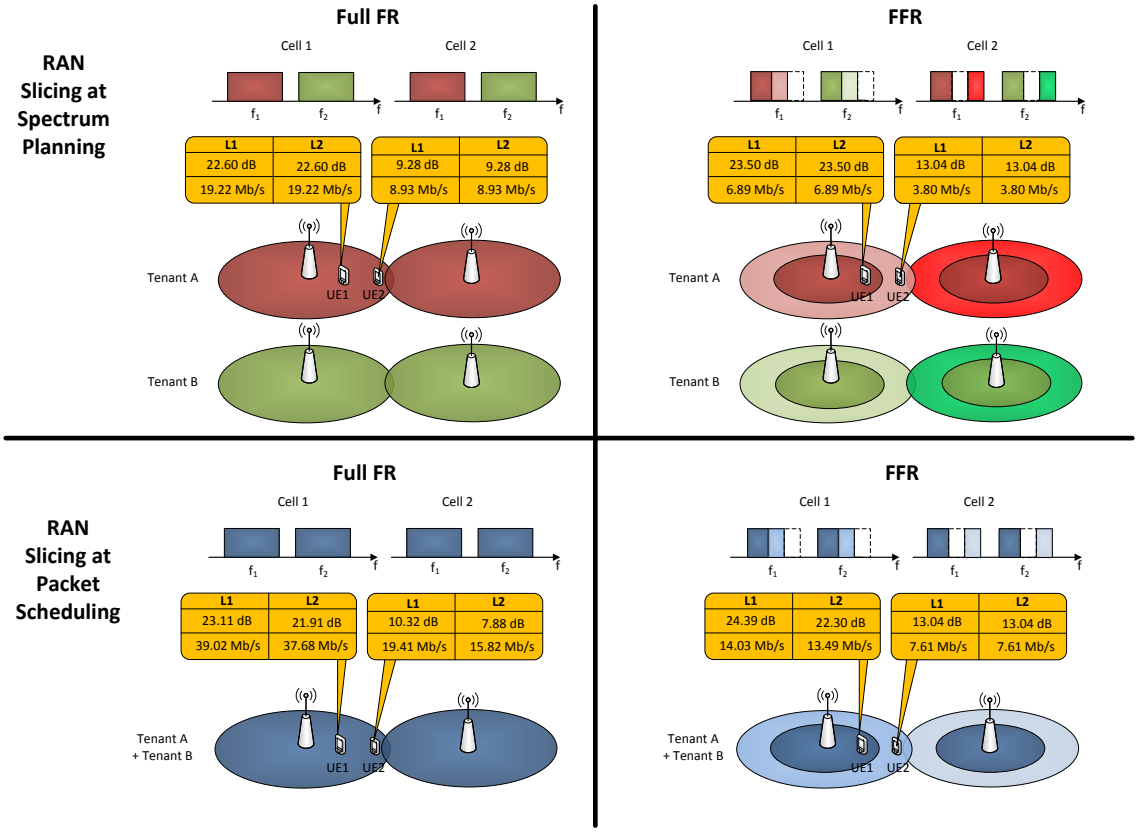


Figure 51: Analysis of radio-electrical isolation for different RAN slicing approaches

3.2.3 Conclusions

This section has analysed the application of the network slicing concept to a multi-cell RAN that is shared among multiple tenants, as considered in SESAME. Due to the characteristics of the radio channel and the potential influence that any transmitter may have on any receiver, the slicing of the RAN is a challenging process as it needs to consider isolation among tenants from two different perspectives, namely the traffic isolation and the radio-electrical isolation. Based on this, the RAN slicing problem has been analysed from a comprehensive perspective including these two concepts and four possible RAN slicing approaches have been presented that differ on the RRM functions used as a support for splitting the radio resources between slices.

The identified alternatives have been compared from different perspectives. The highest level of radio-electrical isolation is obtained with the "RAN slicing at spectrum planning" and "RAN slicing at ICIC" approaches, which prevent co-channel inter-cell interference among tenants, as a difference from "RAN slicing at PS" or "RAN slicing at AC", in which potential inter-cell interference among tenants may appear. Instead, these two techniques offer a higher granularity and more flexibility in the assignment of radio resources to tenants. In turn, the lowest level of traffic isolation is provided by the "RAN slicing at AC", while the other approaches ensure a higher isolation because they limit the amount of RBs that can be used by each tenant. The different approaches also offer different degrees of customisation among tenants, because they establish the RRM and associated self-x functions that have to be common to all the tenants and those that can be implemented following tenant-specific policies. In this respect, the highest customisation is provided by the "RAN slicing at spectrum planning", which allows particularizing the spectrum planning, ICIC, PS and AC functions on a per-tenant basis. Instead, the lowest customisation is obtained with the "RAN slicing at AC", where only the AC can be made tenant-*specific*.

3.3 Artificial Intelligence-*based* framework for Self-X

This section presents a general framework for the introduction of Artificial Intelligence (AI) mechanisms in the development of self-x functions. AI intends to develop intelligent systems able to perceive and analyse the environment and take the appropriate actions. The inclusion of AI in self-x provides the ability to smartly process input data from the environment and come up with knowledge that can be formalized in terms of models and/or structured metrics that represent the network behavior. This will allow gaining in-depth and detailed knowledge about the whole ecosystem, understanding hidden patterns, data structures and relationships, and using them for a more efficient network management. This enables to shift the Self-X functionalities, which currently exhibit an intrinsic reactive design approach and a lack of end-to-end (E2E) knowledge of the network [8], towards a more proactive behaviour able to exploit the huge amount of data available and to incorporate additional dimensions coming from the characterisation of end-user experience and end-user behaviour. In this respect, the higher level of knowledge about the network and its users constitutes a key differential factor between self-x in future 5G and in legacy systems.

Based on these considerations, this section establishes the general framework and identifies candidate tools for knowledge discovery together with the associated knowledge models that can be extracted. On this basis, the applicability of these models to a comprehensive range of-self-x functions across the categories of self-planning, self-optimization and self-healing is analyzed. Later on, in *section 3.4*, specific applicability use cases derived from this framework will be analyzed for some self-x functionalities.

Figure 52 illustrates the considered framework for AI-*based* self-x. Three main stages are identified:

- The acquisition and pre-processing of input data exploiting the wide variety of available data sources. In the 5G and big data era [9], it is feasible going far beyond the network data (e.g. performance measurements, network counters, etc.) that has been traditionally used in legacy systems, and consider also other dimensions such as the user-specific data, the content associated to applications or even external data from outside the MNO domain (e.g. planned events, weather forecasts, etc.).
- The Knowledge Discovery stage supported by AI-*based* tools that will smartly process the gathered input data to come up with exploitable knowledge models that represent the network/user behaviour in a way that can be directly used to make smart network planning and optimization decisions. AI-*based* tools rely on machine learning to carry out the mining of the input data and extract relevant knowledge models at different levels: cell level (contains the characterisation of the conditions on a per cell basis), cell cluster level (characterisation of groups of cells built according to their similarities) and user level (contains the characterisation of the conditions experienced by individual users). The general goal of machine learning is to build computer systems that can adapt and learn from their experience [10]. Specific machine learning functions that are relevant here are [11]:
    - *Classification:* It is the process of finding a model or function that describes and distinguishes data classes or concepts. The obtained model (i.e. the classifier) is then used to determine the class to which an object belongs. The object is the entity to be classified and it is usually represented by a tuple that includes a set of attribute values (e.g. an object could be a cell and the attributes could be different performance measurements associated to that cell). Classification process assumes that the possible classes are predefined in advance. Then, the classifier model is usually obtained from a supervised learning algorithm that analyses a set of training tuples associated with

known classes. Table 38 briefly describes different classification tools.

- *Prediction:* It intends to find models to anticipate future values of a certain parameter. Prediction models usually exploit the trend analysis of input data in terms of four major components, namely long-term movements that indicate the general direction in which a time-series graph is moving over a long interval of time, cyclic movements that refer to oscillations about a trend line, seasonal movements that are systematic or calendar related, and irregular or random movements that characterize the sporadic motion of time series due to random or chance events. Table 39 summarizes some relevant prediction tools.
- *Clustering:* It is the process of grouping a set of objects in a way that objects within the same cluster are similar to one another and dissimilar to the objects in other clusters. Clustering mechanisms do not rely on a set of predefined classes but these are obtained through an unsupervised learning process. Different clustering tools are described in Table 40.

- The Knowledge Exploitation stage will apply the obtained knowledge models to drive the decision-making of the actions associated to the self-x functionalities.

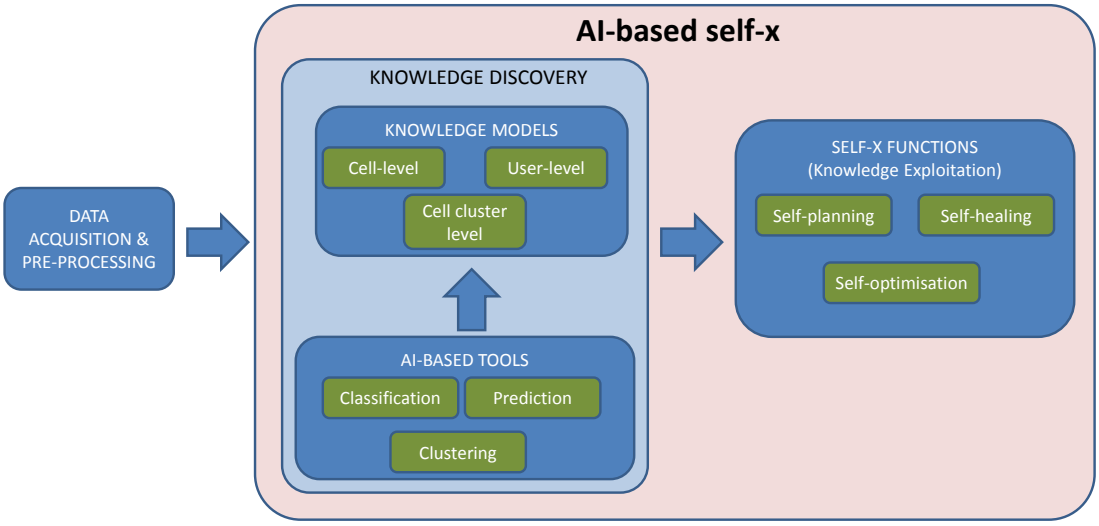


Figure 52: General framework for AI-based Self-X

Classification tools	
Decision Tree Induction[36]	Classification uses a flow-chart structure where each node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves represent classes. The flow-chart structure is built during the supervised learning stage through a top-down recursive divide-and-conquer manner, starting from the training set which is recursively partitioned into smaller subsets.
Bayesian classification[37]	Classification process evaluates the probability that a given tuple belongs to a class based on its attributes and selects the class with the highest probability. The probability computation is done using Bayes' theorem whose terms are obtained from the analysis of the training set.
Rule-*based* classification[38]	Classification uses a set of if/then rules obtained from decision trees or directly from the training data.
Fuzzy Logic[39]	Classification is similar to rule-*based* classification but allowing "fuzzy" thresholds to be defined for each class. Then, the classification is given in terms of a value between 0 and 1 that represents the degree of membership that a certain attribute value has in a given category.
Neural networks	Classification uses a feed-forward neural network that consists of an input layer, one or more hidden layers and an output layer. Each layer is made up of processing units called neurons. The input attributes of the object to classify are fed into the neurons making up the input layer. These inputs pass through the input layer and are then weighted and fed simultaneously to a second layer. The process is repeated until reaching the output layer, whose neurons provide the selected class. The weights of the connections between neurons are learnt during the supervised learning phase using a back propagation algorithm.
Support Vector Machines (SVM)	Classification uses the optimal boundary that separates the input tuples of the training set in their corresponding classes. This boundary is found during the training stage through a nonlinear mapping to transform the original training data into a higher dimension so that the optimal boundary becomes a hyperplane. SVM classifier is originally intended to do a binary classification, but multi-class SVM classifiers can be built by hierarchically combining multiple binary SVM classifiers.
K-nearest neighbour (k-NN)[40]	Classification is done by comparing the test tuple to classify with the set of training tuples and finding the k training tuples with shortest distance to the test tuple. The assigned class is the most common class among these k training tuples.

Table 38: AI-*based* tools for classification

[36] For more related information see, *for example*: https://en.wikipedia.org/wiki/Decision_tree_learning
[37] See, *for example*: https://en.wikipedia.org/wiki/Naive_Bayes_classifier
[38] See, *for example*: http://slidewiki.org/deck/1601_rule-based-classification#tree-0-deck-1601-1-view
[39] For an introduction to the concept of fuzzy logic see: https://en.wikipedia.org/wiki/Fuzzy_logic
[40] For more related information see, *for example*: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

Prediction tools	
Seasonal Auto-Regressive Integrated Moving Average[41] (ARIMA)	In ARIMA the time series is modelled as a linear combination of its past values and the past values of an error series. Seasonal ARIMA extends this with the inclusion of one or multiple seasonal factors to capture e.g. weekly variations, monthly variations, etc. The setting of the model parameters can be done automatically based on past input data applying algorithms such as the minimization of Akaike Information Criterion[42] (AIC).
Holt-Winters Exponential Smoothing[43] with seasonal patterns	The time series is modelled by a local mean, a local trend and a local seasonal factor that are updated by exponential smoothing. Either one or multiple seasonalities can be considered. The setting of the model parameters based on the past observations can be done automatically through the minimization of the AIC.
Support Vector Regression[44] (SVR)	The prediction function is obtained by solving a convex optimization problem based on the SVM concept [12]. It does not need to assume linear relationship of the predicted parameter with respect to previous values, so it can work in linear, non-linear, stationary and non-stationary systems.

Table 39: AI-*based* tools for prediction

Clustering tools	
Partitioning[45] methods	Clusters are formed to optimize an objective partitioning criterion such as a dissimilarity function based on distance. Existing partitioning algorithms include k-means[46], Partitioning Around Medoids[47] (PAM) or Clustering LARge Applications (CLARANS) [11]. The number of clusters has to be known in advance.
Hierarchical methods	Objects are grouped into a tree of clusters that is iteratively modified in an agglomerative (i.e. starting with clusters of one object and then merging them according to some similarity criterion) or a divisive way (i.e. starting with a single cluster that contains all the objects and progressively subdividing it into smaller clusters). Hierarchical algorithms do not require the number of clusters to be known in advance but it can be dynamically found by the process.
Density-*based* methods	Clusters are considered as dense regions of objects in the data space that are separated by regions of low density.
Grid-*based* methods	Clustering uses a multiresolution grid that quantizes the object space into a finite number of cells.
Self-Organizing Map[48] (SOM)	Clustering relies on a neural network model. Each neuron has an associated weight with as many components as the number of attributes of the input objects. An unsupervised learning process updates the values of the weights applying the Kohonen's algorithm[49] over each input object [13]. At the end, the weight of each neuron

[41] For more related information see: https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average
[42] See, *for example*: https://en.wikipedia.org/wiki/Akaike_information_criterion
[43] For more related information see, *for example*: https://en.wikipedia.org/wiki/Exponential_smoothing
[44] https://en.wikipedia.org/wiki/Regression_analysis
[45] For more relevant information see: https://en.wikipedia.org/wiki/Partition_(database)
[46] For more related information see, *for example*: https://en.wikipedia.org/wiki/K-means_clustering
[47] See, *for example*: http://www.unesco.org/webworld/idams/advguide/Chapt7_1_1.htm
[48] For more related information see: https://en.wikipedia.org/wiki/Self-organizing_map
[49] See, *for example*: https://en.wikipedia.org/wiki/Self-organizing_map#cite_note-KohonenMap-1

	captures the attribute values of a cluster of input objects.
Geospatial clustering	Clustered objects are associated to geographical coordinates and clustering is done by grouping points based on their connectivity, density and/or reachability in geographical space. Specific tools include Kernel Density Estimation[50] (KDE)[14], which is based on estimating the density of points, or the Geo-ProZones (GPZ) algorithm, which applies an adaptive partitioning method to detect regions with a high concentration of points [15].

