iJOIN

INFSO-ICT-317941

INFSO-ICT-317941 iJOIN

D4.1

Report on SotA and requirements for network-layer algorithms and network operation and management

Editor:	Yingli Sheng, University of Surrey
Deliverable nature:	Public
Suggested readers:	iJOIN GA
Due date:	October 31st, 2013
Delivery date:	November 4th, 2013
Version:	1.0
Total number of pages:	93
Reviewed by:	GA members
Keywords:	NET/MM State-of-the-art, DMM, SDN, Backhaul NET, RANaaS, iJOIN
Resources consumed	42.42 PM

Abstract

This deliverable presents an overview of the researches in Work Package 4 (WP4) during the first twelve months of iJOIN project. The document provides a comprehensive state-of-art review of the main topics to be explored in WP4 scope. The set of candidate technologies to be investigated is also presented with assumptions, technical requirements, as well as first details of solution related to the work package 4 "Network-layer solutions and system and operation management". Finally, the WP4 functional architecture is introduced, including the interactions between candidate technologies. This comprises the description of novel algorithms and includes a high-level description of interactions between backhaul, wireless access and terminal.

List of authors

Company	Author
NEC	Peter Rost (Peter.Rost@neclab.eu)
	Dani Camps (Daniel.CampsMur@neclab.eu)
TUD	Vinay Suryaprakash (vinay.suryaprakash@ifn.et.tu-dresden.de)
	Jens Bartelt (jens.bartelt@ifn.et.tu-dresden.de)
UC3M	Carlos J. Bernardos (cjbc@it.uc3m.es)
TID	Ignacio Berberana (ibfm@tid.es)
IMC	Umer Salim (umer.salim@intel.com)
	Nikolaos Sapountzis (Nikolaos.Sapountzis@eurecom.fr)
UNIS	Yingli Sheng (y.sheng@surrey.ac.uk)

History

Modified by	Date	Version	Comments
Yingli Sheng	November 4th, 2013	1.0	Final version

Table of Contents

List of authors ... 2
History ... 3
Table of Contents ... 4
List of Figures ... 7
Abbreviations ... 9
1 Introduction ... 11
2 Executive Summary ... 12
3 State of the Art ... 14
3.1 Software Defined Networking in mobile networks ... 14
3.2 Mobility Management in 3GPP ... 16
3.2.1 IP Mobility ... 16
3.2.2 Mobility in 3GPP ... 18
3.2.3 New approaches to provide Mobility Management in Small Cell Networks ... 18
3.2.4 Mobility Management and Software Defined Networking ... 19
3.3 Network Energy Optimisations ... 20
3.3.1 Energy Saving Concept in 3GPP ... 20
3.3.2 Energy Efficiency Efforts in EU Projects ... 21
3.3.3 Other Works to Improve Energy Consumption in Mobile Backhaul Networks ... 21
3.3.4 Other Works to Improve Energy Consumption in Cellular Networks ... 22
3.4 Congestion Control ... 23
3.5 SON in Backhaul Networks ... 25
3.6 Analysis of Backhaul Networks based on Stochastic Geometry ... 25
3.7 Load Aware Network Architecture ... 26
3.8 Traffic Management in Backhaul Networks ... 27
4 Description of Candidate Technologies ... 28
4.1 CT 4.1: Distributed IP Anchoring and Mobility Management ... 28
4.1.1 Motivation ... 28
4.1.2 Assumptions ... 29
4.1.3 Technical Requirements ... 30
4.1.4 Description of Candidate Technology ... 30
4.1.4.1 Distributed Anchoring and Mobility Management ... 31
4.1.4.2 Mobility Management using the Virtual Cell Concept ... 35
4.2 CT 4.2: Network Wide Energy Optimisation ... 44
4.2.1 Motivation ... 44
4.2.2 Assumptions ... 44
4.2.3 Technical Requirements ... 44
4.2.4 Description of Candidate Technology ... 45

4.2.4.1 Network Wide Energy Optimisation using SDN Approach 45
4.2.4.2 Energy Optimisation Solution Considering QoS Constrains and Traffic Profile 50
4.3 CT 4.3: Joint Path Management and Topology Control 53
4.3.1 Motivation 53
4.3.2 Assumptions 54
4.3.3 Technical Requirements 54
4.3.4 Description of Candidate Technology 54
4.4 CT 4.4: Routing and Congestion Control Mechanisms 57
4.4.1 Motivation 57
4.4.2 Assumptions 58
4.4.3 Technical Requirements 58
4.4.4 Description of Candidate Technology 59
4.4.4.1 iJOIN Congestion Control 59
4.4.4.2 Joint Routing and Congestion Control 61
4.5 CT 4.5: Network Wide Scheduling and Load Balancing 62
4.5.1 Motivation 62
4.5.2 Assumptions 63
4.5.3 Technical Requirements 63
4.5.4 Description of Candidate Technology 64
4.6 CT 4.6: Backhaul Analysis based on Viable Metrics and “Cost” Functions using Stochastic Geometry 68
4.6.1 Motivation 68
4.6.2 Assumptions 68
4.6.3 Technical Requirements 68
4.6.4 Description of Candidate Technology 68
5 WP4 Functional Architecture and Interactions of CTs 70
5.1 Overview 70
5.2 Interface Specification 72
5.2.1 Consolidated List of Required Input of WP4 CTs with Source Information 72
5.2.2 Consolidated List of Provided Output of WP4 CTs with Sink of Information 73
5.3 Module Specification: Interaction of WP4 Modules 75
5.3.1 Network Bootstrap Procedure 75
5.3.2 UE Attachment Procedure 76
5.3.3 Congestion Management Procedures 76
5.3.4 Energy Optimisation Procedures 78
6 Additional Considerations on the Use of Software Defined Networking 80
6.1 Impact of the Use of SDN 80
6.2 SDN Protocols 83
7 Summary 85

Acknowledgements and Disclaimer 86
References 87

List of Figures

Figure 3-1 SDN architecture as proposed by the Open Networking Foundation (ONF) 15
Figure 3-2 Centralised IP mobility approaches 16
Figure 3-3 Sketch of the four trade-offs relations without and with practical concerns networks [75] 22
Figure 3-4 Congestion control in a mobile network.. 24
Figure 3-5 Realization of a point process.. 26
Figure 4-1 Reference scenario for IP mobility related technologies 29
Figure 4-2 Outline of the DMM solution 31
Figure 4-3 MSC of the attachment procedure (anchor and mobility entities).. 32
Figure 4-4 Anchor selection algorithm (simplified view) 32
Figure 4-5 MSC of the intra-anchor handover procedure (anchor and mobility entities) 33
Figure 4-6 MSC of the inter-anchor handover procedure (anchor and mobility entities) 33
Figure 4-7 MSC of the new anchor assignment procedure (anchor and mobility entities) 34
Figure 4-8 MSC of an energy-triggered handover 35
Figure 4-9 UE's involvement in iJOIN WP4.. 35
Figure 4-10 Mobility management: concept of virtual cells 36
Figure 4-11 Outline of the mobility management solution 37
Figure 4-12 Overview of the handover scenarios.. 38
Figure 4-13 MSC of 3GPP X2-based handover in Femtocells [99].. 39
Figure 4-14 MSC of intra anchor handover with temporary mobility anchor in iSC.................................... 40
Figure 4-15 Intra anchor handover scenario with temporary mobility anchor in iSC.................................... 41
Figure 4-16 MSC of intra anchor handover with iTN mobility anchor... 41
Figure 4-17 Intra anchor handover scenario with iTN mobility anchor 42
Figure 4-18 MSC of inter anchor handover... 43
Figure 4-19 Inter anchor handover scenario.. 43
Figure 4-20 Outline of the network-wide energy optimisation solution 46
Figure 4-21 System overview of the SDN-based energy optimisation solution.. 47
Figure 4-22 MSC of SDN-based energy optimisation solution... 48
Figure 4-23 Network Wide Energy Optimisation 51
Figure 4-24 Traffic Profile of an iSC 52
Figure 4-25 Power-consumed profile of two different iSC 52
Figure 4-26 Coverage gaps due to turned-off iSC.. 53
Figure 4-27 Exemplary radio access and backhaul network deployment with heterogeneous backhaul technologies.. 55
Figure 4-28 Protocol view on radio access and backhaul network 56
Figure 4-29 Q in Q 56
Figure 4-30 Example for QinQ tag used in iJOIN.. 57
Figure 4-31 iJOIN congestion control use case (example)... 58