Table 40: AI-*based* tools for clustering

3.3.1 Knowledge models to support AI-*based* Self-X

Different knowledge models derived from the AI-*based* tools in the knowledge discovery stage of Figure 52 are discussed in the following, for the three abovementioned cell, user and cell cluster levels. In turn, Table 41 analyses the potential applicability of these knowledge models for different self-x functions.

1) Cell-level models

This level includes the knowledge that characterizes the existing conditions in a cell. This encompasses the following models:

- *Traffic characterisation*. From a time-domain standpoint, this defines how the traffic of a cell varies as a function of time. The traffic can be measured in different ways, such as the load factor, the total number of users, the total data rate, etc., and it can be aggregated or split among Quality of Service (QoS) classes. From a space-domain perspective, the traffic characterisation can be done in different terms, such as the geographical distribution of the users, traffic load, services/applications or QoS class. In general, given the proliferation of multiple small cells that can be located indoors and deployed in tall buildings, a 3D characterisation can be required. Then, AI-*based* techniques will be applied over past observations of traffic in order to provide:
    - Classification of the time domain traffic pattern: Time correlations in the traffic evolution of a given cell should be detected to identify existing seasonalities at different levels (e.g. intra-day variations, variations during the week between working days and weekend, variations in the traffic between winter and summer, etc.) and classify the cell accordingly. The classification tools of Table 38 can be used to extract these models.
    - Learning the traffic behaviour in the time domain: This refers to the identification of a model that captures the cell traffic at different periods of time (e.g. hours, days of the week, etc.) and allows identifying time periods exhibiting similar traffic levels. Candidate tools to extract this knowledge include some of the clustering techniques listed in Table 40, such as partitioning, hierarchical methods or SOM.
    - Prediction of the future traffic: A prediction model can be extracted to anticipate future values of the traffic evolution in a cell. This can feed various decision-making processes regarding planning (e.g., in order to anticipate the need to deploy additional network nodes) and optimization (e.g., in order to tune handover parameters in neighbouring cells to absorb traffic if the cell is anticipated to be overloaded), mainly depending on the time scale at which

[50] See, *for example*: https://en.wikipedia.org/wiki/Kernel_density_estimation

the prediction is conducted. Candidate tools are the techniques listed in Table 39.

- Clustering spatial traffic (hot-spots): It targets the identification of concentrations of users in limited geographical areas. Candidate tools include the geo-spatial clustering techniques of Table 40.
- Learning mobility patterns: This intends to identify if the traffic follows some specific mobility patterns inside the cell that can be characterized in terms of prototype or representative trajectories followed by many of the users (e.g. trajectories directed towards specific points such as a metro station, etc.). Candidate tools to extract this knowledge include some of the clustering mechanisms of Table 40 like the partitioning methods or the SOM.

- *Performance characterisation*. The assessment of a cell's performance involves multiple measurements and Key Performance Indicators (KPIs) that can be organized under different categories depending on the specific performance criterion. Typically, it can be distinguished among accessibility KPIs (e.g., success rate in the set-up of new calls), retainability (i.e., how often an end-user abnormally loses a call or a session), mobility (e.g., number of handovers to different target cells, handover types or handover causes), QoS-*related* KPIs and associated measurements (e.g., cell bit rate, throughput per user, latency and packet loss rate), resource utilisation related measurements (e.g., transmit power per carrier, percentage of time that all the resource blocks devoted to traffic have been used), and RF measurements (e.g., distributions of the Channel Quality Indicators (CQIs), of the received power levels, etc.). Traditionally, measurements such as accessibility rate or dropping rate are aggregated on a cell basis and averaged along relatively long-term periods (e.g. days or weeks). These averaged values are used to trigger various optimization processes. Instead, the AI-*based* approach presented in this work intends to attain much deeper exploitation of these KPIs and extract additional knowledge based on the time and spatial domain analysis of the abovementioned indicators. Some possibilities are listed in the following:
  - Learning the time domain pattern of a performance indicator: This characterizes the time evolution of a given performance indicator, with the objective of identifying existing hidden patterns that would remain undetected if only aggregated measurements along several days/weeks were considered. For example, it can be automatically detected if the dropping rate in a cell exceeds certain thresholds during some specific hours, and if this situation exhibits some regularity, meaning that actions should be triggered to optimize the performance for those specific hours. Clustering tools listed in Table 40 like partitioning, hierarchical methods or SOM are candidates to extract this knowledge.
  - Prediction of a performance indicator: This involves the definition of a prediction model to anticipate future values of a performance indicator at different time scales. Prediction should be based on past observations of the indicator, but it can also consider past observations of other related indicators as additional features. For example, the prediction model of the throughput per user can take as inputs the observations of the signal to noise and interference conditions seen by the users and the resource usage. Prediction models of Table 39 can be used to extract this knowledge.
  - Learning space-domain black-spots: The characterisation of the performance

indicators in the spatial dimension allows the identification of specific areas where the desired performance limits are not met, such as black-spots where there is a high concentration of dropped calls, reduced throughput levels or low signal strength. Candidate tools include the geo-spatial clustering tools of Table 40.

- Classification of the general performance status of the cell: Given the high number of existing performance metrics, and the fact that hidden correlations between them can exist, it can be useful to combine these metrics into a simpler indicator that reflects the overall performance behaviour of the cell. Supervised classification tools like those of Table 38 can be useful here.

2) User-level models

The capabilities offered by big data and big data analytics technologies for processing unprecedented amounts of data will enable the exploitation of the user-data dimension in the different management processes of future 5G networks. Besides supporting relevant business processes, such as customer experience management, valuable knowledge extracted from the characterisation of the network usage made by the individual users can be exploited, not only for increasing the efficiency of the network optimization decision making processes, but also for a better personalization of the network services offered to the different users. Clearly, the extraction of this knowledge involves a trade-off between the achievable degree of user personalization and the complexity associated to the huge volume of data that needs to be processed to achieve it. In general, all the cell-level models described above admit disaggregation and analysis at user level. Nevertheless, *in particular*, the following components are envisaged so that to have high interest and applicability for the user-level characterisation:

- *Time-domain traffic pattern characterisation*: This should reflect the behavioural patterns of an individual user when experiencing mobile service in terms of the type of services consumed at different periods of time or the traffic volume generated. Examples include the classification of the time domain traffic pattern of the user (e.g. primarily use of streaming services at night, primarily use of web navigation at noon, etc.), learning the time traffic pattern of a user in order to identify regularities, the extraction of a prediction model at user level to anticipate the service demand of individual users, etc.

- *Spatial-domain traffic pattern characterisation*: This captures the behavioural patterns of an individual user from the spatial perspective, reflecting which cells the user has been connected to, together with the type of services/applications and the traffic volume generated by the user in each cell. Besides, the analysis of the order in which the user connects to the different cells along a service session can provide information about the trajectories followed by the user (e.g. travelling from home to the office, going from home to the gym, etc.) and this can be used to predict the next cells that a user will be connected to, detect the prototype trajectories that the user follows along the day, etc.

- *Performance characterisation*: This identifies the performance experienced by an individual user in terms of the different KPIs discussed above, such as accessibility, dropping rate, throughput, latency, etc. Besides, the comparison between the performance experienced by the user and the performance at the cell level can be a useful indicator to decide if specific actions need to be carried out, since it may happen that the overall cell performance (i.e. aggregated or averaged for all the users) is adequate but a specific user is repetitively affected by bad performance (e.g. because the user is very often located in an area of the cell with poor coverage, etc.) so that

actions can be triggered (e.g., increase priority level in packet scheduling mechanisms) to enhance user's satisfaction.

3) Cell clustering models

Cell clustering refers to the process of identifying groups of cells that exhibit certain similarities, so that a more efficient management can be carried out by considering the group of cells "as a whole" rather than on considering each cell individually. The cell clustering model can capture different perspectives:

- Clustering cells that exhibit a certain degree of mutual interaction, e.g. in the form of inter-cell interference, coverage overlapping, neighbouring relationships, etc. In this case, cell clustering will be closely linked to the geographical proximity between cells. The adequate cluster size or the specification of the cluster borders need to be considered. Acquiring knowledge about commonalities affecting a certain group of cells can lead to more efficient decisions taken at area level rather than at cell level both in terms of planning (e.g., high load levels in several neighbouring cells may advice the addition of a new site in the area) and operations (e.g., high dropping affecting a number of neighbouring cells can be associated to an external source of interference affecting a wide area).

- Clustering cells that exhibit similar characteristics or a similar behavior, regardless of whether they are closely located or not. The rationality of this approach is that the knowledge learnt for one cell can be valid for the rest of cells of the same cluster. In other words, SON functions can benefit from the knowledge of cells that exhibit a similar behaviour, because the actions that are learnt to be good (bad) for one cell can also be good (bad) for other cells of the same cluster. The clustering process according to the cell similarities can be done based on different dimensions: (i) performance (e.g. group cells that offer similar accessibility, retainability, QoS KPIs, etc.); (ii) traffic characteristics (e.g. group cells that exhibit similar traffic volume, service distributions, etc.); (iii) RF characteristics (e.g. group cells with similar received power distributions, interference, etc.).

  While the performance-*based* clustering may lead to a highly aggregated vision, since it will be the result of multiple different effects, the clustering based on e.g. traffic or RF may provide a more detailed vision of the cell behaviour. The identification of cells that are similar from the RF perspective and with similar traffic types may allow extrapolating effects from one cell to the other. For example, the behaviours of a cell in front of a given traffic level, can be extrapolated to other cells that have similar characteristics, so it is possible to predict how these cells will behave when they reach the traffic volume as the initial cell.

  When doing this clustering, each cell is considered as an object characterized by a number of features including aspects such as the RF measurements, traffic patterns, static attributes, user-level characteristics, etc. The clustering algorithm will then analyse these features and will group the cells according to their similarity, usually measured in terms of a distance function. Candidate tools to perform this process include the techniques discussed in Table 40. Specifically, hierarchical methods can be particularly suitable as long as the number of possible clusters is not known a priori but it has to be derived from the observed behaviours of the involved cells.

Knowledge Models / Self-x functions		Cell-level									User-level			Cell clustering	
		Classification of the traffic pattern	Learning the traffic behavior in time domain	Traffic prediction	Clustering spatial traffic (Hot-spots)	Learning mobility patterns	Learning time pattern of a performance indicator	Prediction of a performance indicator	Learning space-domain black-spots	General performance status of a cell	User time-domain traffic pattern	User spatial-domain traffic pattern	User performance	Cell clustering based on interaction	Cell clustering based on similarities
Self-planning	Planning a new cell			●	●	●		●	●	●					
	RF planning of a new cell				●	●			●						
	Spectrum planning	●	●	●			●	●		●	●			●	●
Self-optimization	Coverage and Capacity Optimization				●	●			●					●	●
	Automatic Neighbor Relations					●	●		●					●	
	Intercell Interference Coordination		●	●	●		●	●	●			●	●	●	
	Mobility Robustness Optimization				●	●	●	●	●			●	●	●	
	Mobility Load Balancing	●	●	●	●	●	●	●	●					●	
	Optimization of admission control / congestion control / packet scheduling parameters		●	●			●	●			●	●	●		●
	Energy Saving	●	●	●	●		●	●	●		●	●	●	●	
	SON coordination	●	●	●	●	●	●	●	●		●	●	●	●	●
Self- healing	Cell Outage Detection	●	●				●		●	●					
	Cell Outage Compensation	●	●	●	●		●	●	●					●	

Table 41: Applicability of the Knowledge Models in different Self-X functions

3.3.2 Conclusions

This section has presented a vision of an AI-*based* self-x framework that processes input data from very different sources and extracts, through learning-*based* classification, prediction and

clustering models, relevant knowledge models used to drive the self-x decisions. Following a taxonomy of self-x functions and a detailed list of AI-*based* tools that could empower the framework, a number of potential knowledge models and their applicability to the different Self-X functions have been identified.

3.4 Applicability of AI-*based* framework for self-x

3.4.1 Classification of cell-level time domain traffic

Based on the general framework for AI-*based* self-x that has been discussed in *section 3.3*, this section addresses the applicability of the knowledge extracted from the classification of the cell-level time domain traffic [19].

The cell-level time domain traffic defines how the traffic of a cell (e.g. a CESC in the context of SESAME) varies as a function of time. Traffic can be measured in different ways, such as the load factor, the total number of users connected to the cell, the total data rate, etc., and it can be aggregated or split among QoS classes and/or tenants. The traffic in a cell will be tightly related with the environment where the cell is deployed and with the characteristics and profiles of the users served by the cell. This will lead to time correlations in the traffic evolution of a given cell at different levels (e.g. intra-day variations in which the traffic can substantially differ between mornings or nights, variations during the week between working days and weekend, etc.). The detailed analysis of these correlations will allow extracting valuable knowledge that can be used for making management decisions regarding the configuration of a cell. In this respect, this section focuses on the application of classification techniques to extract this knowledge. In particular, the cells will be classified based on their historical traffic samples. The possible classes will indicate certain behaviours of the cells that are relevant for different Self-X functions. In the following parts we start by providing the general classification methodology and then we particularize it according to its applicability in some selected use cases, providing some results obtained using data extracted from a real mobile network.

3.4.1.1 General classification methodology

The input data for each cell i is a time series $\mathbf{X_i}=(x_i(t), x_i(t-1), ..., x_i(t-(N-1)))$ composed of N samples of the measured traffic in the cell i at different times t. The objective of the classifier is to make an association between the input time series $\mathbf{X_i}$ and a class $C(\mathbf{X_i})$ that characterizes the behavior of the cell's traffic in the time domain. The number and the type of classes will depend on the specific applicability of the classification outcomes, as it will be detailed in the use cases that will be presented later on.