Figure 4-32 Use of alternative routes for off-loading congested nodes 60
Figure 4-33 Congestion Control process message exchange 61
Figure 4-34 Outline of joint routing and admission/congestion control........ 62
Figure 4-35 Overview of Load Balancing and Scheduling 63
Figure 4-36 Load Balancing example 64
Figure 4-37 SITA overview [109]........ 65
Figure 4-38 MSC of Load Balancing 66
Figure 4-39 Scheduling example........ 66
Figure 4-40 MSC of Scheduling........ 67
Figure 4-41 LB and Energy saving procedures 67
Figure 4-42 Network Model considered for Analysis 69
Figure 5-1 WP4 functional architecture 71
Figure 5-2 Interaction during Network Bootstrap 75
Figure 5-3 Interaction during UE Attachment........ 76
Figure 5-4 Interaction during Congestion Management (no UE relocation needed) 77
Figure 5-5 Interaction during Congestion Management (UE relocation needed) 78
Figure 5-6 Interaction during Energy Optimisation implying switch-off of iTN (no UE relocation)........ 78
Figure 5-7 Interaction during Energy Optimisation implying switch-off of iSC (and UE relocation)........ 79
Figure 6-1 Downlink scheduling process in LTE........ 81
Figure 6-2 Common scheduling across base stations........ 81
Figure 6-3 Scheduling through an external node........ 82
Figure 6-4 SDN architecture for common scheduling........ 83
Figure 6-5 SDN protocols architecture........ 83

Abbreviations

3GPP	3rd Generation Partnership Programme
ANDSF	Access Network Discovery and Selection Function
API	Application Programming Interface
AQM	Active Queue Management
BGP	Border Gateway Protocol
BS	Base Station
CCO	Capacity and Coverage Optimisation
CSG	Closed Subscriber Group
DGW	Distributed Gateway
DMM	Distributed Mobility Management
DNS	Domain Name Service
DSMIP	Dual Stack Mobile IP
ECN	Explicit Congestion Notification
eNB	Evolved Node B
EPS	Evolved Packet System
FBS	Femto Base Station
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GTP	GPRS Tunnelling Protocol
HSS	Home Subscriber Server
IETF	Internet Engineering Task Force
IFOM	IP Flow Mobility
iJOIN	Interworking and JOINt Design of an Open Access and Backhaul Network Architecture for Small Cells based on Cloud Networks
iNC	iJOIN Network Controller
IP	Internet Protocol
IT	Information Technology
LEDBAT	Low Extra Delay Background Transport
LGW	Local Gateway
LIMONET	LIPA Mobility and SIPTO at the Local Network
LIPA	Local IP Access
LMA	Local Mobility Anchor
LTE	Long Term Evolution
MAG	Mobility Access Gateway
MIMO	Multiple Input Multiple Output
MIPv6	Mobile IP
MME	Mobility Management Entity

MN	Mobile Node
MPLS	Multiprotocol Label Switching
MSC	Message Sequence Chart
NAS	Non-Access Stratum
NB-IFOM	Network-Based IP Flow Mobility
NFV	Network Function Virtualization
NGMN	Next Generation Mobile Network Alliance
OAM	Operations, Administration, Maintenance
ONF	Open Networking Foundation
OPEX	Operational Expenditures
OSG	Open Subscriber Group
OSPF	Open Shortest Path First
PDN	Packet Data Network
P-GW	Packet Gateway
P2P	Peer to Peer
PMIP	Proxy Mobile IP
PQ	Phantom Queue
QoE	Quality of Experience
QoS	Quality of Service
RACH	Random Access Channel
RAN	Radio Access Network
RANaaS	RAN as a Service
RAP	Radio Access Point
RAT	Radio Access Technology
RED	Random Early Detection
RNL	Radio Network Layer
RRM	Radio Resource Management
SDN	Software Defined Networks
SIPTO	Selected IP Traffic Offload
SLA	Service Level Agreement
SON	Self-Organised Networks
UE	User Equipment
WLAN	Wireless Local Area Network

1 Introduction

The iJOIN project aims on tackling the increasing rate demands in future networks. Its approach incorporates two key concepts: very dense small-cell networks and (partly) centralised processing of Radio Access Network (RAN) functionality. Further to that, it considers the application of heterogeneous backhaul which requires a flexible assignment of functionality within the mobile network. The mobile network consists not only of the radio access part but also the backhaul network which connects radio access network and core network. Therefore, iJOIN pays special emphasis on the interworking and joint optimisation of radio access and backhaul network. This joint operation will also require a joint management of networks in order to avoid contradicting measures applied to either one. Furthermore, iJOIN considers the partial remote execution of RAN functionality which will impose further requirements on the network management in order to guarantee minimum requirements for the centralisation of functionality. Hence, the focus of this deliverable is to analyse the network management and orchestration of the iJOIN system.

The unique attribute of a mobile network is the mobility of its users, i.e., users may connect from different physical locations. This implies temporal and spatial changes of the traffic distribution. As networks become denser in order to provide the required capacity, also their management becomes more difficult. In order to orchestrate the dense network efficiently the management includes the mobility support and its required signalling efforts, energy optimisations, admission and congestion control, scheduling and load balancing, and traffic routing. In iJOIN we have adopted a logically centralised approach, based on Software Defined Networking (SDN), to orchestrate, coordinate and execute the different network mechanisms. This approach is aligned with current trends in the industry [1] and we plan to influence and impact ongoing research and standardisation efforts, by analysing the specific requirements exhibited by mobile backhaul networks, as well as by the functional split due to partial remote execution of RAN functionality.

Beside the active network management after deployment (maintenance), also deployment and dimensioning aspects need to be taken into account. This includes the deployment density, connectivity, and performance of backhaul links which are required to satisfy the key performance indicators of iJOIN, i.e.,:

- Energy-efficiency: for instance, if base stations (eNB) are turned off, energy saving potential is only exploited in the radio access network. However, this impacts the backhaul network, which needs to provide sufficient diversity in order to allow for exploiting further energy saving potential in both the RAN and the backhaul network.

- Cost-efficiency: for instance, deploying a very dense small cell network will also require a very dense backhaul network able to carry the required data rates. This requires a dimensioning of the backhaul network that is cost-efficient but still allows for high data rates. Efficient mobility support and routing algorithms to optimise the congestion are key elements.

- Utilization-efficiency: for instance, RAN and backhaul need to be utilised optimally instead of being dimensioned for peak-throughput. This requires to exploit diversity effects and to deploy efficient load balancing and routing solutions. The logical centralisation provided by the use of an SDN-based architecture enables to achieve a close to optimal solution.

- Spectral efficiency: for instance, available backhaul resources should be used as efficiently as possible in order to reduce costs and increase system capacity. This is related to energy efficiency solution in the backhaul networks regarding how to efficiently switch on/off backhaul network entities/links.

In addition, future deployments will be very diverse, i.e., as much as services evolve, also the scenarios and use cases will further diversify. Therefore, iJOIN focuses on four different scenarios, i.e., dense hotspot in a stadium, dense hotspot on city square, wide area coverage with high data rates, and dense indoor deployments, as described in D5.1 [2]. All these deployment scenarios will require different network management functions and abilities.

2 Executive Summary

This deliverable summarises the main research activities by WP4 during the first twelve months of the iJOIN project. The main objective of this document is to present the state-of-art literature review for the topics related to the WP4 scope "Network-layer solutions and system and operation management", and to provide an overview of the candidate technologies to be further explored in the iJOIN project. The structure of this deliverable is described next.

Section 3 introduces the state of the art within those topics that are of particular relevance for this work package, i.e., software defined networking, mobility management, energy optimisation, load aware network architecture, self-organised networks, traffic engineering and stochastic analysis of backhaul networks. The analysis of the state of art does not only point currently deployed protocols and solutions but also outlines their deficiencies with respect to the requirements posed by use cases considered by iJOIN.

Based on the analysis of the state of the art, Section 4 introduces a set of candidate technologies which will be further investigated in the course of this project. Each candidate technology is introduced first with a brief motivation. Furthermore, the general assumptions and technical requirements which are imposed by each candidate technology are listed, and a detailed description of each candidate technology is provided. The design of each candidate technology will be further developed within the project lifetime. In particular, the following candidate technologies are introduced:

- Distributed IP anchoring and mobility management.
- Network-wide energy optimisation.
- Joint path management and topology control.
- Routing and congestion control mechanisms.
- Network wide scheduling and load balancing.
- Backhaul analysis based on stochastic geometry.

In Section 5, the WP4 functional architecture and interactions between candidate technologies are summarised. This functional architecture is heavily influenced by the Software Defined Networking (SDN) approach that has been adopted as a common WP4 platform. The main intelligence of the WP4 candidate technologies is logically centralised in the iJOIN Network Controller (iNC). The iNC is a key network entity in charge of configuring, monitoring and driving the operation of the rest of the radio access and backhaul network entities. An iJOIN-extended OpenFlow controller located at the iNC, takes care of all the protocol interactions with the rest of the WP4 network entities. Inside the iNC there is basically one module per WP4 candidate technology, except for the stochastic backhaul network analysis candidate technology, which deals with planning and pre-provisioning aspects of the network and does not require any runtime module. The interactions of WP4 modules within iNC are presented and analysed in this section. The interface specifications are summarised in a list, which is provided as input to WP5.

Finally, Section 6 provides a first assessment of the impact of using SDN technologies in the iJOIN architecture, as well as explores the associated benefits. It also presents the SDN protocols that are currently foreseen to be adopted by WP4 candidate technologies and the potential needs to extend them to fit the WP4 scope.

During this first project phase the SotA has been revised and analysed, basic specifications for the iJOIN project have been defined and promising candidate technologies have been investigated, by first clearly defining a set of goals and identifying the main characteristics of the mechanisms to be further developed in the next months. However, the research results achieved within these first twelve months have already allowed us to lead to publications or submissions to prestigious international conferences and journals, as well as to contribute to relevant standardisation organisations. Subsequently, the main results of these publications are briefly described.

In [3] we summarise the main challenges that WP4 will tackle during the iJOIN project lifetime, namely those related to the joint design of the backhaul and radio access network in a cloud-based mobile network. Some of the mechanisms that are being researched are also introduced.

In [4] we present and analyse (using simulations and also an experimental platform with running code) some of the mobility mechanisms that we plan to further develop in the next months, adapting them to the iJOIN specific scenario.

In terms of standardisation, we have especially active. We are co-authoring the two main documents of the IETF Distributed Mobility Management (DMM) Working Group (WG): the one describing the requirements that DMM solutions should met [5], and the one identifying the gaps between existing DMM practices and the aforementioned requirements [6]. Additionally, we are also co-authors of the WG document analysing how to apply congestion exposure (CONEX) mechanisms to 3GPP mobile networks [7].

In the Open Networking Foundation (ONF), iJOIN has also contributed to some of the use cases that have been included for consideration by the recently created Wireless & Mobile Working Group (WMWG). This is a very important result for iJOIN, as some of the work to be carried out in the next months within the project will have a direct application to the work conducted at the WMWG.

3 State of the Art

3.1 Software Defined Networking in mobile networks

Software Defined Networking (SDN) is defined as “a network architecture in which the network control plane is decoupled from the physical topology.” But beyond this general declaration, several trends can be identified:

- Separation of hardware from software, i.e., choosing hardware based on necessary features and software based on protocol requirements.
- Logically centralised network control, which is considered to be more deterministic, efficient and fault tolerant.
- Automation: separate monitoring, management, and operation.

SDN allows for making currently rather static networks more flexible by tailoring and optimising specifically them for different use cases. This creates a dynamic network environment which is adapted to the needs of the applications running on top of it [8].

According the Open Networking Foundation[1], the SDN architecture should be:

- **Directly programmable**: Network control is directly programmable because it is decoupled from forwarding functions.
- **Agile**: Abstracting control from forwarding lets administrators dynamically adjust network-wide traffic flow to meet changing needs.
- **Centrally managed**: Network intelligence is (logically) centralised in software-based SDN controllers that maintain a global view of the network, which appears to applications and policy engines as a single, logical switch.
- **Programmatically configured**: SDN lets network managers configure, manage, secure, and optimise network resources very quickly via dynamic, automated SDN programs, which they can write themselves because the programs do not depend on proprietary software.
- **Open standards-based and vendor-neutral**: When implemented through open standards, SDN simplifies network design and operation because instructions are provided by SDN controllers instead of multiple, vendor-specific devices and protocols.

The SDN framework, as proposed by the Open Networking Foundation (and in a simplified way), is represented in Figure 3-1.

[1] http://www.opennetworking.org/

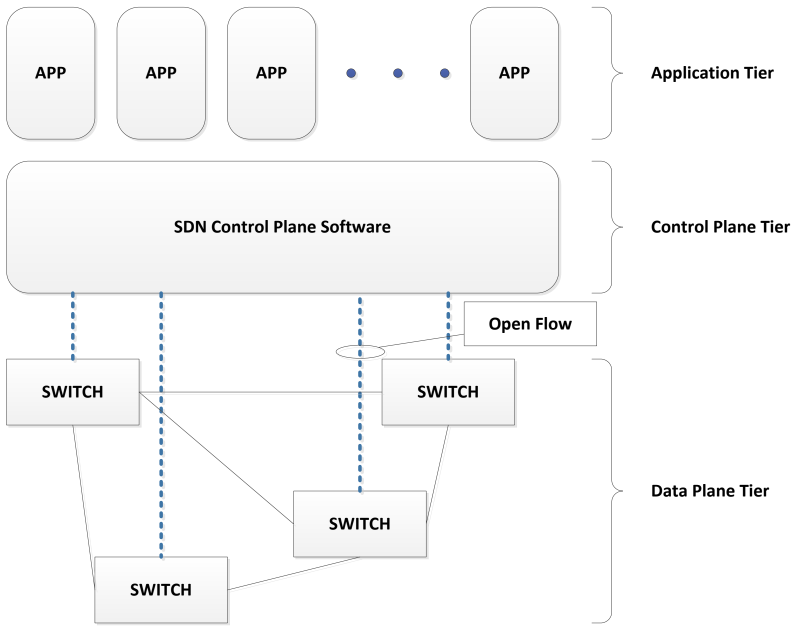

Figure 3-1 SDN architecture as proposed by the Open Networking Foundation (ONF)

There are three tiers that compose the framework of the SDN architecture:

- Application tier, e.g., virtual network overlays, network slicing (delegation), tenant-aware broadcast, application-aware path computation, integration with other software packages, policy, security, traffic engineering.
- Control plane tier, e.g., data plane resource marshalling, common libraries (e.g., topology, host metadata, state abstractions).
- Data plane tier, e.g., packet forwarding (as per flow table), packet manipulation (as per flow table), and collection of statistics.