Since the number of time samples N will typically be a very large value (e.g. reflecting the traffic measured in a cell in periods of some minutes and collected during several weeks, months, etc.), it will not be feasible to use the time series $\mathbf{X_i}$ directly as input of a classifier tool. Therefore, an initial processing is carried out to come up with a vector $\mathbf{F(X_i)}$ of shorter dimension M that preserves the relevant characteristics of the traffic pattern. This vector will be the input of the classifier. Following the usual terminology in classification [11], vector $\mathbf{F(X_i)}$ represents the tuple to be classified and each of its components represents a feature or attribute. Again, the definition of the mapping between $\mathbf{X_i}$ and $\mathbf{F(X_i)}$ will be dependent on the specific applicability of the classification, so it will be detailed later on when analyzing the different use cases.

The classifier will perform the association between the input $\mathbf{F(X_i)}$ and the class $C(\mathbf{X_i})$, as illustrated in Figure 52. The internal structure of the classifier will be given by the specific classification tool being used and its settings will be automatically configured through a supervised learning process executed during an initial training stage. This training will use as input S time series $\mathbf{X_j}$ j=1,...,S of some cells whose associated classes $C(\mathbf{X_j})$ are pre-defined by an expert. In this way, the training set will be composed by the S tuples $\mathbf{F(X_j)}$, j=1,...,S and their associated classes $C(\mathbf{X_j})$. The supervised learning process will analyze this training set to determine the appropriate configuration of the classification tool. The overall process is illustrated in Figure 52.

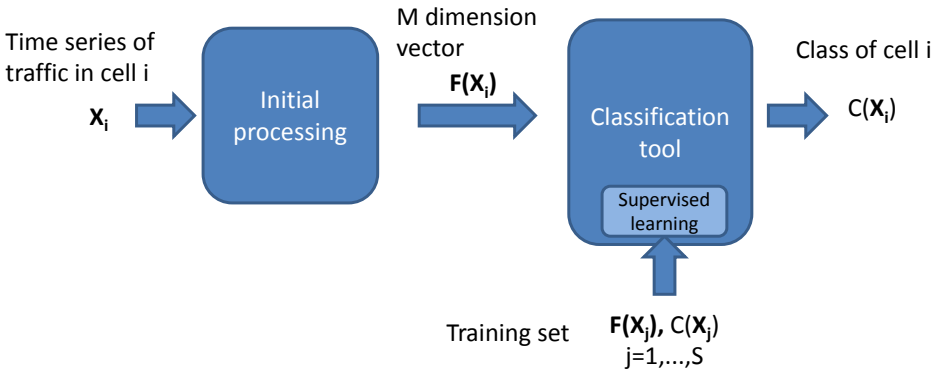


Figure 53: General Classification Methodology

3.4.1.2 Classification tools

Regarding the classification tool, the following alternatives are considered [11]:

- *Decision tree induction*: The classification is done by means of a decision tree, which is a flow-chart structure where each node denotes a test on a feature value, i.e. a component of vector $\mathbf{F(X_i)}$, each branch represents an outcome of the test, and tree leaves represent the classes. The tree structure is built during the supervised learning stage through a top-down recursive divide-and-conquer[51] manner, starting from the training set which is recursively partitioned into smaller subsets.
- *Naive Bayes classifier*[52]: In this case the classifier evaluates the probability Prob($C(\mathbf{X_i})|\mathbf{F(X_i)}$) that a given cell $\mathbf{X_i}$ belongs to a class $C(\mathbf{X_i})$ based on the values of the features $\mathbf{F(X_i)}$. The resulting class is the one with the highest probability. The computation of this probability is done using Bayes' theorem[53] under the "naive" assumption of class conditional independence, which presumes that the effect of a feature value on a given class is independent of the values of the other features. In turn, the different terms in the computation of the Bayes' theorem are obtained from the analysis of the training set.
- *Support Vector Machine (SVM)*: A SVM is a classification algorithm based on obtaining, during the training stage, the optimal boundary that separates the vectors $\mathbf{F(X_j)}$ of the training set in their corresponding classes $C(\mathbf{X_j})$. The obtained boundary is then used to perform the classification of any other input vector $\mathbf{F(X_i)}$. To find this optimal boundary, it uses a nonlinear mapping to transform the original training data into a higher dimension so that the optimal boundary becomes an hyperplane. Although SVM classifier is originally intended to do a binary classification, a multi-class SVM classifier can easily built by hierarchically combining multiple binary SVM classifiers. Each of these binary classifiers specifies whether the cell belongs or not to a given class.
- *Neural Network*: The classification is done by means of a feed-forward neural network[54] that consists of an input layer, one or more hidden layers and an output layer. Each layer is made up of processing units called neurons. The inputs to the classifier, i.e. each of the components of vector $\mathbf{F(X_i)}$, are fed simultaneously into the neurons making up the input layer. These inputs pass through the input layer and are then weighted and fed simultaneously to a second layer. The process is repeated until reaching the output layer, whose neurons provide the se-

[51] For more relevant information see, *for example*: https://en.wikipedia.org/wiki/Divide_and_conquer_algorithms
[52] See, for example: https://en.wikipedia.org/wiki/Naive_Bayes_classifier
[53] For more relevant information see, *for example*: https://en.wikipedia.org/wiki/Bayes%27_theorem
[54] https://en.wikipedia.org/wiki/Feedforward_neural_network

lected class C($\mathbf{X}_i$). The weights of the connections between neurons are learnt during the training phase using a back propagation algorithm.

The abovementioned general classification methodology presents applicability in different self-x functions. In the following sub-sections, the methodology is particularized for two use cases associated to two different functions.

3.4.1.3 Use case 1: Energy saving

This use case aims at reducing the energy consumption in the deployed cellular network. According to the Mobile's Green Manifesto report [20], approximately 80% of the energy consumption and Green House Gas emissions of mobile operators is caused from their networks. From an economical perspective, if all networks with above-average energy consumption were improved to the industry average, there is a potential energy cost saving for mobile operators of $1 billion annually at 2010 prices. In case of improving to levels of the top quartile the cost saving could be more than $2 billion a year [20]. Therefore, techniques intended to reduce the energy consumption are relevant for operators of current and future networks.

In this use case, the energy reduction is done by switching off the cells that carry very little traffic at certain periods of the day (e.g. at night) and making the necessary adjustments in the neighbor cells so that the existing traffic can be served through some other cell. In this context, the classification methodology of *section 3.4.1.1* can be used to identify candidate cells to be switched-off based on their traffic patterns. The automation of this procedure based on expert criteria captured in the training set becomes particularly useful considering that networks in the envisaged ultra-dense scenarios for future 5G systems can comprise several tens of thousands of cells. Therefore, it is not practical that a human expert can make this classification manually. It is worth mentioning, *however*, that the final decision on whether or not to switch off a cell would make use of this classification as well as other possible inputs which are out of the scope of this work (e.g. the neighbor cell lists to ensure that a call that is generated in a cell that has been switched-off can be served through another cell).

In this use case a cell can be classified in two different classes:

- Class A: Candidate cell to be switched off, and;
- Class B: Cell that cannot be switched off.

a) Data acquisition and pre-processing

In this case, the components of vector $\mathbf{F(X_i)}$ correspond to the average normalized traffic of the cell during the nights (i.e. from 0h to 8h), the mornings (i.e. from 8h to 16h) and the afternoons (i.e. from 16h to 24h) for each day of the week (Monday to Sunday). This leads to a total of M=21 components that can be easily obtained by normalizing the time series $\mathbf{X_i}$ so that the traffic ranges from 0 to 1 and by averaging the time series in each of the abovementioned periods.

To assess the behavior of the classification methodology in this use case, a set of real traffic measurements for a total of 419 cells deployed by an operator in a certain geographical region has been used. For each cell i, the time series $\mathbf{X_i}$ is composed by the data traffic measurements done every 15 min, and collected during a whole week. Therefore, each time series is composed by N=672 traffic samples. The traffic in a period of 15 min is given by the average number of users in the cell with an active data session.

b) Knowledge discovery

The different classification tools discussed in *section 3.4.1.2* have been implemented by means of RapidMiner Studio Basic [21]. The different parameters have been manually adjusted to obtain good accuracy levels of the different classification tools. In particular, the SVM is configured with radial kernel type, complexity constant which sets the tolerance for misclassification C=30, kernel cache 200 MB, convergence precision 0.001, a maximum of 10^5 iterations and the loss function is defined with complexity constants equal to 1 for both positive and negative examples

and insensitivity constant equal to 0. The neural network classifier is configured with one hidden layer, 500 training cycles, learning rate 0.3, momentum 0.6 and the optimization is stopped if the training error gets below 10^{-5}. The decision tree is configured with maximal depth 20, minimal leaf size 2, confidence level 0.25, minimal size for split 4, minimal gain 0.1 and applying pruning and pre-pruning with 3 alternatives. Finally the Naive Bayes classifier is configured with Laplace correction[55], greedy estimation mode and 10 kernels.

c) Results

To illustrate the expert criteria to be learnt by the classification tool, Figure 54 plots the time series $\mathbf{X}_i$ of 4 example cells included in the training set. Two of them are classified by the expert as A and two of them are classified as B. Then, different training sets have been built including these cells together with other examples in order to train the classification tools.

First, several tests have been done to derive the accuracy of the considered classification tools as a function of the training set size S. For a given S, the accuracy is measured by executing the classification over the cells of the training set and calculating the percentage of cells that are classified in the same category that was declared by the expert in the training. The test has been applied for all 4 classification tools and training set sizes ranging from S=10 to S=200. The best accuracy is obtained by the SVM, which provides 100% accuracy in all the cases, followed by the Neural Network and Decision Tree[56], which exhibit accuracy above 98.5%. The worst behavior is obtained with the Naive Bayes classifier with a minimum accuracy of 96.4%.

After completing the training process, the classification of the 419 available cells is performed. Then, as a first result that illustrates the operation of the classification process, Figure 55 depicts the time series $\mathbf{X}_i$ of two example cells that did not belong to the training set: Cell 260, which is classified as Class A by all 4 classification tools considered, and Cell 240, which all 4 classification tools categorize as Class B. From visual inspection, and by comparing these cells with the examples given by the expert in Figure 54, it appears an adequate decision given that Cell 260 exhibits relatively long periods at night serving no traffic at all and Cell 240 has traffic during all the time periods in the week.

Figure 56 presents the total number of cells that are classified as A by each classification tool as a function of the training set size S. It is observed that, for low values of S (e.g. S=10) roughly half of the cells are classified as A and half are classified as B by all the tools. This indicates that, due to the low number of examples in the training set, the classification tools are not able to clearly distinguish the traffic patterns and the classification exhibits high randomness. Instead, when increasing the training set size S, the number of cells belonging to class A is substantially reduced for all the classifiers (e.g. for the case of the largest training set size S=200 the number of cells classified as A ranges from 46 with SVM up to 90 for the Naive Bayes case). It is worth emphasizing that the SVM exhibits a more efficient operation compared to the rest of classification tools since it is less sensitive to the value of S: as soon as the training set is $S \geq 20$, the result of the classification is very similar (i.e., there are around 50 cells classified as A).

[55] https://en.wikipedia.org/wiki/Additive_smoothing

[56] See, *for example*: https://www.quora.com/What-are-some-advantages-of-using-neural-networks-over-decision-trees

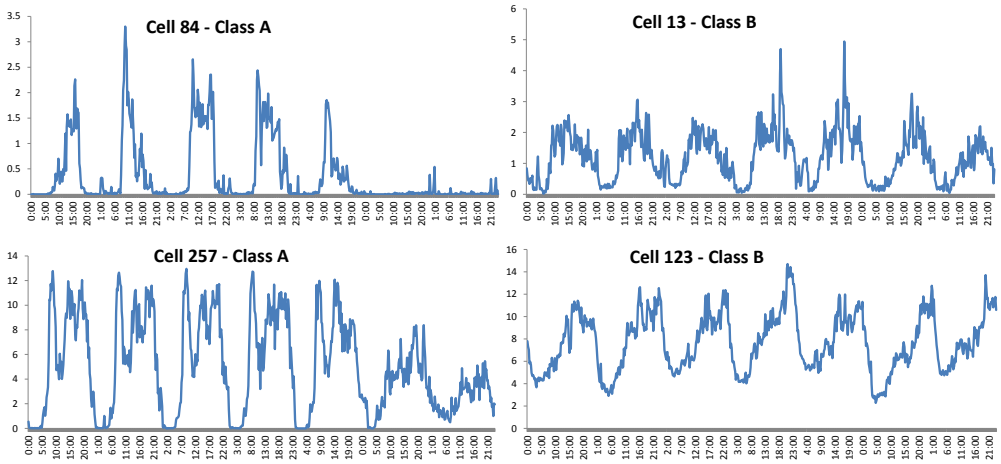


Figure 54: Examples of cells of the training set belonging to classes A and B

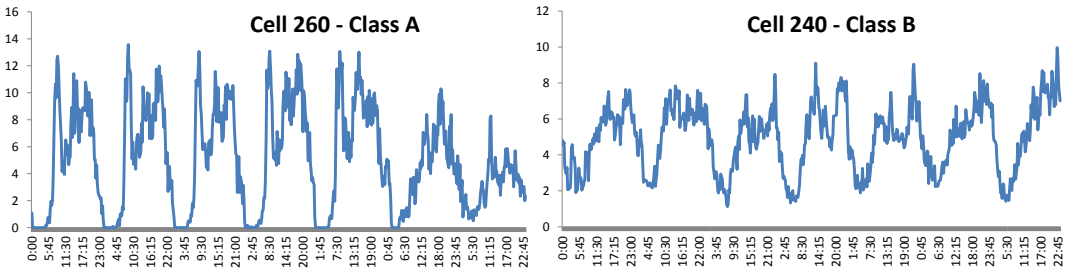


Figure 55: Examples of two cells classified as A (Cell 260) and B (Cell 240)

Table 44 compares the outcomes of the different classification tools by presenting the percentage of coincidences between every pair of tools for the case S=200. For example, the table shows that 91% of the cells (i.e. 381 out of 419 cells) have been classified equally by the SVM and the Neural Network. The table also presents the "Expert validation", which measures the percentage of coincidences with respect to the classification made by the expert. It can be observed that the largest percentages of coincidences are obtained with SVM.

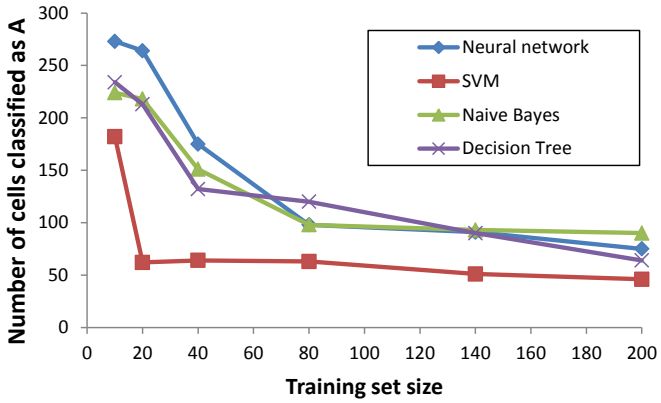


Figure 56: Examples of two cells classified as A (Cell 260) and B (Cell 240)

	SVM	Neural Network	Naive Bayes	Decision Tree	Expert validation
SVM	--	91%	88%	93%	98%
Neural Network	91%	--	87%	91%	91%
Naive Bayes	88%	87%	--	88%	87%
Decision Tree	93%	91%	88%	--	94%

Table 42: Percentage of total coincidences by every pair of classification tools with S=200

3.4.1.4 Use case 2: spectrum planning

In light of the more advanced spectrum management models envisioned for future 5G systems, the provisioning of the spectrum resources to be exploited at a given time and cell should be considered from a wider perspective. Specifically, although licensed spectrum remains operators' top priority to deliver advanced services and better user experience, other elements need to be explored as complements to meet the ultra-high capacity foreseen to be needed by future systems. These elements include the use of unlicensed spectrum considered in initiatives such as LTE-U (Unlicensed LTE) [22], [23], as well as the use of shared spectrum on a primary/secondary basis in which the operator is allowed to access a certain spectrum band owned by a different primary user, as long as certain conditions are met in order not to interfere the primary users. With all these considerations, the use case considered here intends to decide whether it is possible or not to boost the capacity of a cell by exploiting unlicensed spectrum bands. This decision will exploit the knowledge about the time evolution of the cell's traffic, in the sense that typically unlicensed spectrum could be adequate to cope with sporadic traffic increases. Then, this use case intends to classify the cells according to the following classes:

- Class A: Candidate cell to boost capacity through additional unlicensed spectrum.
- Class B: Cell that does not need capacity boost through unlicensed spectrum.

a) Data acquisition and pre-processing

This use case has been assessed considering a total of 300 cells deployed in an urban area, under the rationality that this type of scenario is where capacity boosting will be more likely needed. Besides, assuming that spectrum demands will be mainly associated to the periods of the day when there is more traffic, in this use case the components of vector $\mathbf{F(X_i)}$ correspond to the average traffic of a cell on a per hour basis, between 6h and 22h. This leads to a total of M=16 components. As a difference from the previous use case, here the traffic is not normalized, since the absolute value of the traffic is also relevant to decide whether additional unlicensed spectrum may be needed.

b) Knowledge discovery

The same classification tools as in *section 3.4.1.3* are considered here.

c) Results

Figure 57 plots the components of vector of F(Xi) for 2 cells of the training set categorized as A and B by the expert. Class A cells use to exhibit peaks of high traffic levels while class B cells exhibit lower traffic values and more homogeneity. Like in the previous use case, different training set sizes have been used to train the considered classification tools. After the training process, the 300 cells have been classified. Figure 58 depicts two example cells that were not included in the training set and that are classified as A and B by all the considered classifiers. It is observed that both cells present similar characteristics like the cells of the training set shown in Figure 57, meaning that the classification tools have been able to identify also the relevant characteristics of the time evolution in this use case.

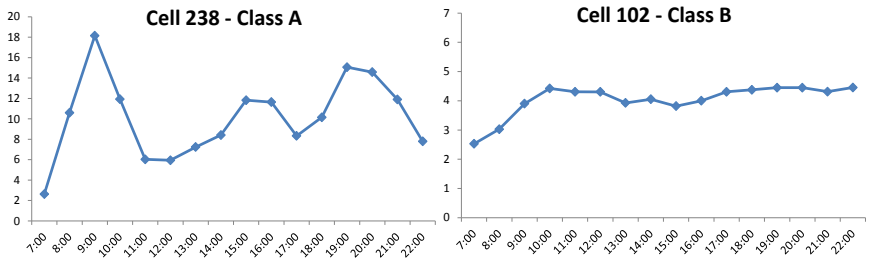


Figure 57: Examples of cells of the training set belonging to classes A and B

Figure 59 presents the number of cells classified as A by each classifier as a function of the training set size with the different classifiers. Like in the previous use case, it is observed that the SVM is able to converge more quickly than the other classifiers when the training set size is small. It is also noticed that for the case of S=140 very small differences are observed between the classifiers. This can also be corroborated in Table 45 that presents the percentage of coincidences between every pair of classifiers and with the expert validation. It can be observed that the percentages of coincidence with the expert in this use case are higher than in the previous one. This reflects that the characteristics that make a cell to be classified as A (e.g. sporadic traffic peaks) are more easily distinguishable than in the previous use case. Table 45 also shows that the best performance in terms of coincidences with the expert validation is achieved by both SVM and Neural Network classifiers.

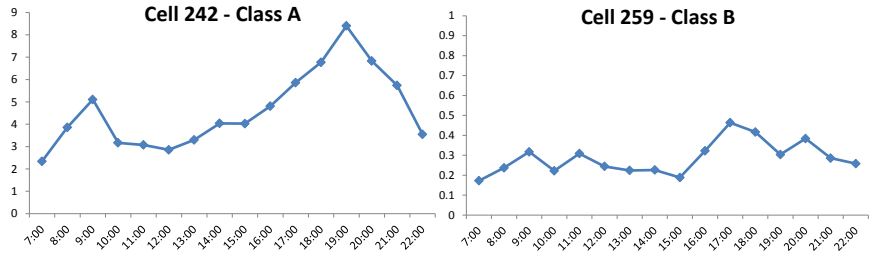


Figure 58: Examples of two cells classified as A and B

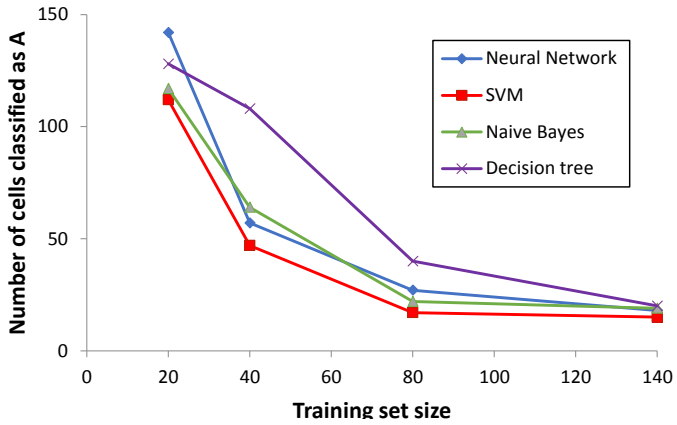


Figure 59: Number of cells classified as A as a function of the training set size

	SVM	Neural Network	Naive Bayes	Decision Tree	Expert validation
SVM	--	97%	98.7%	97.7%	99.7%

Neural Network	97%	--	98.4%	99.4%	99.7%
Naive Bayes	98.7%	98.4%	--	99%	98.4%
Decision Tree	97.7%	99.4%	99%	--	97.4%

Table 43: Percentage of total coincidences by every pair of classification tools with S=140

3.4.1.5 Conclusions

This section has focused on the application of AI and data mining concepts to support the radio access management in future cellular networks, where automatization is fundamental to cope with the huge number of cells that an operator can deploy, so manual intervention from a human expert becomes impractical. In particular, the focus has been put on extracting knowledge about the time domain traffic pattern of the cells. A general methodology for supervised classification of this traffic pattern has been presented and particularized in two applicability self-x use cases, addressing energy saving and spectrum planning. In both cases the outcomes of different classification tools are assessed, concluding that the SVM technique is in general the one that best captures in the classification process the expert knowledge provided in the examples of the training set.

3.4.2 Learning mobility patterns

In the context of the AI-*based* self-x framework presented in *section 3.3*, this section focuses on the knowledge discovery stage and, *in particular*, on automatically learning the mobility patterns of the mobile users, trying to identify if the traffic across the cells in a scenario follows specific patterns that can be characterized in terms of prototype trajectories followed by many users[24].

Different works of the literature have addressed the analysis of trajectories in different contexts such as hurricane trajectories, animal movements, public transportation, etc. Various tools have been considered, such as Self-Organizing Maps (SOMs) together with visual analysis [25], density-based clustering ([26], [27]) or Principal Component Analysis [28]. In wireless networks, [29] proposed a trajectory prediction strategy to deal with routing in mesh sensor networks. It is based on clustering similar trajectories followed by wireless nodes and using them for making predictions of other nodes. However, the concept of trajectory in [29] is defined by the set of nodes that a mobile node would associate with to send or receive data along a path, but not by the geographical locations. Instead, in our work we intend to derive a deeper knowledge about trajectories based on analyzing the geographical coordinates. In turn, [30] and [31] address the problem of classifying the trajectory followed by a mobile terminal based on a set of reference trajectories in order to optimize the handover process in LTE. However, while [30] and [31]use a simple method for building the set of reference trajectories, based on monitoring certain users with a given probability and adding their trajectories to the set, in our approach we propose the use of clustering techniques, which are more powerful for identifying the most representative trajectories.

In this context, the approach proposed in this work considers the use of clustering techniques, namely K-means and SOM, to learn the mobility patterns existing in a cellular network. These patterns are materialized in a database of prototype trajectories obtained after having observed multiple trajectories of mobile users. Different applicability areas for these patterns in the context of self-x are discussed and, *in particular*, a methodology is proposed for predicting the trajectory of a mobile user.

3.4.2.1 Mobility pattern knowledge discovery

Current cellular networks like 4G already include the capability that the User Equipments (UEs) provide geolocation information, including both geographical coordinates and altitude, as part of the radio measurement reporting processes [32]. Location information can be obtained from UEs in connected mode, who periodically transmit measurement reports to the network. Furthermore, thanks to the use of Minimization of Drive Tests (MDT) feature [33], UEs in idle mode can log measurements and transmit them later on when the UE enters in connected mode. These capabilities enable MNOs to collect large amounts of data that include valuable knowledge about the spatio-temporal traffic distribution across the cells. This work proposes a methodology to analyse this data and identify the existing mobility patterns of the UEs.

The approach for learning mobility patterns is graphically illustrated in Figure 60. It operates on a long-term basis after having observed a large amount of connected and idle mode UEs in different time periods of a certain geographical area and analyzes the collected location information from these UEs to identify the existence of prototype trajectories. As shown in Figure 60, the first step is the pre-processing, which analyzes consecutive reports for each UE and extracts the geolocation information in order to build a trajectory for this UE. A trajectory is defined here as the concatenation of N coordinates at consecutive time instants $t_1,...,t_N$. Then, assuming for simplicity two-dimensional (2D) coordinates (x,y), the trajectory for the j-th UE is given by the vector of dimension B=2N denoted as $\mathbf{r}_j=[x_j(t_1),y_j(t_1),...,x_j(t_N),y_j(t_N)]$. The result of the pre-processing task will be a total of J trajectories $\mathbf{r}_j$, j=1,...,J.

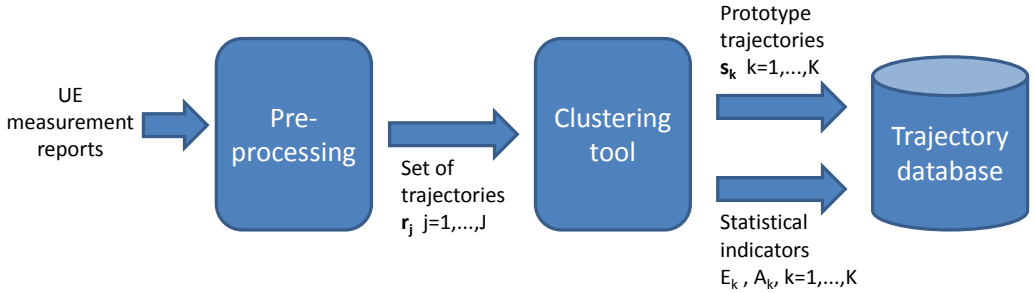


Figure 60: Procedure for Learning Mobility Patterns

The second step is the clustering, which processes the set of J trajectories by grouping them in K clusters in a way that trajectories of the same cluster are similar among them and different from the trajectories of the rest of the clusters. Two alternative clustering techniques are considered in this work:

- *K-means*: This strategy belongs to the family of partitioning methods. It groups the J input trajectories in K clusters by trying to maximize that the similarity between trajectories of the same cluster and to minimize the similarity between trajectories of different clusters, using the Euclidean distance[57] as a metric of similarity. The process can be summarized as follows (see [11] for further details): (a) The algorithm starts by selecting randomly K out of the J input trajectories. Each of these K trajectories represents an initial cluster. For each cluster k, the algorithm computes the centroid $\mathbf{s}_k$. At this initial stage, where each cluster contains only one trajectory, the centroid $\mathbf{s}_k$ equals the selected trajectory for the k-th cluster. (b) Each of the remaining J-K trajectories is assigned to the cluster to which it is the most similar, based on Euclidean distance between the trajectory and the centroid of each cluster $|\mathbf{r}_j-\mathbf{s}_k|$. Once all the J trajectories have been clustered, the new values of the centroids $\mathbf{s}_k$ are recomputed. In particular, the i-th component of $\mathbf{s}_k$ is the average of the i-th components of all the trajectories belonging to the k-th cluster. (c) Using the new values of the centroids $\mathbf{s}_k$, each of the J trajectories $\mathbf{r}_j$ is reassigned to the cluster with lowest distance $|\mathbf{r}_j-\mathbf{s}_k|$. The new centroids are recomputed and this step is iteratively repeated until convergence (i.e. until there are no

[57] https://en.wikipedia.org/wiki/Euclidean_distance

changes in the obtained clusters after two consecutive iterations). (d) At the end of the process, each cluster k=1,...,K will contain a number of input trajectories N_k and its centroid $\mathbf{s_k}$ will be the so-called prototype trajectory that is taken as a representative of all the trajectories belonging to this cluster.

- *Self-Organizing Map (SOM)*: This clustering strategy relies on a neural network model with a total of K neurons and where each neuron is characterized by a B-dimensional weight vector $\mathbf{s_k}$. The process can be summarized as follows (see [13] for details): (a) The weight vectors $\mathbf{s_k}$ are initialised. This can be done randomly or through the linear initialization method described in [13]. (b) An iterative unsupervised learning process is used to update the values of the weight vectors $\mathbf{s_k}$ of the different neurons according to the Kohonen's algorithm [13] based on the input trajectories $\mathbf{r_j}$. In essence, at iteration t the algorithm identifies, for each trajectory $\mathbf{r_j}$ the winning neuron as the one with the lowest Euclidean distance $|\mathbf{r_j} - \mathbf{s_k}|$. Then, the algorithm updates the weight vector of this winning neuron k as $\mathbf{s_k}(t+1)=s_k(t)+\alpha(t)(\mathbf{r_j} - \mathbf{s_k}(t))$ where $\alpha(t)$ is a scalar-valued adaptation gain that decreases with successive iterations. A similar update is performed for the weight vectors of the rest of neurons $k' \neq k$ but in this case the adaptation gain $\alpha(t)$ is multiplied by a neighborhood function that decreases with the distance between neurons k' and k. The process is repeated for a certain number of iterations. (c) At the end of the process, all the input trajectories that have neuron k as winning neuron form the k-th cluster. The number of trajectories in the k-th cluster is N_k, and the prototype trajectory of this cluster is the weight vector $\mathbf{s_k}$.