Between the application tier and the control plane tier a number of open Application Programming Interface (API) are defined, whose level of standardisation is being explored by the Open Networking Foundation (ONF) [9]. Between the data plane tier and the control plane tier the communications is carried out according to a standardised protocol, namely OpenFlow, which is promoted by the ONF. OpenFlow allows direct access to and manipulation of the forwarding plane of network devices such as switches and routers, both physically and virtually (hypervisor-based). The protocol specifies basic primitives that can be used by an external software application to program the forwarding plane of network devices, just like the instruction set of a CPU would program a computer system. OpenFlow is based on the approach known as “Match-Action". Roughly, a subset of packet bytes are matched against a table; the matched entry specifies a corresponding action(s) that are applied to the packet. Using this approach, Openflow has made significant progress by establishing *i)* flow tables as a standard data-plane abstraction for distributed switches, *ii)* a protocol for the centralised controller to install flow rules and query states at switches, and *iii)* a protocol for a switch to forward to the controller packets not matching any rules in its switch-local flow table.

Current access networks (last-mile backhaul networks) operate rather inefficiently [10]. The closer the aggregation network, the more efficient because diversity effects may be used (and need to be controlled). Furthermore, the move towards IP based backhaul networks also implies to move from fixed to non-deterministic bandwidth planning and management for continuously extending backhaul capacity.

Recently, some efforts have been done to apply SDN concepts in 3GPP networks. Some of them propose to virtualise the control plane side of the entities of the Evolved Packet Core (EPC), so they can run in a cloud environment, and use a simpler and flatter network architecture in the data plane, following an SDN approach to manage the forwarding and allowing to benefit from enhanced flexibility [11] [12] [13] [14]. Note that the work performed so far by the ONF has been limited to wired Ethernet-alike networks, although the creation of Wireless and Mobile Group has been recently approved (October 2013) to discuss progression of methods by which OpenFlow can be used to control wireless and mobile Radio Area Networks (RAN) and core networks. SDN is applied to backhaul and radio access networks [15] [16] and allows for flexible adaptation of both, e.g., if different backhaul technologies are used and need to be controlled, or if physical links may be used for fronthaul and backhaul based on the functional split. This is in line with the backhaul requirements stated by NGMN in [17], which state that backhaul networks need a higher degree of configurability, e.g., granularity of information rates, resource sharing and prioritization of operators, traffic shaping, admission control, and load balancing. One could even consider SDN as part of RANaaS which applies the optimal functional split based on the underlying backhaul technology.

A complementary technology to SDN is Network Function Virtualisation (NFV) [18] which refers to implementing and virtualising network functionality on standard IT hardware. This is of particular interest for iJOIN where radio access network functionality is virtualised and should be accessible in a decentralised way on eNBs as well as in a centralised way in data centres. This functional split and flexible assignment requires virtualising radio access network functionality, which is also investigated within the iJOIN project.

3.2 Mobility Management in 3GPP

Within 3GPP, multiple approaches are standardised to cope with User Equipment's (UE) mobility. This covers both mobility within a given access network technology (e.g., intra-3GPP) and also between different radio access technologies (RATs).

3.2.1 IP Mobility

In order to integrate an heterogeneous set of access technologies, mobility can no longer be considered an intra-technology issue managed at link level. New network-layer mobility functions need to be introduced in order to extend mobility capabilities beyond the link-based procedures available at each access technology. Mobility solutions can be adopted at two different levels, depending on the terminal's degree of involvement in the mobility process: global or host-based mobility and local or network-based mobility.

Global or host-based mobility is based on the ability of the terminal to maintain a persistent and globally accessible address independent of its current point of attachment to the network. As its name indicates, it requires the terminal's awareness and involvement in the mobility process. Usually, a mobility management protocol is applied to map the persistent address to the temporary local address available at each moment. In further modifies accordingly the end-to-end routing of packets in order to maintain connectivity. This is addressed by Mobile IPv4/6 protocols [19] [20] standardised by IETF. Apart from host involvement, it also requires a Global Mobility Anchor (GMA) entity in the home network, where the permanent address of the terminal is attached and the mapping with the temporary address is kept.

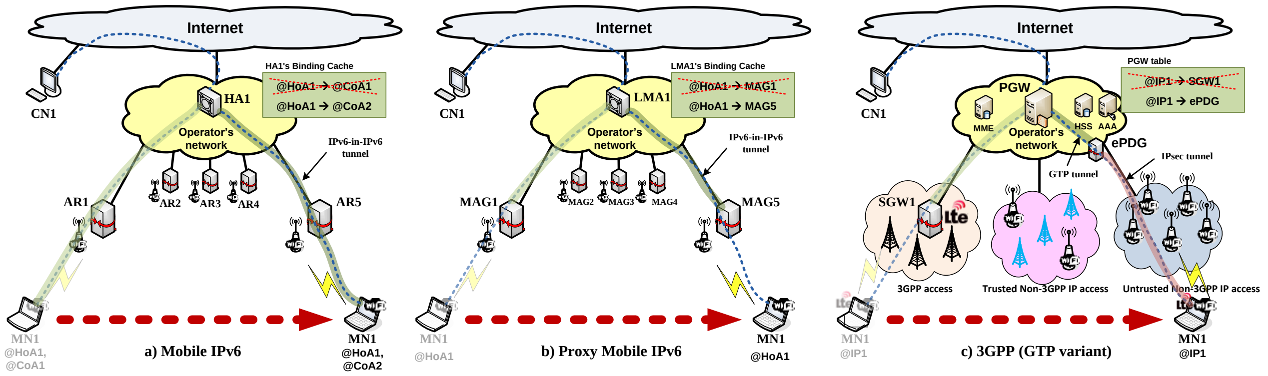

Figure 3-2 Centralised IP mobility approaches

Local or network-based mobility is intended to allow terminals to maintain connectivity when moving across a certain network area, known as local mobility domain. This type of mobility is provided by the network with no need of terminal intervention or awareness. Local mobility is based on the presence of two network entities: the Mobility Access Gateway (MAG) and the Local Mobility Anchor (LMA). The MAG is in charge of tracking terminal location and updating it on the LMA on behalf of the terminal by means of a mobility management protocol, whereas the LMA is devoted to keeping the terminal's persistent address and routing inbound and outbound traffic.

Figure 3-2 summarises the operation of the main centralised IP mobility approaches: a) Mobile IPv6 (global/client mobility), b) Proxy Mobile IPv6 (local/network mobility), and c) 3GPP (GTP variant).

Current IP mobility protocols (GTP [21], PMIPv6 [22] and DSMIPv6 [23]) are centralised and require all user data traffic to traverse the mobile operators' core network. This implies several limitations [24]:

a) Sub-optimal routing: Mobile nodes are anchored at a central entity that leads to IP routing-paths that are generally longer than necessary.

b) Scalability problems: core networks are dimensioned to support peak data traffic.

c) Reliability: the central entity/core network is a single point of potential failure.

d) Lack of fine granularity of the mobility management service.