As shown in Figure 60, the prototype trajectories obtained as a result of the clustering will be stored in the database. In addition, two statistical indicators are also included for each cluster to assess how representative this cluster is:

- Percentage of hits ($A_k=N_k/J$): It is the percentage of input trajectories that belong to the cluster k. The prototype trajectories of clusters with a high value of A_k will be more frequent and representative of the scenario.
- Average squared Euclidean distance of the trajectories in k-th cluster (E_k): It is a metric that captures the degree of similarity between trajectories of the same cluster with respect to the prototype trajectory $\mathbf{s_k}$ of the cluster. A high value of E_k reflects a higher dispersion in the cluster, meaning that the prototype trajectory is less representative of the clustered trajectories. It is defined as:

$$E_k = \sum_{j \in \text{Cluster k}} \left|\mathbf{r_j} - \mathbf{s_k}\right|^2 \quad (1)$$

3.4.2.2 Exploitation of mobility patterns for self-x

It is envisaged that the identification of prototype trajectories as explained in previous section can have applicability for different self-x functions.

For example, prototype trajectories can be used in the context of self-planning to decide appropriate cell locations and antenna settings. For example, if there is a well identified representative trajectory, a sector of a cell site can be pointed in the direction of this trajectory. Typically, this can be the case of a cell site providing coverage over a main street. Despite one could argue that a radio engineer could easily identify such a situation and take such a common sense decision, the interest of the proposed use case remains in the fact that self-x involves automatization. That is, self-planning and self-configuration means the capability for the system to automatically identify the trajectories and propose the adequate values for the parameters of a new cell.

Similarly, the learnt mobility patterns can also have applicability in the self-optimization of several functions such as handover, load balancing or admission control. For example, by identifying the trajectory of a UE or group of UEs in relation to a known prototype trajectory it is possible to anticipate the cell that the UEs are heading to and configure these functions so as to avoid call droppings and overload situations. In the following, we focus on proposing a methodology to predict the future positions of a certain UE based on analyzing the actual locations reported by the UE in relation to the learnt prototype trajectories.

The proposed approach is illustrated in Figure 61 and is executed on an individual UE basis. The criterion to decide which specific UEs are analyzed is out of the scope of this work and it will depend on the specific self-optimization function under consideration. For example, the optimization of load balancing may predict the trajectory of UEs that demand a high bit rate in order to anticipate the arrival of these UEs to a cell and take the appropriate actions to ensure there are sufficient resources for these UEs in the cell. Similarly, it is also possible to predict the trajectory of high priority UEs to ensure that they will not experience problems in handovers, etc.

The process of Figure 61starts from the measurement reports provided by the UE whose trajectory is being predicted. First, pre-processing stage is carried out to extract the geolocation information and build the trajectory **u** that is currently being observed for this UE. The trajectory **u** is a vector of dimension C=2M composed by the concatenation of M pairs of coordinates followed by the UE at consecutive time instants $\mathbf{u}=[x(t_1),y(t_1),\ldots,x(t_M),y(t_M)]$. Without loss of generality, let us consider that the dimension of **u** is lower than the number of elements of the prototype trajectories $\mathbf{s_k}$ (i.e. C≤B). This reflects that, in case that the UE was following a prototype trajectory, the actual location of the UE is somewhere within the prototype trajectory.

The mobility prediction process of Figure 61 intends to determine the likelihood that the UE is following one of the learnt prototype trajectories. This is done by assessing the similarity between the trajectory **u** followed by the UE and the prototype trajectories $\mathbf{s_k}$ according to the Euclidean distance. Given that C≤B, all the possible portions of C consecutive elements of the vectors $\mathbf{s_k}$ (k=1,…,K) need to be considered when assessing this similarity. The α-th portion of $\mathbf{s_k}$ is then defined as the vector $[s_k(1+\alpha),\ldots, s_k(C+\alpha)]$ with α=0,…,B-C, where $s_k(i)$ denotes the i-th component of $\mathbf{s_k}$. Then, the squared Euclidean distance between the α-th portion of $\mathbf{s_k}$ and trajectory **u** is computed as:

$$d_{u,k}(\alpha) = \sum_{c=1}^{C}[u(c) - s_k(c+\alpha)]^2 \text{ with } \alpha=0,\ldots,B-C \quad (2)$$

Then, the similarity between **u** and $\mathbf{s_k}$ is computed as the minimum Euclidean distance between **u** and the possible portions of the prototype trajectory $\mathbf{s_k}$, that is:

$$m_k = \min_{\alpha} d_{u,k}(\alpha) \quad (3)$$

A low value of m_k indicates that the trajectory **u** is very similar to some portion of vector $\mathbf{s_k}$. Then, the likelihood L_k that the UE is following the prototype trajectory $\mathbf{s_k}$ is defined here as:

$$L_k = \frac{1/m_k}{\sum_{k=1}^{K}(1/m_k)} \quad (4)$$

A high value of L_k reflects that the UE is following a trajectory very similar to a portion of $\mathbf{s_k}$. Therefore, $\mathbf{s_k}$ provides information about the positions that the UE may likely follow in the future.

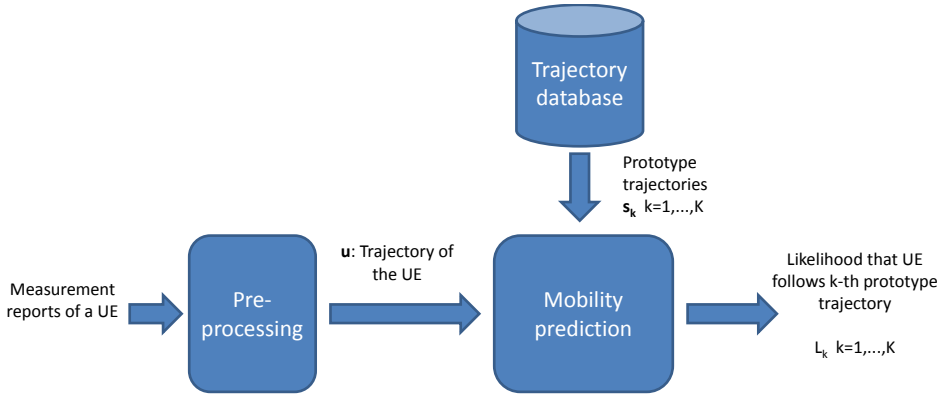


Figure 61: Exploitation of learnt patterns for predicting the trajectory of a UE

3.4.2.3 Results

This section provides some results to illustrate the performance of the proposed approach. The considered scenario is shown in Figure 62 and represents an urban area in the intersection between two main streets. The mobility of multiple UEs has been considered including a wide variety of situations, as shown Figure 62. For example, some UEs move straight along a street, others move straight and turn right, left or move back. For each kind of trajectory, 100 realizations have been generated by considering UE trajectories that are not perfectly straight but they have lateral movements simulating e.g. cars changing the lane in the road. It is assumed that the distance between two consecutive positions of the trajectory is a random value (simulating that the user speed may be variable). Moreover, 100 realizations of users that move a short distance and stop at a particular position (represented by red arrows in Figure 62) have been also generated. Finally, a group of 100 static users (represented by black dots in Figure 62) have also been placed randomly in each of the four corners of the scenario. After the preprocessing of the UE measurements, there are a total of J=2100 trajectories. Each trajectory $\mathbf{r}_j$ consists of N=40 positions.

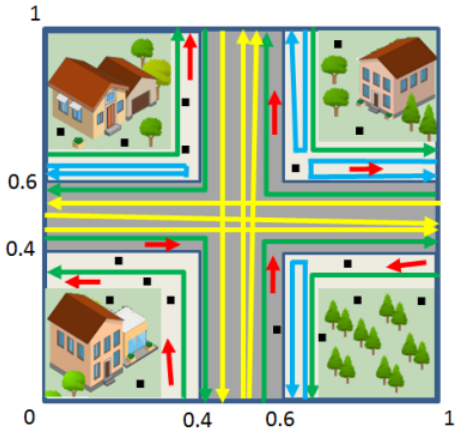


Figure 62: Illustration of the considered scenario. Distances are normalized between 0 and 1

a) Clustering process

The K-means and SOM clustering techniques have been implemented by means of RapidMiner Studio [21]. The K-means algorithm is configured with 1000 runs and a maximum of 100 iterations for each run (i.e. the process explained in section 2 is repeated 1000 times with different initial random selections, and the best result among all runs is kept). In turn, a SOM with one dimension is configured with 10000 iterations and initial adaptation rate equal to 0.1 and final

adaptation rate 0.01. The neighborhood function is defined by an initial adaptation radius of 2 and a final adaptation radius equal to 0.01.

First, the impact of the number of clusters K has been analyzed for both K-means and SOM techniques. The Davies-Bouldin index [34] is considered as a relevant metric to assess the quality of the clustering process. This index takes into account how similar are all the trajectories that belong to the same cluster and how different are the prototype trajectories of the different clusters. Low values of the Davies-Bouldin index reflect a better quality of the clustering process. Figure 63 presents the Davies-Bouldin index as a function of the number of clusters for both K-means and SOM methodologies. As shown, for the considered use case, the minimum value of the Davies-Bouldin index is observed with K=20 clusters for both methodologies. For this case, Figure 64 illustrates the prototype trajectories $\mathbf{s_k}$ obtained by the K-means methodology. The same prototypes are obtained by the SOM methodology with K=20. The red point marked in each prototype trajectory in Figure 64 indicates the initial position of the trajectory while the black point indicates its final position (e.g. the prototype trajectory 1 represents a user moving from the left to the right while the prototype trajectory 2 represents a user moving from the right to the left). Note that some shorter prototype trajectories represent users who move on a specific direction and then go back (e.g. prototype trajectory 13 represents to users who move from the left to the right in the scenario, and then go back from the right to the left). Other prototype trajectories, such as prototype 17, represent the centroid of some static users located around this area.

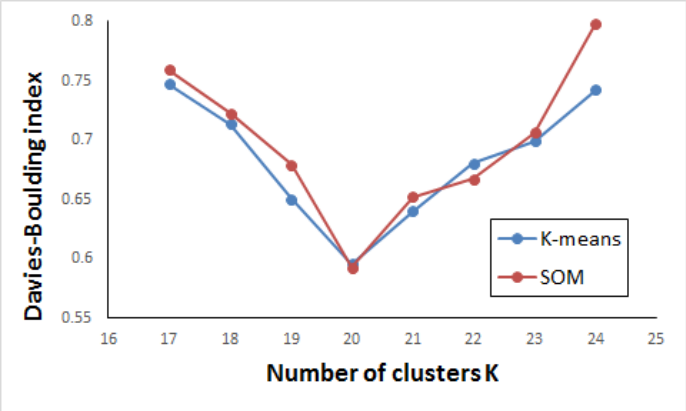


Figure 63: Davies-Bouldin index for different numbers of clusters

Figure 65a) illustrates the percentage of hits A_k for each cluster with both K-means and SOM, while Figure 65b) represents the average squared Euclidean distance E_k. All the clusters corresponding to long trajectories (i.e. clusters 1 to 12 of Figure 64) exhibit low E_k. This indicates that these trajectories are well-clustered and their corresponding prototype trajectories are good representatives of the cluster. In turn, clusters 13, 14, 15 and 16 of Figure 64 include users that move straight and go back, users that move short distances and even some static users. As a consequence, higher percentage of hits A_k and higher values of E_k are observed. Finally, clusters 17, 18, 19 and 20 of Figure 64 are formed by static users scattered around the four corners of the scenario. These are characterized by high values of E_k, meaning that some static users of these clusters may be located at a relatively high distance of the centroid. A very similar clustering is done by both K-means and SOM methodologies as shown in Figure 65a) and Figure 65b).

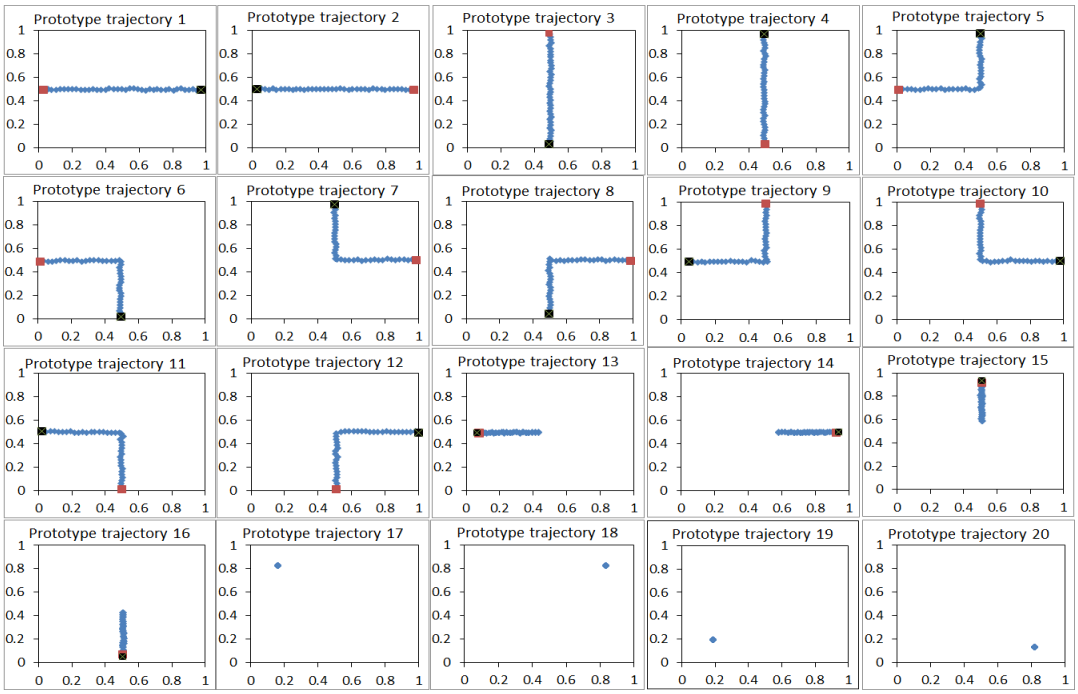
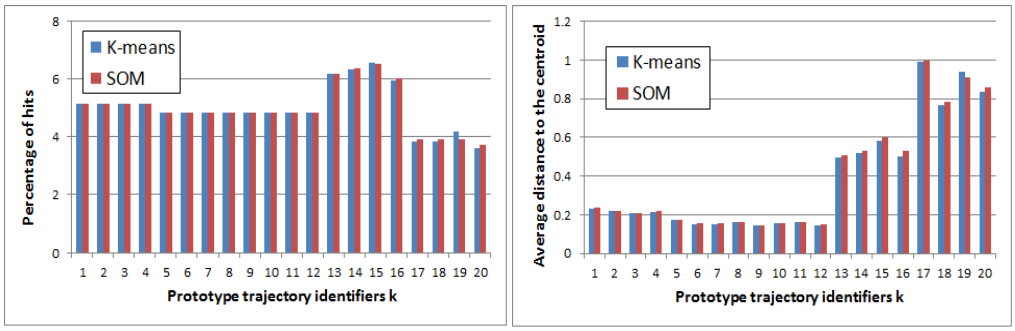


Figure 64: Prototype trajectories obtained with K-means (K=20)



(a) (b)

Figure 65: (a) Percentage of hits Ak for the different clusters; (b) Average square Euclidean distance to the centroid Ek for the different clusters

b) Mobility prediction

This section presents several examples to illustrate the behavior of proposed mobility prediction approach. Figure 66 presents the trajectories followed by four different UEs. UE A has a trajectory that consists of 20 positions representing a movement from the left of the scenario to the right. UE B has a trajectory of 20 positions moving straight and then turning in the intersection. UE C has a shorter trajectory of 10 positions while UE D is static and contains 10 samples of the same position.

Figure 66 shows the likelihood L_k that each of the four UEs is following each prototype trajectory. As shown, the likelihood that UE A is following prototype trajectory 1 is almost 100%. As seen in Figure 64, this prototype trajectory corresponds to the users that move from the left to the right. Similarly, for UE B there also is a very high likelihood that it is following prototype trajectory 5. For UE C, the likelihood L_1, L_5 and L_6 are similar, because, with the trajectory followed by UE C so far, it may correspond to either trajectories 1, 5 or 6. Finally, trajectory D is not similar to

any of the prototypes obtained in the clustering process (see Figure 64). For this reason, the prediction process provides a very low likelihood for all the clusters.

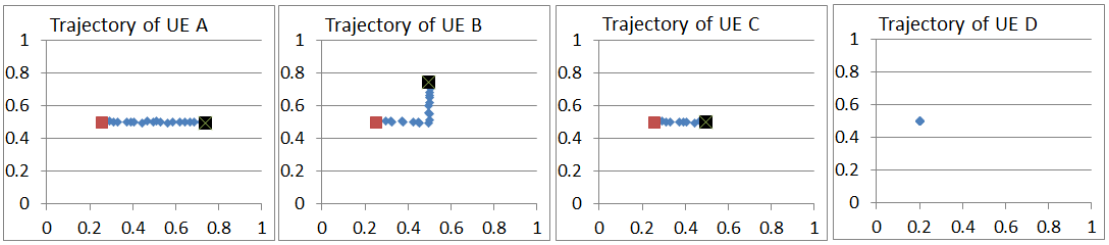


Figure 66: Example of UEs' trajectories

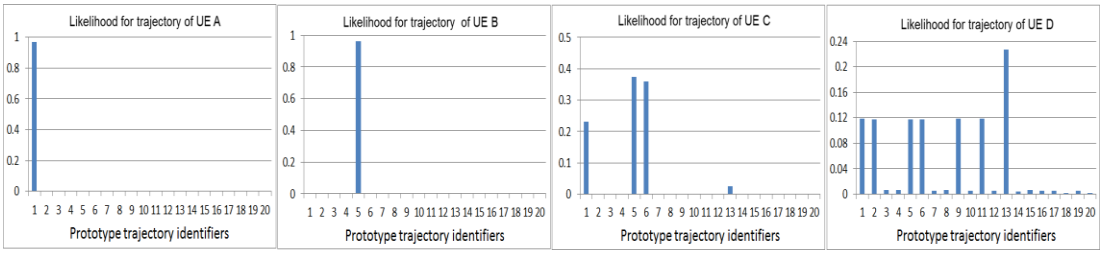


Figure 67: Likelihood L_k that the UEs are following the learnt prototype trajectories.

3.4.2.4 Conclusions

The section has proposed a methodology for learning mobility patterns in wireless networks based on clustering techniques such as K-means and SOM. Learnt trajectories present applicability in different areas, such as self-planning and self-optimization. In this respect, a strategy has been proposed for predicting the mobility of specific users based on the obtained prototype trajectories. Results reflect that both K-means and SOM techniques are able to properly identify the different trajectories existing in the considered scenario.

3.5 Initial studies on caching within the CESC Cluster

One of the key propositions of the SESAME project is to develop cost effective solutions for 5G relying on the potentials of mobile edge computing. In this regards, the possibility to deploy virtual caches within the edge cloud infrastructure provided by the Light DC constitutes an added value of the project. In specific, the SCNO can bridge together users and Over-The-Top (OTT) service providers. The SESAME infrastructure can host several virtualized functions, among which virtual caches (vCaches) are particularly appealing. Users aim to have always superior quality of experience when they consume mobile services, with particular attention to video streaming. In order to enable the users with a high quality of experience, contents must be rapidly accessible and the content itself must reach the end users in a very reliable manner. If, on the one hand, this is not always possible within the coverage of the macrocell base station due to unreliability in the received signal, within the SESAME domain a high quality of experience can be achieved with a much higher probability. Indeed, to achieve these goals, it is well understood that a high capacity and low latency backhaul connection is required to enable sufficiently fast access to contents of remote service providers (YouTube[58], Netflix[59], broadcasting services, etc.). In particular, current trends of consuming mobile services is growing so quickly [52] that in few years, mobile devices are likely to become the main platform over which these services are made available to the public. In this respect, one of the key problems is exactly the economic feasibility of deploying high capacity backhaul connections on a large scale. In this landscape mobile edge computing services leveraged over the SESAME infrastructure can contribute to circumvent this limitation.

The CESC Manager (see Figure 1) is the entity responsible for managing the CESC cluster. Particularly, the deployment of VNFs and Network Services (NSs) is under the control of the NFV Orchestrator (NFVO), which are deployed in cooperation with the Virtualised Infrastructure Manager (VIM). Service Providers (SPs) can establish specific SLA with the SCNO, which becomes the distributor of different services. The possibility to deploy caches within the cloud environment of the Light DC brings several benefits, including low latency in accessing cached contents for the users and the possibility to relieve traffic over the backhaul connection. In this respect, one possible approach which becomes feasible in SESAME consists in developing competitive schemes for SPs in order to acquire storage space in the Light DC. Therefore, different schemes can be developed to assign them storage space. In particular, this can be seen as one possible self-optimising function of the SESAME system. Since the Light DC can be seen as single storage space made of the distributed storage capability available at the servers composing the CESC cluster, the origin of the video stream is less crucial in SESAME and at the same time mobility can be handled in a much more efficient manner. When users move from a small cell coverage to that of an adjacent cell, the RAB just has to be redirected towards the right small cell area but it does not necessarily require to move the vCaches from one server to another for this reason. Indeed, vCaches can be migrated from one micro-server to another due to the need of optimizing the utilisation of storage and compute resources within the Light DC, but in general SESAME is capable of decoupling location of users and contents, at least in relatively concentrated installations (shopping malls, stadiums, etc.).

Given the limitations of the micro-servers' facility, cached content is not generally stored for an infinite period of time but rather either removed or overwritten after some time. This is the approach taken in the initial studies on competitive schemes for edge caching developed within SESAME. This approach reflects the evidence that resources of the Light DC have to be used efficiently and in addition that mobile users requesting contents might radically change even during the time of the day and popularity of contents might change accordingly. Recently, besides traditional approaches to caching (e.g. large content providers such as Akamai[60]), the context of 5G networks has provided new impulse for research on this topic as shown in [53], [54], [55], [56], [57], [58], [59] and [60].

[58] https://www.youtube.com/
[59] https://www.netflix.com/gr/
[60] https://www.akamai.com/

In [53] SC base stations (BSs) are distributed according to a Poisson Point Process[61] (PPP). Contents to be cached minimize a cost function which depends on the expected number of missed cache hits. Content delay is optimized in [54] by performing joint routing and caching. In [55] the authors model a wireless content distribution system where contents are replicated at multiple sites depending on popularity, so as to maximally create network-coding opportunities during delivery. In [56] the authors consider a device-to-device (D2D) network and derive throughput scaling laws under cash coding and spatial reuse. In [57] a model for caching contents over a D2D network is proposed. A convex optimization problem[62] is obtained and solved using a dual optimization algorithm[63]. In [58] a Stackelberg game[64] is investigated to study a caching system consisting of a content provider and multiple ISPs. In this model, the content provider leases its video to the ISPs to gain profit and ISPs aim at save the backhaul costs by placing popular video on local caches. In [59] a distributed matching algorithm based on the deferred acceptance algorithm provides association of users and SC base stations based on latency figures. In [60] optimal allocation of contents into caches is proved NP-hard[65], even when content's popularity is known, and a coded caching strategy that optimizes contents' placement based on SC association patterns is developed.

More in detail, the initial study on caching proposed in the context of SESAME is based on applying the generalized Kelly mechanism[66]. In other words, the competitive scheme of SPs purchasing caching space is modeled as a "game". Therefore, the model of a cost function is developed, which depends upon the missed cache rate per content provider. In this game, the caching price is fixed by the network provider and it admits a "Nash equilibrium"[67] (i.e. optimal strategy for all players). The work assumes the presence of C SPs, also referred to as content providers. Each provider $c \in C$ pays to store bc contents up to a maximum BC. The total storage space is denoted by N and c' total contents amount to Nc. In this context it is assumed that N≤Nc, or in other words that the available storage space is lower than, or at most equal, the possible contents to be stored. The strongest assumptions made in this work is that small cells are distributed over space according to a homogeneous Poisson Point Process (PPP) of intensity λf. Furthermore, the amount of time a content is stored is modeled whereby a random variable (r.v.) with mean 1/δ and the state of the system is represented by the amount of available storage in the Light DC. The utility function is hence defined as

$$U(b_c, b_{-c}) = g_c \exp\left(-\pi r^2 \lambda_f \frac{N}{N_c} \frac{b_c}{b_c + b_{-c} + \delta} \right), \quad (5)$$

where g_c is the rate at which customers request contents from provider *c*, b_{-c} stands for the contents cached by all other providers other than provider *c*, and *r* is the geographical distance between a UE and the serving small cell. It is worth pointing out that under this model b_c represents the strategy of provider *c*, whereas b_{-c} the strategy of the ensemble of all providers other than *c*.

The work is developed in a way to minimize the utility function $U(b_c, b_{-c})$ and using known results from game theory, the existence of the Nash equilibrium is guaranteed by the work in [61], since the set of possible strategies (i.e. b_c cached contents for each provider) represents a compact set in $\mathbb{R}^C$. Results on this work are currently under development.

---

61 For relevant information see, *for example*: https://en.wikipedia.org/wiki/Poisson_point_process
62 See, *for example*: https://en.wikipedia.org/wiki/Convex_optimization
63 See: https://en.wikipedia.org/wiki/Duality_(optimization)
64 https://en.wikipedia.org/wiki/Stackelberg_competition
65 https://en.wikipedia.org/wiki/NP-hardness
66 See: F.P. Kelly (1998). Charging and rate control for elastic traffic. Europ*ean Transactions on Telecommunications, 8,* 33–37.
67 For more explanations see: https://en.wikipedia.org/wiki/Nash_equilibrium

4 Virtualisation Aspects

4.1 VNF Orchestration

Based on the inputs from D2.2 [62] and D2.4 [64], Network Function Virtualisation (NFV) implementation empowers (V)SCNOs to flexibly manage both PNFs and VNFs (SC-VNF and service VNFs). Additionally, the SESAME Light DC has the elasticity to piece together functionality for dynamic configuration of VMs and RESTful resources, which is the virtual infrastructure manager's (VIM) responsibility. On top of that, the NFVO provides the capability to control and assemble services. In light of this, NFVO needs to address the following challenges:

- Network Service Orchestration, which includes the lifecycle management of NSs (i.e., chain of SC / service VNFs and PNFs). It empowers the SCNO to respond automatically to changes in service demand requirements and KPIs. The up/out scaling of VNFs is supported by the VNF Manager, as defined by the ETSI Management and Orchestration (MANO) architecture.
- Resource allocation on NFV infrastructure (NFVI) for VNF placement, i.e., the Light DC in the SESAME ecosystem, offers resources for the NS instantiation. NFVO via coordination with virtual VIMs handles the VM instantiation and is capable of instructing the SDN controller to form the NS.

With SESAME network service orchestration, (V)SCNOs have to simultaneously manage services coming from both physical and virtual network functions, and this will be one of the main breakthroughs of the SESAME NFVO. Note that the given level of control to each stakeholder is subject to the business model. NFVO is also able to manage the lifecycle of network services across the SESAME platform. The orchestrator is a meant to ensure network service elastic scalability and automatic update of service connectivity. This requires a stateful inventory that provides persistent and accurate data for all processes, including its native services information model, and topology. More details about NFVO's role on the SESAME platform, its logical architecture and implementation roadmap are available in Deliverable 6.1 [65].

5 Conclusions

This document has described the parts of the SESAME architecture associated with the CESCs, with a focus on the Proof of Concept demonstrator using the S1 functional split.

The demonstrator delivers the envisaged benefits of SESAME, for applying aspects of virtualisation and Mobile Edge Computing within a distributed RAN environment:

- Separation of VSCNOs,
- Management of their resource share – not just data and UEs but also CESC resources by means of virtualisation,
- Per-VSCNO KPI reports,
- SLAs that affect both what is provisioned and how it is measured,
- Service Chaining.

This document also demonstrates the robustness and flexibility of the architecture:

The triplet of SC-PNF, SC-Common-VNF and SC-VNF can accommodate different functional splits of the Physical Small Cell. These splits place different requirements on fronthaul and processing power, and result in different levels of RRM and Self-X possible in each component, but they do not change the basic architecture.

While they are not essential, additional 3GPP RAN functions could be optionally included as VNFs, including HeNB GW (per VSCNO) functionalities, X2-interface functionalities (to simplify handover procedure) and X2 GW functionalities (to simplify connectivity between CESCs in a cluster).

The document has also presented the initial studies on Self-X functions carried out in the context of the project.

First, the implications of multi-tenancy on the RRM and self-x functions that support mobility control have been discussed, considering handover, automated neighbour relations, mobility robustness optimisation and mobility load balancing functions. Then, the document has presented the application of the network slicing concept to a multi-cell RAN shared among multiple tenants. The RAN slicing problem has been analysed from a comprehensive perspective including the traffic isolation and the radio-electrical isolation, and four different RAN slicing options have been presented that differ on the RRM functions used as a support for splitting the radio resources between slices. The different approaches offer different degrees of customisation among tenants, because they establish the RRM and associated self-x functions that have to be common to all the tenants and those that can be implemented following tenant-specific policies.

The document has also presented an AI-*based* self-x framework that processes input data from very different sources and extracts, through learning-*based* classification, prediction and clustering models, relevant knowledge models used to drive the self-x decisions. Following taxonomy of self-x functions and a detailed list of AI-*based* tools that could empower the framework, a number of potential knowledge models and their applicability to the different Self-X functions have been identified. The applicability of the framework has been discussed in terms of different use cases. In particular, the use of supervised classification of the time domain traffic pattern has been presented and particularized in two applicability self-x use cases, addressing energy saving and spectrum planning. Besides, a methodology for learning mobility patterns based on clustering techniques has also been presented, which has applicability in different areas of self-planning and self-optimisation.

Finally, the document has presented some initial studies on the application of self-optimisation for content caching, exploiting the deployment of virtual caches within the cloud environment of the Light DC and considering competitive schemes for acquiring storage space for the different service providers.