Within the IETF, the working group on Distributed Mobility Management (DMM)[2] was chartered in March 2012 to work on a more generic framework for distributed mobility management. The group is currently working on identifying challenges and the scope of potential solutions [5][6]. Most of the solutions presented within the IETF community re-use existing definitions and operations specified for MIPv6 and PMIPv6, respectively for client- and network-based solutions. The solution described in [25] modifies the MIPv6 home agent in order to deploy it in each access network as default router for the mobile node (MN). This allows the MN to have several anchors, enabling better path establishment and improving handover latency. The authors in [26] introduce two possible approaches as network-based solutions, i.e., a partially distributed solution, where a central database supports mobility-featured access routers as a mobility sessions store, and a fully distributed solution, where the central mobility database is removed. In [27], the central session server is maintained, but the signalling is changed. These solutions provide dynamic mobility and fast handover management, but standard access routers need to be enriched with mobility functionalities inherited by the PMIPv6's local mobility anchor and mobile access gateway. A different approach is presented in [28], where the main idea is to deploy many small PMIPv6 domains and to define a signalling protocol for LMA-to-LMA communications. Many benefits from the well-known PMIP protocol can be gained, but the design does not meet the flatness envisioned for future network architectures.

An overview of the DMM impact in standardisation, both for IETF and 3GPP is given in [29], while other non-standard related, but more generic, solutions were proposed in [30]-[37]. The articles [30]-[32] explore similar solutions as those in [25] [26], focusing on extensions and giving more attention to examples of deployment use cases. In [33] and [34], authors propose and evaluate a network-based mechanism for DMM without using dedicated signalling. Similarly to the solutions mentioned above, mobility capable access routers can re-configure the routing in the network to grant session continuity for a moving terminal. The mechanism relies on inspecting the traffic generated by the terminal. Furthermore, peer-to-peer strategies are evaluated in [35] [36]. In [35] [36], access routers still need some mobility features to maintain the mobility database and to anchor IP flows. The signalling is derived from peer-to-peer (P2P) technologies, such as employing distributed hash tables and creating a P2P overlay. The authors of [37] propose to handle mobility management relying on routing and DNS updates, based on Interior Border Gateway Protocol (iBGP), Border Gateway Protocol (BGP) and dynamic DNS protocols. Opposite to the previous designs, here routers do not need modifications, but the performance of the location update is bounded by the convergence time of the routing protocols.

Another effort that is gaining traction lately in the IETF is the separation of the data and control plane signalling in the main IETF IP protocols, such as for Proxy Mobile IPv6 [38]. IETF IP mobility protocols exhibit a coupled data and control plane signalling, which differs from the 3GPP approach followed by GTP and the use of the MME as pure control plane entity. The IETF is now looking at how to extend their protocols to allow such a separation, which would also help enabling the virtualisation of the evolved packet core in the 3GPP architecture, as proposed in [39], where a virtual EPC (vEPC) architecture is run in a cloud infrastructure.

While all these solutions can provide some gains over existing centralised approaches, they do not take into account the particularities of the iJOIN use cases:

- Current DMM efforts ignore very dense deployments, and this might have a significant impact, for example in how to properly select an anchor point, because changes of attachment points will be more frequent in a small cell based environment.
- Interactions with the backhaul network, as for example to enable energy efficiency optimisations, or to allow for jointly optimising access and backhaul networks.

[2] http://datatracker.ietf.org/wg/dmm/

These aspects, however, imply additional constraints and challenges to mobility-related procedures, such as access discovery and selection. The current ANDSF framework was not designed to operate well in these scenarios. Other solutions such as IEEE 802.21 [40] would also need extensions to properly operate when the density of potential attachment points increases significantly.

3.2.2 Mobility in 3GPP

The UE mobility state in 3GPP systems is classified into two states: idle mode and connected mode [41] [42]. In idle mode [42], the cell selection and re-selection is performed for the mobility management of UE. When a UE is turned on, the UE searches for a suitable cell and chooses this cell to provide available services and tunes to its control channel. This is referred to as “camping on the cell.” If the UE finds a more suitable cell, according to the cell re-selection criteria, it selects the more suitable cell and camps on it, which is referred to as cell re-selection. When a call is initiated, the idle mode is transited to the connected mode.

LTE utilises a network-controlled handover procedure which is assisted by the UE in connected mode [41]. The UE measures the signal strength and sends a measurement report to the serving eNB. The serving eNB then performs the handover decisions based on the measurement reports. The handover procedure consists of three steps: handover preparation, handover execution and handover completion.

In case of 3GPP access, the Serving Gateway (S-GW) terminates the GPRS Tunnelling Protocol (GTP) interface towards the 3GPP radio access networks, and therefore it is the extended equivalent of the Gateway GPRS Support Node (GGSN) of previous 3GPP releases. In addition, the S-GW provides IP routing features and takes the MAG role for the provision of network-based mobility. Conversely, the Packet Gateway (P-GW) provides the interface towards PDNs and takes the LMA role. Local mobility management between the S-GW and the P-GW may be based on GTP or on Proxy Mobile IPv6 (PMIPv6) [22], which is a local mobility implementation based on the MIPv6 protocol. There are also plans for global mobility support under 3GPP access, thus making the P-GW take also the GMA role and basing global mobility management between the UE and the P-GW on Dual Stack MIPv6 (DSMIPv6) [23], which is a MIPv6 variant for dual stack (IPv4 + IPv6) hosts.

At present, operators are looking for more distributed approaches that are cheaper and more efficient. The 3GPP is working on approaches that permit offloading traffic from the operators’ core network. These will result in two main solutions: Selected IP Traffic Offload (SIPTO) and Local IP Access (LIPA) [43]. SIPTO enables an operator to offload certain types of traffic at a network node close to that UE’s point of attachment to the access network.

LIPA enables an IP capable UE that is connected by means of a femto-cell to access other IP capable entities in the same residential/enterprise IP network. In order to achieve this, a Local GW (LGW) that is collocated with the femto-cell is used. Both SIPTO and LIPA have very limited mobility support, especially in 3GPP specifications of up to Rel-10. In Rel-11, a work item on “LIPA Mobility and SIPTO at the Local Network (LIMONET)” [44] studied how to provide additional, but still limited, mobility support to SIPTO and LIPA mechanisms, albeit mainly restricted to a localised area and requiring PDN connections to be deactivated and re-activated when not moving locally.

Furthermore, the research community has proposed extensions to current 3GPP mechanisms, such as [45] and [46]. These works mainly deal with relocating the P-GW in order to contrast the sub-optimal path that data packets need to take to traverse the P-GW. Authors of [47] propose instead to introduce an additional entity in the EPS called Distributed GW (D-GW), specifically designed to handle mobility in a distributed way, taking into account also non 3GPP access.

While 3GPP specifications have been making efforts to enhance mobility management in femtocell networks, there are still some remaining issues which are to be improved further. iJOIN project will focus on proposing a distributed mobility management scheme which reduces the signalling towards EPC and minimize the handover latency.