6 References

[1] J. Ramiro, K. Hamied, Self-Organizing Networks. Self-planning, self-optimization and self-healing for GSM, UMTS and LTE, John Wiley & Sons, 2012.
[2] 3GPP 32.500 v12.1.0, "Self-Organizing Networks (SON); Concepts and requirements (Release 12)", September, 2014.
[3] Small Cell Forum, "SON API for small cells", Document 083.05.01, March, 2015.
[4] 3GPP TS 36.300 v13.2.0 "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN);Overall description; Stage 2 (Release 13)", December, 2015.
[5] 3GPP TS 36.420 v13.0.0, "E-UTRAN; X2 general aspects and principles (Release 13)", December, 2015.
[6] 3GPP TS 32.522 v11.7.0, "Self-Organizing Networks (SON) Policy Network Resource Model (NRM) Integration Reference Point (IRP); Information Service (IS) (Release 11)", September, 2013.
[7] Qualcomm, "LTE Small Cell SON Test Cases. Functionality and interworking", June, 2015.
[8] A. Imran, A. Zoha, A. Abu-Dayya, "Challenges in 5G: How to Empower SON with Big Data for Enabling 5G", IEEE Network, November/December, 2014, pp. 27-33.
[9] I. Chih-Lin, L. Yunlu, H. Suangfeng, W. Sihai, L. Guangyi, "On Big Data Analytics for Greener and Softer RAN", IEEE Access, vol. 3, August, 2015.
[10] R.A. Wilson, F.C. Keil, The MIT Encyclopedia of the Cognitive Sciences, MIT Press, 1999.
[11] J. Han, M. Kamber, "Data Mining Concepts and Techniques", 2nd edition, Elsevier, 2006.
[12] N.I. Sapankevych, R. Sankar, "Time Series Prediction: Using Support Vector Machines: A Survey", IEEE Computational Intelligence Magazine, May, 2009.
[13] T. Kohonen, "Essentials of the self-organizing map", Neural Networks, Vol.37, pp.52-65, 2013.
[14] A.C. Gatrell, T.C. Bailey, P.J. Diggle, B.S. Rowlingson, "Spatial point pattern analysis and its application in geographical epidemiology", Transactions of the Institute of British Geographers, Vol.21, No.1 (1996), pp.256-274.
[15] A.J. Brimicombe, "A dual approach to cluster discovery in point event data sets", Computers Environment and Urban Systems, 31(1), pp.4-18, 2007.
[16] T. Chen, et *al.*, "Software Defined Mobile Networks: Concept, Survey, and Research Directions", IEEE Comm. Mag., Nov. 2015, pp.126-133.
[17] ETSI GS NFV-MAN 001 (V1.1.1) "Network Function Virtualisation (NFV); Management and Orchestration", December, 2014.
[18] O. Sallent, J. Pérez-Romero, R. Ferrús, R. Agustí, "Small Cell as a Service: From Capacity Provisioning to Full Customisation", Proc. EuCNC, 2016.
[19] J. Pérez-Romero, J. Sánchez-González, O. Sallent, R. Agustí, "On Learning and Exploiting Time Domain Traffic Patterns in Cellular Radio Access Networks", 12th International Conference on Machine Learning and Data Mining (MLDM), New York, July, 2016.
[20] GSMA, Mobile's Green Manifesto, 2nd Edition, June 2012.
[21] RapidMiner Studio, http://www.rapidminer.com
[22] 3GPP workshop on LTE in unlicensed spectrum, Sophia Antipolis, France, June 13, 2014. http://www.3gpp.org/ftp/workshop/2014-06-13_LTE-U/
[23] A. Al-Dulaimi, S. Al-Rubaye, N. Quiang, E. Sousa, "5G Communications Race: Pursuit of More Capacity Triggers LTE in Unlincensed Band", IEEE Vehicular Technology Magazine, Vol.10, Issue 1, pp.43-51, February 2015.
[24] J. Sánchez-González, J. Pérez-Romero, R. Agustí, O. Sallent, "On Learning Mobility Patterns in Cellular Networks", 5G-PINE workshop at the 12th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Thessaloniki, Greece, September, 2016.

[25] T. Schreck et *al.*, "Visual Cluster Analysis of Trajectory Data With Interactive Kohonen Maps", IEEE Symp. on Visual Anal. Science and Techn., October, 2008, Columbus, USA.
[26] J-G. Lee, J. Han, K-Y. Whang, "Trajectory Clustering: A Partition-and-Group Framework", SIGMOD, 2007, China.
[27] G. Andrienko, et *al.*, "Interactive Visual Clustering of Large Collections of Trajectories", IEEE Symp. on Visual Anal. Science and Techn., October, 2009, Atlantic City, USA,
[28] E. Masciari, "A Complete Framework for Clustering Trajectories", 21st IEEE International Conference on Tools with Artificial Intelligence, 2009.
[29] H. J. Lee et *al.*, "Data Stashing: Energy-efficient Information Delivery to Mobile Sinks through Trajectory Prediction", ACM/IEEE IPSN Conference, 2010.
[30] B. Sas, K. Spaey, C. Blondia, "Classifying Users based on their Mobility Behavior in LTE networks", 10th Int. Conf. on Wireless and Mob. Comms. (ICWMC), 2014.
[31] B. Sas, K. Spaey, C. Blondia, "A SON function for Steering Users in Multi-Layer LTE Networks based on their mobility behavior", VTC Spring Conference, 2015.
[32] 3GPP TS 36.331 v12.7.0, "Radio Resource Control (RRC); Protocol Specification (Release 12)", September, 2015.
[33] W.A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, B. Sébire, "Minimization of Drive Tests Solution in 3GPP", IEEE Communications Magazine, June, 2012.
[34] D.L. Davies, D.W. Bouldin, "A Cluster Separation Measure", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol: PAMI-1, issue 2, pp.224–227, April 1979.
[35] 3GPP TR22.891 v2.0.0 "Feasibility Study on New Services and Markets Technology Enablers; Stage 1 (Release 14)", February, 2016.
[36] V. Del Piccolo, A. Amamou, K. Haddadou, G. Pujolle, "A Survey of network isolation solutions for multi-tenant data centers", IEEE Communications Surveys and Tutorials, IEEE Early Access Articles, 2016.
[37] C. Liang, F.R. Yu, "Wireless Network Virtualization: A survey, some research issues and challenges", IEEE Communications Surveys and Tutorials, Vol.17, No.1, 1st Quarter, 2015.
[38] R. Kokku, R. Mahindra, H. Zhang, S. Rangarajan, "NVS: A substrate for Virtualizing Wireless Resources in Cellular Networks", IEEE/ACM Transactions on Networking, Vol.20, No.5, October, 2012.
[39] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, "Radio Access Network Virtualization for Future Mobile Carrier Networks", IEEE Communications Magazine, July, 2013.
[40] K. Spaey et *al.*,"SON functions for Multi-layer LTE and Multi-RAT networks (Final Results)", Deliverable D4.2 of the SEMAFOUR project, August, 2014, available at http://www.fp7-semafour.eu/en/public-deliverables/
[41] T.D. Novlan, R.K. Ganti, A. Ghosh, J.G. Andrews, "Analytical Evaluation of Fractional Frequency Reuse for OFDMA Cellular Networks", IEEE Transactions on Wireless Communications, Vol.10, No.12, December, 2011, pp. 4294-4304.
[42] 3GPP TR 36.942 v12.0.0, "Radio Frequency (RF) system scenarios", September, 2014.
[43] 3GPP TS 32.425 v10.7.0, "Performance Management (PM); Performance measurements Evolved Universal Terrestrial Radio Access Network (E-UTRAN), June 2012".
3GPP TS 36.314 v9.1.0, "Evolved universal Terrestrial Radio Access. (E-UTRA); Layer 2 Measurements", June 2010.
[44] S. Hernan, S. Lambert, and T. Ostwald, "Uncover Security Design Flaws using The STRIDE Approach" msdn. microsoft. com," Design, no.1, pp.1–8, 2006.
[45] D. Kreutz, F.M.V. Ramos, and P. Verissimo, "Towards secure and dependable software-defined networks," *Pro*c. Second ACM SIGCOMM Work. Hot Top. Softw. Defin. Netw. - HotSDN '13, p.55, 2013.
[46] D. Liu et *al.*, "User Association in 5G Networks: A Survey and an Outlook", IEEE communications Surveys & Tutorials, Vol.18, No.2, Second Quarter 2016.

[47] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energyefficient context-aware user association for outdoor small cell heterogeneous networks," in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2014, pp. 1614–1619.

[48] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2012, pp.2457–2461.

[49] H. Zhou, S. Mao, and P. Agrawal, "Approximation algorithms for cell association and scheduling in femtocell networks," IEEE Trans. Emerging Topics Comput., vol.3, no 3, pp.432–443, Sep. 2015.

[50] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," IEEE J. Sel. Areas Commun., vol.28, no.9, pp.1479–1489, Dec. 2010.

[51] [50] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," IEEE J. Sel. Areas Commun., vol.28, no.9, pp.1479–1489, Dec. 2010.

[52] "Ericsson Mobility Report: On the Edge of a Networked Society," Whitepaper, Ericsson, June 2014.

[53] B.N. Bharath, K.G. Nagananda, and H.V. Poor, "A Learning-Based Approach to Caching in Hetero-genous Small Cell Networks," CoRR, abs/1508.03517, 2015.

[54] M. Dehghan, A. Seetharam, B. Jang, T. He, T. Salonidis, J. Kurose, D. Towsley, and R.K. Sitaraman, "On the Complexity of Optimal Routing and Content Caching in Heterogeneous Networks," in Proc. of IEEE INFOCOM 2015, pp.936–944, April 26 -May 1, 2015.

[55] J. Hachem, N. Karamchandani, S. Diggavi, "Multi-Level coded caching," in Proc. of IEEE INFOCOM, pp.756-764, April 26th - June 1st 2015.

[56] M. Ji, G. Caire, A.F. Molisch, "Fundamental Limits of Distributed Caching in D2D Wireless Networks," in Proc. of IEEE ITW, pp.1-5. IEEE, 2013.

[57] H.J. Kang and C.G. Kang, "Mobile Device-to-Device (D2D) Content Delivery Networking: A Design and Optimization Framework," Journal of Communications and Networks, vol.16(5), pp.568–577, Oct. 2014.

[58] J. Li and W. Chen, "Efficient Video Pricing and Caching in Heterogeneous Networks," IEEE Trans. On Vehicular Technology, 2015.

[59] F. Pantisano, M. Bennis, W. Saad, M. Debbah, "Cache-Aware User Association in Backhaul-Constrained Small Cell Networks," in Proc. of IEEE WiOPT, pp.37-42, May 2014.

[60] A. Sengupta, S. Amuru, R. Tandon, R. Buehrer, T. Clancy, "Learning Distributed Caching Strategies in Small Cell Networks," in Proc. of IEEE ISWCS, pp.917–921, Aug. 2014.

[61] J.B. Rosen, "Existence and Uniqueness of Equilibrium Points for Concave N-person Games," Econometrica, vol.33(3), July 1965.

[62] H2020 5G-PPP 671596 SESAME, Deliverable D2.2, "Overall System Architecture and Interfaces", March 2016.

[63] H2020 5G-PPP 671596 SESAME, Deliverable D2.3, "Specification of the CESC components –First Iteration", March 2016.

[64] H2020 5G-PPP 671596 SESAME, Deliverable D2.4, "Specification of VIM and CESCM functions".

[65] H2020 5G-PPP 671596 SESAME, Deliverable D6.1, "Orchestrator Components, Interfaces Design and Specifications".

[66] 3GPP TS 32.592: Telecommunication Management; HeNB Operation, Administration, Maintenance and Provisioning; Information Model for Type 1 interface HeNB to HeNB Management Systems.

[67] 3GPP 32.411: Technical Specification Group Services and System Aspects; Telecommunication management; Performance Management (PM) Integration Reference Point (IRP): Requirements.

[68] 3GPP 36.413: Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP).

[69] 3GPP 32.342: Technical Specification Group Services and System Aspects; Telecommunication management; File Transfer (FT) Integration Reference Point (IRP); Information Service (IS).

[70] 3GPP 32.435: Technical Specification Group Services and System Aspects; Telecommunication management; Performance measurement; eXtensible Markup Language (XML) file format definition.

[71] 3GPP 32.346: Technical Specification Group Services and System Aspects; Telecommunication management; File Transfer (FT) Integration Reference Point (IRP): Solution Set (SS) definitions.

[72] 3GPP 23.003: Technical Specification Group Core Network and Terminals; Numbering, addressing and identification.

[73] 3GPP 32.111-1: Technical Specification Group Services and System Aspects; Telecommunication management; Fault Management; Part 1: 3G fault management requirements.

[74] 3GPP 32.111-2: Technical Specification Group Services and System Aspects; Telecommunication management; Fault Management; Part 2: Alarm Integration Reference Point (IRP): Information Service (IS).

[75] 3GPP 32.111-3: Technical Specification Group Services and System Aspects; Telecommunication management; Fault Management; Part 3: Alarm Integration Reference Point (IRP): Common Object Request Broker Architecture (CORBA) Solution Set (SS).

[76] 3GPP 32.111-5: Technical Specification Group Services and System Aspects; Telecommunication management; Alarm Integration Reference Point (IRP): eXtensible Markup Language (XML) definitions.

[77] 3GPP 32.111-6: Technical Specification Group Services and System Aspects; Telecommunication management; Fault Management; Part 6: Alarm Integration Reference Point (IRP): Solution Set (SS) definitions.

[78] 3GPP 32.111-7: Technical Specification Group Services and System Aspects; Telecommunication management; Alarm Integration Reference Point (IRP): SOAP Solution Set (SS).

[79] TR-069: CPE WAN Management Protocol, Issue: 1 Amendment 5, November 2013, CWMP Version: 1.4.

[80] TR-196: Femto Access Point Service Data Model, Issue: 2, November 2011.

[81] ITU-T Recommendation X.730: Information Technology - Open Systems Interconnection - Systems Management: Object Management Function.

[82] ITU-T Recommendation X.731: Information Technology - Open Systems Interconnection - Systems Management: State Management Function.

[83] ITU-T Recommendation X.732: Information Technology - Open Systems Interconnection - Systems Management: Attributes for representing Relationships.

[84] ITU-T Recommendation X.733: Information Technology - Open Systems Interconnection - Systems Management: Alarm Reporting Function.

[85] 3GPP TS 23.401 v13.6.1 “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access”, May, 2016.

[86] SCF159.07.2.01: Small Cell Forum Document 159.07.02, “Small cell virtualization functional splits and use cases”, January 2016.

7 Appendix A – Supported PM Counters

The PM counters described in this appendix are supported by the SC-PNF.