3.2.3 New approaches to provide Mobility Management in Small Cell Networks

The authors of [48] propose two mobility management schemes applied to the Femto-GW at Radio Network Layer (RNL) for LTE Femto-to-Femto handover. The first proposal suggests that the Femto-GW could act as a mobility anchor which makes handover decisions. When the Femto-GW receives a handover request from the source cell, it checks the target cell ID. If the target cell is under its control, it will handle the handover directly. By contrast, the second proposal describes a Femto-GW which operates as a transparent node which simply forwards all handover messages between the Femto cell and the MME. After handover completion, the S-GW is notified about the change of the attachment point. The first proposal is more suitable for enterprise use, because it reduces the signalling traffic within the core network. On the other hand, the second proposal is more suitable for home use, because more signalling messages are exchanged.

In [49] and [50], an adapted signalling flow is proposed for the three types of handovers in heterogeneous LTE networks, i.e., handover from macro cell to small cell, handover from small cell to macro cell, and handover between small cells. The proposed scheme considers the movement prediction mechanism as an additional parameter for handover decision, which effectively makes it a client-based handover. Reactive and proactive handover procedures are proposed to trigger the handover, because the handover procedure may be initiated by femto-cell, macro/micro-cell, and the UE. In reactive handover, the handover is triggered when the UE has almost lost its serving cell signal or the most likely position of the UE can be predicted. Reactive handover aims to postpone the handover as long as possible to prevent frequent and unnecessary handover, and mitigate the generated overhead of handover. However, in proactive handover, the handover may occur at any time before the handover-condition is met, e.g., via estimate of the exact position of the UE. Proactive handover is expected to minimise packet loss and latency during handover.

In [51], a method for access and handover management for OFDMA femto-cell networks is proposed. In a Closed Subscriber Group (CSG) scenario, when a UE comes near a femto-cell, its serving macro-BS will check the UE’s ID. If it is within the allowed list, the macro-BS informs the femto-BS to start the handover procedure. Otherwise, the macro-BS should notify the femto-BS to start a proactive interference management procedure. The authors also propose a hybrid access to the same scenario. After a non CSG UE enters a femto-cell, the cell measures the UE’s signal strength and decides whether the potential interference caused by the UE is above the interference threshold or not. If so, the femto-BS will request a handover procedure from the serving macro-BS for the UE and indicates that this is done to avoid interference. The CSG scenario reduces unnecessary handovers and signalling load. However, in the hybrid scenario, the number of HOs is increased.

To solve the same problem as presented in [51], it is proposed in [52] a pseudo handover based on the direct information exchange between base stations. This exchange includes sub-channel and power adaptation information in order to avoid excessive interference. The pseudo handover is executed in the Radio Access Network (RAN) and does not imply any signalling exchange with the MME in order to reduce signalling overhead significantly. When the UE tries to camp on a CSG FBS where it belongs to, the regular handover is triggered. Otherwise, the pseudo handover is triggered. The FBS will set up and maintain a table all non-CSG UEs that tried to camp on it (and executed a pseudo-handover).

Although there are many proposed solutions for handover management in small cells, most of the solutions have targeted only one or two parts of the handover procedure, such as handover preparation, handover decision parameter, handover signalling, macro cell to small cell hand-in algorithm. In [48]-[50] [53]-[57], schemes are proposed for the signalling flow during the handover process with different additional parameters used to reduce the number of unnecessary handovers. For instance, [48] supports CSG and Open Subscriber Group (OSG) scenarios for the handover between femto-cells. The scheme uses the user speed, QoS, and load balancing as additional parameters for the handover decision. A comprehensive handover procedure is required to achieve optimal handover performance.

While a lot of existing research work have been making efforts to design the handover procedures in femtocell networks, there are still some issues which need to be worked on further. iJOIN project will focus on proposing a complete handover procedure in order to achieve better handover performance. And other issues such like congestion control will be considered as criteria during the handover procedure.

3.2.4 Mobility Management and Software Defined Networking

As mentioned above, the Software Defined Networking (SDN) paradigm is gaining momentum in the last years, being applied to many different aspects of traditional networking. While it started in the "wired" world, for example to improve the performance of data centres (as Google is currently doing in its production networks), the use of SDN approaches is also being evaluated for mobile wireless networks.

In [11], the concept of software defined network mobile network (SDMN) is introduced, an architecture that builds on the decoupling of data and control in the mobile network user plane and the definition of a OpenFlow based approach, called *MobileFlow*, to control the forwarding in the IP/Ethernet transport network. While this is probably the most recent article reporting on the use of SDN approaches to provide mobility and efficient control and data plane separation in a 3GPP network, there have been some previous works proposing the use of SDN in wireless networks (such as Wi-Fi ones) [58].

On the standards world, early attempts to apply SDN concepts to mobile networks can also be identified already. For example, in the IETF, [59] addresses the problem of mobility in SDN networks, though it tackles this issue quite superficially, being a mere trigger for discussion in the IETF community. Similar efforts can be found in the IEEE, where the IEEE 802 OmniRAN EC Study Group is looking at the use of SDN techniques to perform access network abstraction in a multi heterogeneous network based on IEEE 802 technologies. Last, but not least, the ONF is also discussing extensions of OpenFlow for mobile and wireless networks, as introduced before in this document.

iJOIN attempts to go even further, as our use cases involve not only the use of SDN to tackle mobility management, but also to handle additional mechanisms (such as congestion control, energy optimisations, load balancing) in a mobile backhaul.

3.3 Network Energy Optimisations

3.3.1 Energy Saving Concept in 3GPP

Currently, most mobile network operators aim at reducing their energy consumption by several methods of optimising the network planning and management. In the new generation radio access networks such as LTE, the Energy Saving Management (ESM) function is required especially when network operators want to reduce transmit power, switch on/off cell based on network measurements which shows that there is no need to maintain active full set of network elements capacities. Energy efficiency in mobile networks can be envisioned as follows:

- To optimise the number of sites, which maintaining coverage, capacity and quality of service. This can be integrated with network planning procedures.
- To optimise energy efficiency of these sites and minimise energy consumption of equipment.
- To explore renewable energy sources such as wind and solar energy.

In the context of LTE technology combined with Self-optimised Networking (SON) functionalities, the main focus is to identify mechanisms to optimise E-UTRAN equipment energy consumption (second point as above). Network Operations, Administration, Maintenance (OAM) system plays an essential role in the optimisation process, either directly by locating the optimisation functions within the management system or indirectly by providing relevant performance information to optimisation functions elsewhere [60].

Three architectures are candidate to offer energy saving functionalities [60]:

- Distributed architecture, where network elements collect relevant information and trigger appropriate self-optimising algorithms when it is needed. In the distributed architecture, there is no OAM involvement.
- Centralised architecture, where OAM collects relevant information from network elements, triggers appropriate self-optimisation algorithms and decides the further actions to be taken on the network elements.
- Hybrid architecture, which is a combination of the distributed and centralised architectures.

The specification requirements for distributed architecture and centralised architecture are analysed in [61]. The procedures of each architecture as mentioned above are described in detail in [60] [61]. Two use cases using the ESM concept have been studied in [61]: capacity limited network use case which is to cope with peak time traffic demand and to be under-utilised in off-peak times, and eNB overlaid use case in which legacy systems like 2G/3G provide radio coverage together with E-UTRAN. More use cases using either centralised or distributed ESM concepts have been analysed in [62]. In [62], three categories are considering: intra-eNB energy saving, inter-eNB energy saving and inter-RAT energy saving. Comparisons of centralised and distributed ESM concepts for each category are also provided, with advantages and disadvantages analysed.

3.3.2 Energy Efficiency Efforts in EU Projects

The EU FP7 EARTH project aims at addressing the environmental challenges by investigating and proposing effective mechanisms to reduce energy wastage and improve energy efficiency of mobile broadband communication systems with 50% without compromising users' QoS and system capacity [63]. Many green radio technologies have been investigated by the EARTH project. This involves energy efficiency algorithms in physical layer to improve the transceivers for small cell/macro cell stations, technologies to improve antenna systems and MIMO technology. EARTH also explores the energy saving potential from network perspective by optimising cell sizes and network planning procedures.

Low traffic conditions allow saving power by switching off some base stations if the base stations remaining on can serve the demands from the users. Turning on/off base stations in the dense base station deployments is done in order to cope with both of the peak-time and off-peak time traffic demands. The cell on/off technology has been investigated in the EARTH project. The on/off decision is made based on a linear predictor, which requires instantaneous and historical data regarding load conditions in the network. And an machine to machine environment, acting over the multiple vehicles that drive on a highway during specific amount of time, is considered for validation. The results show that up to 25% peak energy savings can be achieved on the homogeneous cell type distribution, meanwhile the average energy saving is 14% in the case of randomised daily traffic curves. The results are similar to the ones obtained when using voice or data traffic, so it can be concluded that the particular application handling data traffic from vehicles on the highway is suitable to deploy ON/OFF mechanisms [63].

The cell on/off schemes have also been investigated in Heterogeneous Networks (HetNet) and multi-RAT networks by the EARTH project. Based on the network utilisation figures, the energy saving potential by switching off empty small cells of the capacity layer, including the traditional macro coverage and the macros with Remote Radio Head (RRH), is investigated. The results show that 25%-40% and 44%-58% energy saving can be achieved in 1Mbps and 100kbps cell bitrate guaranteed coverage layer respectively. And relatively more gain can be achieved in case of RRH (up to 7% on daily average), since in this case the energy consumption of the coverage layer is greatly reduced. Thus, any reduction of the capacity layer has a great impact on the total system wide energy consumption [63]. The signalling procedures to instruct the on/off decisions between eNBs have also been studied.

3.3.3 Other Works to Improve Energy Consumption in Mobile Backhaul Networks

Apart from the efforts of 3GPP and the European Commission, various standardisation bodies have also introduced energy efficiency specifications for emerging telecom systems. ETSI energy measures for wireless access equipment address energy efficiency within RANs [64]. For fixed access network, energy efficient schemes have been proposed by ETSI energy efficiency for telecommunication equipment [65], IETF power states [66], as well as ITU-T power saving GPON [67] and IEEE energy efficient Ethernet addressing equipment requirements and protocol specifics [68]. However, a network wide energy saving scheme combining both radio and fixed aspects is still missing.

In mobile networks the backhaul's contribution to the total power consumption is usually neglected because of its limited impact compared to that of the radio base stations. However, the exponentially increasing mobile data traffic indicates that backhaul networks will take a significant share of the energy consumption in future mobile networks. And their actual contribution will depend on the radio station deployment scenario as well as on the technology and topology of the backhaul itself [69]. The study in [70] has shown that the relative effect of the backhaul power consumption is non-negligible for scenarios with large number of small base stations. This means that the total power consumption of a heterogeneous network deployment to a larger degree is affected by the backhaul architecture. This raises the need for further investigation on how different architectural options for the backhaul affect the total backhaul power consumption.

In [69], an analysis of the power consumption for today's heterogeneous network scenarios, including the effect of two established backhauling technologies, i.e., fiber and microwave, and impact of backhaul network topologies, i.e., ring topology, star topology and tree topology, is provided. In order to evaluate the power consumption, two power models have been introduced: one for the microwave use case and one for the fibre use case. The results are leveraged to see how different architectural options for a given backhaul technology affect both the total backhaul and the total mobile network power consumption. Simulation results show that in today's heterogeneous networks, from a mere energy consumption perspective, fibre-based backhaul solution might be preferable due to its relatively low impact on extra power consumption.

3.3.4 Other Works to Improve Energy Consumption in Cellular Networks

In [71], a research about renewable energy resources is investigated; especially in locations in which electrical grids are not available or are unreliable (e.g., Africa, Northern Canada). In such places, renewable energy resources such as sustainable biofuels, solar and wind energy seem to be more viable options to reduce the overall network expenditure. Recently, a program called “Green Power for Mobile” to use renewable energy resources for Base Stations has been started by 25 leading telecoms including MTN Uganda and Zain, united under the Global Systems for Mobile communications Association (GSMA). This program is meant to aid the mobile industry to deploy solar, wind, or sustainable biofuels technologies to power 118,000 new and existing off-grid BSs in developing countries by 2012. Powering that many BSs on renewable energy would save up to 2.5 billion litres of diesel per annum (0.35% of global diesel consumption of 700 billion litres per annum) and cut annual carbon emissions by up to 6.8 million tonnes.

In addition, regarding the renewable energy saving, because of the limited battery capacity, generated energy should be utilised well to avoid energy overflow [72]. Therefore, how to optimise the utilisation of renewable energy in cellular networks is not a trivial problem. By envisioning future BSs powered by multiple types of energy sources (e.g., on-grid energy and renewable energy), it is investigated the optimal energy utilisation strategies in cellular networks with both on-grid and off-grid BSs. In such cellular networks, BSs are powered by renewable energy if they have enough renewable energy storage; otherwise, the BSs switch to on-grid energy to serve mobile users. To optimise energy utilization, BSs cooperatively determine their cell size through energy-aware cell breathing based on the proposed cell size optimisation (CSO) algorithm [71] [73]. The core idea of the CSO algorithm is that optimises the cell sizes of BSs and maximises the utilisation of renewable energy in order to reduce the on-grid energy consumption.

Furthermore, in [72], the authors state that coordinated multipoint (CoMP) transmission (multiple BSs cooperatively transmit data to mobile users to improve their receiving signal quality) is a key technology for future mobile communication systems. Three different techniques are analysed: (i) **joint transmission** (exploits the cooperation among neighbouring BSs to transmit the same information to individual users located at the cell edge [74]), (ii) **cooperative beamforming** (a BS forms radio beams to enhance the signal strength of its serving users while forming null steering toward users in its neighbouring cells) and (iii) **cooperative relaying** (by taking advantage of cognitive radio techniques, cooperation can be exploited between a primary cell and a secondary cell).