7.1 Supported 3GPP 36.314 Counters

Reference	Counter
4.1.1.1	Total PRB Usage
4.1.1.2	PRB usage per traffic class
4.1.2	Received Random Access Preambles
4.1.3.1	Number of Active UEs in the DL per QCI
4.1.3.2	Number of Active UEs in the UL per QCI
4.1.4.1	Packet Delay in the DL per QCI
4.1.5.1	Packet Discard Rate in the DL per QCI
4.1.5.2	Packet Uu Loss Rate in the DL per QCI
4.1.5.3	Packet Loss Rate in the UL per QCI
4.1.6.1	Scheduled IP Throughput in DL
4.1.6.2	Scheduled IP Throughput in UL

Table 44: Supported PM Counters – 3GPP 36.314

7.2 Supported 3GPP 32.425 Counters

Reference	Counter
4.1.1.1	Attempted RRC connection establishments
4.1.1.2	Successful RRC connection establishments
4.1.1.3	Failed RRC connection establishments
4.1.1.4	Failed RRC connection establishment per failure cause
4.1.2.1	Attempted RRC connection re-establishments
4.1.2.2	Successful RRC connection re-establishments
4.1.2.3	Failed RRC connection re-establishments
4.1.3.1	Mean number of RRC Connections
4.1.3.2	Maximum number of RRC Connections
4.1.4.1	Mean RRC connection setup time
4.1.4.2	Maximum RRC connection setup time
4.1.5.1	Number of UE CONTEXT Release Request initiated by eNodeB/RN
4.1.5.2	Successful UE CONTEXT Release
4.2.1.1	Number of initial E-RABs attempted to setup
4.2.1.2	Number of initial E-RABs successfully established
4.2.1.3	Number of initial E-RABs failed to setup
4.2.1.4	Number of additional E-RABs attempted to setup
4.2.1.5	Number of additional E-RABs successfully established
4.2.1.6	Number of additional E-RABs failed to setup
4.2.1.7	Mean E-RAB Setup time
4.2.1.8	Maximum E-RAB Setup time
4.2.2.1	Number of E-RABs requested to release initiated by eNodeB/RN per QCI
4.2.2.2	Number of E-RABs requested to release initiated by eNodeB per cause
4.2.2.3	Number of E-RABs attempted to release
4.2.2.4	Number of E-RAB successfully released
4.2.2.5	Number of E-RAB failed to release
4.2.2.6	Number of released active E-RABs
4.2.3.1	Number of E-RABs attempted to modify the QoS parameter
4.2.3.2	Number of E-RABs successfully modified the QoS parameter
4.2.3.3	Number of E-RABs failed to modify the QoS parameter
4.2.4.1	In-session activity time for UE
4.2.4.2	In-session activity time for E-RABs
4.2.5.1	Average Number of simultaneous E-RABs.
4.2.5.2	Maximum Number of simultaneous E-RABs.

Reference	Counter
4.3.1.2.1	Attempted outgoing inter-eNB handover preparations
4.3.1.2.2	Attempted outgoing inter-eNB handover executions per handover cause
4.3.1.2.3	Successful outgoing inter-eNB handover executions per handover cause
4.3.1.3.1	Attempted outgoing handovers per handover cause
4.3.1.3.2	Successful outgoing handovers per handover cause
4.3.1.4.1	Attempted outgoing intra-frequency handovers
4.3.1.4.2	Successful outgoing intra-frequency handovers
4.3.1.4.3	Attempted outgoing inter-frequency handovers - gap-assisted measurement
4.3.1.4.4	Successful outgoing inter-frequency handovers - gap-assisted measurement
4.3.1.4.5	Attempted outgoing inter-frequency handovers - non gap-assisted measurement
4.3.1.4.6	Successful outgoing inter-frequency handovers - non gap-assisted measurement
4.3.1.5.1	Attempted outgoing handovers with DRX
4.3.1.5.2	Successful outgoing handovers with DRX
4.3.1.5.3	Attempted outgoing handovers non-DRX
4.3.1.5.4	Successful outgoing handovers non-DRX
4.3.2.1.1	Attempted outgoing inter-RAT handovers per handover cause
4.3.2.1.2	Successful outgoing inter-RAT handovers per handover cause
4.4.1.1	Average DL cell PDCP SDU bit-rate
4.4.1.2	Average UL cell PDCP SDU bit-rate
4.4.1.3	Maximum DL cell PDCP SDU bit-rate
4.4.1.4	Maximum UL cell PDCP SDU bit-rate
4.4.1.5	Average DL cell control plane PDCP SDU bit-rate
4.4.1.6	Average UL cell control plane PDCP SDU bit-rate
4.4.2.1	Average number of active UEs on the DL
4.4.2.2	Average number of active UEs on the UL
4.4.3.1	Average DL PDCP SDU delay
4.4.3.2	DL PDCP SDU drop rate
4.4.4.1	DL PDCP SDU air interface loss rate
4.4.4.2	UL PDCP SDU loss rate
4.4.5.1	IP Latency in DL, E-RAB level
4.4.6.1	IP Throughput in DL
4.4.6.2	IP Throughput in UL
4.5.1	DL PRB Usage for traffic

Reference	Counter
4.5.2	UL PRB Usage for traffic
4.5.3	DL Total PRB Usage
4.5.4	UL Total PRB Usage
4.5.5.1	Mean number of RACH preambles received
4.5.5.2	Distribution of RACH preambles sent
4.5.5.3	Distribution of RACH access delay
4.5.5.4	Percentage of contentious RACH attempts
4.5.5.5	Number of UE RACH reports received
4.5.5.6	Percentage of time when all dedicated RACH preambles are used
4.5.6	Cell Unavailable Time
4.5.7.1	Total Number of DL TBs
4.5.7.2	Error Number of DL TBs
4.5.7.3	Total Number of UL TBs
4.5.7.4	Error Number of UL TBs
4.5.8.1	Maximum carrier transmit power
4.5.8.2	Mean carrier transmit power
4.5.9.1	DL PRB full utilisation
4.5.9.2	UL PRB full utilisation
4.6.1.1	Attempted UE-associated logical S1-connection establishment from eNB to MME
4.6.1.2	Successful UE-associated logical S1-connection establishment from eNB to MME
4.7.1.1	Number of paging records discarded at the eNodeB/RN
4.7.1.2	Number of paging records received at the eNodeB/RN
4.8.1.1	Mean processor usage
4.8.1.2	Peak processor usage
4.9.1	Number of incoming IRAT mobility events per LA
4.10.1.1	Wideband CQI distribution
4.10.1.2	Average sub-band CQI
4.10.2	Timing Advance Distribution

Table 45: Supported PM Counters – 3GPP 32.425

7.3 Extended PM Counters

The following 3GPP PM counters are extended variants of those defined in 3GPP 32.425. They each an additional PLMN ID field so that they report separate, per-PLMN counts. Note that these counters are in addition to the standard, aggregate versions.

Reference	Counter
4.1.1.2	Successful RRC connection establishments *per PLMN*
4.1.2.1	Attempted RRC connection re-establishments *per PLMN*
4.1.2.3	Failed RRC connection re-establishments *per PLMN*
4.1.2.2	Successful RRC connection re-establishments *per PLMN*
4.1.4.2	Maximum RRC connection setup time *per PLMN*
4.1.4.1	Mean RRC connection setup time *per PLMN*
4.1.3.2	Maximum number of RRC Connections *per PLMN*
4.1.3.1	Mean number of RRC Connections *per PLMN*

Table 46: Extended 3GPP PM Counters

The following, per PLMN, user plane usage counters are also defined by SESAME:

Counter
GTP-U Octets Received per PLMN
GTP-U Octets Transmitted per PLMN
GTP-U Packets Received per-PLMN without Sequence Number
GTP-U Packets Received per-PLMN with Sequence Number
GTP-U Packets Transmitted per PLMN

Table 47: Per PLMN User Plane Usage Counters

8 Appendix B – Supported CM IRP Methods

This appendix details the list of methods supported by the northbound Configuration Management Integration Reference Point (IRP) provided by SESAME.

8.1 Standard 3GPP IRPs

The following methods defined by 3GPP are supported by SESAME. For details of their parameters and behaviour, see the associated 3GPP specification.

IRP	Method	Comment
Generic CM IRP (3GPP 32.316)	getIRPVersion	Returns a list of IRPs and versions supported by the solution set
Basic CM IRP (3GPP 32.606)	createMO	Creates a new managed object
	deleteMO	Deletes an existing managed object
	getMOAttributes	Returns a list of attributes and values
	setMOAttributes	Allows the values of certain attributes of a managed object to be changed
Kernel CM IRP (3GPP 32.662)	getIRPVersion	As above
Notification Subscription IRP (3GPP 32.306)	subscribe	Allows the subscription to notification events (such as create, delete and value change) relating to a specific managed object or sub-tree of managed objects.
	unsubscribe	Requests ceasing of a notification subscription.

Table 48: Standard 3GPP CM IRPs

8.2 SESAME Specific IRP

The following IRP methods are defined by SESAME in order to assist VSCNOs with the provisioning of virtual small cells.

Method	Comment
getVirtualSmallCells	Returns a list of distinguished names identifying the virtual small cells owned by the VSCNO invoking the method. Each DN uniquely identifies a Virtual Small Cell managed object. See section 2.1.9.3.1.4. The filter parameter may be used to limit the scope of the returned list to a subset of cells (for example within a specific geographic area).
provisionVirtualSmallCell	Creates a new Virtual Small Cell managed object and assigns it to a host SC-PNF.
removeVirtualSmallCell	Deletes and decommissions a previously provisioned Virtual Small Cell managed object.
retreiveVirtualSmallCell	Retrieves the configuration details of a virtual small cell identified by its Global Cell Id parameter. Allows VSCNOs to refer to the cell without needing to know its DN.
modifyVirtualSmallCell	Allows the modification of the parameters of a virtual small cell identified by its Global Cell Id parameter. Allows VSCNOs to modify the cell without needing to know its DN.

Table 49: SESAME IRP Methods

9 Appendix C – Supported FM IRP Methods and Parameters

The FM IRP supports the following methods defined in 3GPP 32-111-2 [74] :

Method	Description
getAlarmList	Provides alarm information for a single managed object, a set of managed objects or a sub-tree of managed objects, based on the baseObjectInstance and filter parameters of the method.
acknowledgeAlarms	Allows one or more alarms to be acknowledged.
getAlarmCount	Provides a per-severity count of alarm information. The filter parameter allows subsets such as active, cleared and acknowledged to be specified.
unacknowledgeAlarms	Allows the acknowledged status to be removed from one or more alarms.
clearAlarms	Allows one or more alarms to be cleared.
setComment	Allows a comment to be recorded in one or more alarms.

Table 50: Methods supported by the FM IPR

If supported by the Prototype implementation, the Notification IRP provides the following methods defined in 3GPP 32-111-2:

IRP	Method	Description
AlarmIRPNotifications_1	notifiyNewAlarm	Reports that new alarm information has been added to the alarm list.
	notifyAckStateChange	Reports that the acknowledged state of an alarm has changed.
	notifyClearedAlarm	Reports that a previously active alarm has cleared.
AlarmIRPNotifications_2	notifyChangedAlarm	Reports that the severity of an active alarm has changed.
AlarmIRPNotifications_3	notifyComments	Reports that a comment has been recorded against an alarm.

Table 51: Methods provided by the Notification IPR

The following alarm attributes, defined in 3GPP 32-111-2, are supported on the northbound FM interface

Attribute name	**3GPP 32.111 Support Qualifier**
alarmId	M
notificationId	M
alarmRaisedTime	M
alarmClearedTime	M
alarmChangedTime	O
eventType	M
probableCause	M
perceivedSeverity	M
specificProblem	O
additionalText	O
ackTime	M
ackUserId	M
ackState	M

Table 52: Attributes supported by the northbound FM interface

10 Appendix D – Supported File Transfer IRP Methods

Method	Description
listAvailableFiles	This operation allows a northbound client (such as a VSCNO) to list all or specified available management data files stored by the EMS. The beginTime and endTime parameters specify the time window over which results are returned. The return parameters include status (success or failure) and a *fileInfoList* as described below.
fileDownloadIndication	*The method uses the notification IRP to inform a subscribed northbound client of completion of a file exchange procedure.*
notifyFileReady	*After the PM data files have been prepared and are ready in the EMS, the EMS emits this notification to a subscribed northbound client(s) to notify the availability of the file(s).*
notifyFilePreparationError	*The subscribed northbound clients are notified regarding the occurrence of an error during the preparation of the file.*

Table 53: Supported File Transfer IRP Methods

Note: only the *listAvailableFiles* method is available in the prototype implementation. The need to support other methods is FFS.

The parameters of the *fileInfoList* result returned by the *listAvailableFiles* method are as follows:

Parameter	Description
fileLocation	A URL, minus the protocol portion, that specifies a location on the EMS from which the file can be retrieved using either FTP or SFTP (preferred).
fileSize	The size of the file in bytes.
fileReadyTime	The time that the file was uploaded to the EMS.
fileExpirationTime	A time, after which, the file will be no longer available. The EMS retains PM report files for a configurable period and purges that after the allowed time has elapsed.
fileCompression	For further study.
fileFormat	Always set to "http://www.3gpp.org/ftp/specs/archive/32_series/32.435#measCollec pm_report.xsd" in the prototype implementation.

Table 54: Parameters'description of the *fileInfoList*

11 Appendix E – Example Alarm List XML File

The following example file illustrates the XML encoding of an alarm report file.

```
<?xml version="1.0" encoding="UTF-8"?>
<AlarmList>
  <ElementIdentity>TBA</ElementIdentity>

  <Alarm>
    <alarmId>0001</alarmId>
    <notificationId>1</notificationId>
    <alarmRaisedTime>YYYY-MM-DDTHH:MM:SS</alarmRaisedTime>
    <alarmChangedTime>YYYY-MM-DDTHH:MM:SS</alarmChangedTime>
    <alarmClearedTime>YYYY-MM-DDTHH:MM:SS</alarmClearedTime>
    <probabableCause>High Temperature</probabableCause>

    <perceivedSeverity>MAJOR</perceivedSeverity>
    <alarmType>ENVIRONMENTAL ALARM</alarmType>
    <specificProblem>Over Temperature</specificProblem>
    <proposedRepairActions>Check fans and Increase ventila-
       tion</proposedRepairActions>
    <additionalText>Current temperature 74 degrees C</additionalText>
    <additionalInformation></additionalInformation >
  </Alarm>

  <Alarm>
    <alarmId>0002</alarmId>
    <notificationId>3</notificationId>
    <alarmRaisedTime>YYYY-MM-DDTHH:MM:SS</alarmRaisedTime>
    <alarmChangedTime>YYYY-MM-DDTHH:MM:SS</alarmChangedTime>
    <alarmClearedTime>YYYY-MM-DDTHH:MM:SS</alarmClearedTime>
    <probabableCause>Link Failure</probabableCause>
    <perceivedSeverity>CRITICAL</perceivedSeverity>
    <alarmType>COMMUNICATIONS ALARM</alarmType>
    <specificProblem>S1 Link Failure</specificProblem>
    <proposedRepairActions>Check connectors and endpoint address-
       es</proposedRepairActions>
    <additionalText></additionalText>
    <additionalInformation></additionalInformation >
  </Alarm>
</AlarmList>
```