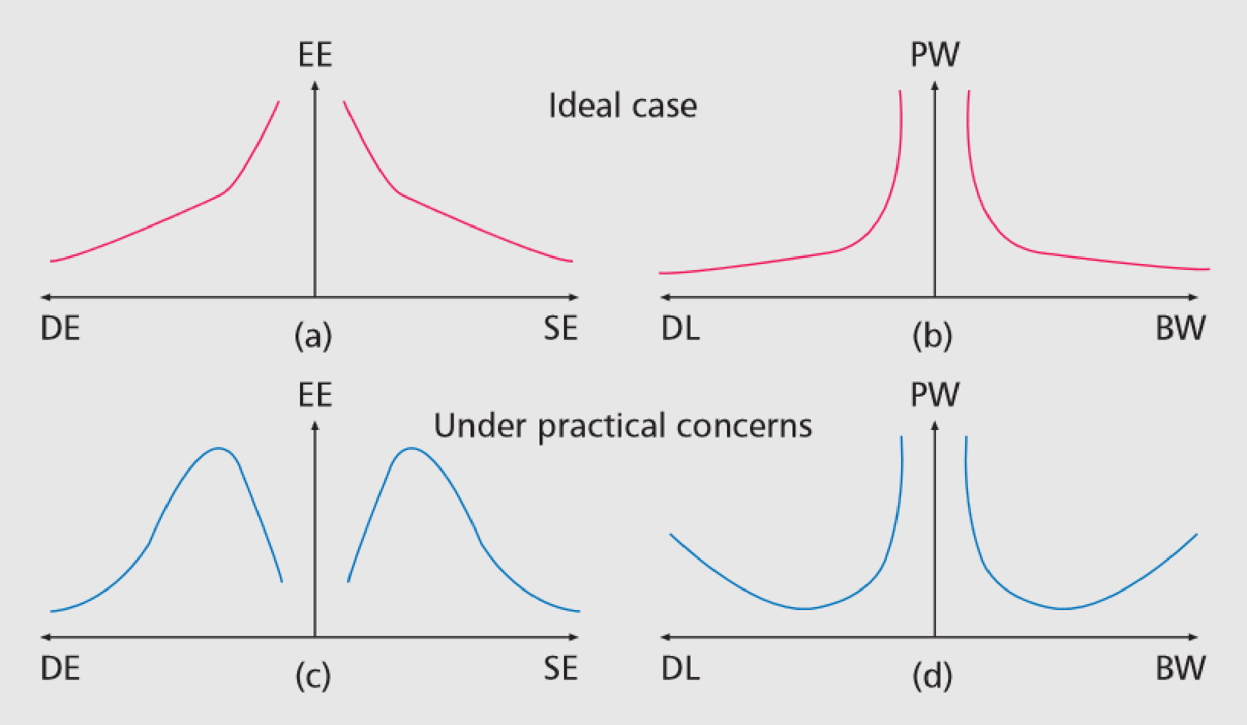

Figure 3-3 Sketch of the four trade-offs relations without and with practical concerns networks [75]

Finally, we underline that there are four *basic* trade-offs on green wireless networks [75], which are also depicted in Figure 3-3:

- Deployment efficiency (DE)–energy efficiency (EE) trade-off: to balance the deployment cost, throughput, and energy consumption in the network as a whole.
- Spectrum efficiency (SE)–EE trade-off: given bandwidth available, to balance the achievable rate and energy consumption of the system.
- Bandwidth (BW)–power (PW) trade-off: given a target transmission rate, to balance the bandwidth utilised and the power needed for transmission.
- Delay (DL)–PW trade-off: to balance the average end-to-end service delay and average power consumed in transmission.

By means of the four trade-offs, key network performance/cost indicators are all strung together.

The iJOIN project attempts to go further beyond the current research work on energy optimization, as our solution involves not only the use of SDN to address the energy saving problem from a network wide perspective, but also to handle the interactions between energy optimization solution and other mechanisms such as mobility management, routing/congestion control and load balancing in a mobile backhaul.

3.4 Congestion Control

The objective of congestion control is to allow network operators to simultaneously achieve high throughput and low average delay. This candidate technology (CT 4.4) analyses the impact of the support of advanced congestion control algorithms and their impact on the performance of the backhaul network in the context of the proposed iJOIN architecture. In this sense, it is important to distinguish between active queue management (AQM) techniques deployed inside the network that, e.g., aim at reducing TCP sending rate by intentional packet drop, and low priority congestion control (LPCC) solutions, that intend to transfer a flow to a lower priority by reacting faster to network congestion using indicators other than packet loss.

The most representative LPCC solution according to IETF, Low Extra Delay Background Transport (LEDBAT) is an experimental delay-based congestion control algorithm that seeks to utilise the available bandwidth on an end-to-end path while limiting the consequent increase in queuing delay on that path. LEDBAT uses changes in one-way delay measurements to limit congestion that the flow itself induces in the network. LEDBAT is designed for use by background bulk-transfer applications to be no more aggressive than standard TCP congestion control (as specified in RFC5681) and to yield in the presence of competing flows, thus limiting interference with the network performance of competing flows.

LEDBAT employs one-way delay measurements to estimate the queuing delay which may indicate that a link is in congestion. End-to-end delay can be decomposed into transmission (or serialization) delay, propagation (or speed-of-light) delay, queuing delay, and processing delay. On any given path, barring some noise, all delay components except for queuing delay are constant. To observe an increase in the queuing delay in the network, a LEDBAT sender separates the queuing delay component from the rest of the end-to-end delay. The latter constitutes the base delay, which is the minimum delay that can be observed on the end-to-end path.

To respond to true changes in the base delay, as can be caused by a route change, LEDBAT uses only recent measurements in estimating the base delay. The duration of the observation window itself is a trade-off between robustness of measurement and responsiveness to change, i.e., a larger observation window increases the chances that the true base delay will be detected (as long as the true base delay is unchanged), whereas a smaller observation window results in faster response to true changes in the base delay.

Assuming that the base delay is constant (in the absence of any route changes), the queuing delay is represented by the variable component of the measured end-to-end delay. LEDBAT measures queuing delay as simply the difference between an end-to-end delay measurement and the current estimate of base delay. The queuing delay should be filtered (depending on the usage scenario) to eliminate noise in the delay estimation, such as due to spikes in processing delay at a node on the path.

LEDBAT can be used as part of a transport protocol or as part of an application, as long as the data transmission mechanisms are capable of carrying timestamps and acknowledging data frequently. LEDBAT can be used with TCP, Stream Control Transmission Protocol (SCTP), and Datagram Congestion Control Protocol (DCCP) with appropriate extensions where necessary. It can be further used with proprietary application protocols such as those built on top of UDP for peer-to-peer (P2P) applications.

It seems reasonable to assume that, although LEDBAT should be implemented at the edges of the connection, its main impact will be observed at the bottleneck link in the connection. And it can be assumed that, in a high percentage of cases, this bottleneck link will be either the radio interface or the last mile backhaul link to the eNB.

The interaction of the LEDBAT congestion control with the radio interface radio management functionalities is very difficult to assess. Radio Resource Management (RRM) mechanisms are expected to act at a different, much shorter time scale than end-to-end congestion control mechanisms. Furthermore, the implementation of some RRM functionalities is vendor dependent, which makes it complicated to determine the impact that they may have on LEDBAT performance.

For LPCC mechanisms, the technical solution requires to implement mechanisms that keep baseline delay as constant as possible when there are no congestion issues to be solved. They may also modify them in order to activate the congestion control mechanisms in a preventive way.

Active Queue Management (AQM) schemes like RED [76] or REM [77] randomly drop or mark packets before the buffer of a network nodes becomes full. Hence, TCP senders can be notified to avoid excessive growth of queues in buffers. AQM is meant to be a general mechanism using one of several alternatives for congestion indication, but in the absence of ECN, AQM is restricted to using packet drops as a mechanism for congestion indication. In this case, AQM drops packets based on the average queue length exceeding a threshold, rather than only when the queue overflows.

ECN based AQM schemes use the two least significant (right-most) bits of the DiffServ field in the IPv4 or IPv6 header to encode four different codepoints:

- 00: Non ECN-Capable Transport — Non-ECT
- 10: ECN Capable Transport — ECT(0)
- 01: ECN Capable Transport — ECT(1)
- 11: Congestion Encountered — CE

If both endpoints support ECN, they mark their packets with ECT(0) or ECT(1). If the packet traverses an AQM queue that is experiencing congestion and the corresponding router supports ECN, it may change the codepoint to CE instead of dropping the packet. This process is referred to as “marking” and its purpose is to inform the receiving endpoint of a potential congestion. At the receiving endpoint, this congestion indication is handled by the upper layer protocol (transport layer protocol) and needs to be echoed back to the transmitting node in order to signal it to reduce its transmission rate.

The use of ECN based AQM for mobile networks it is being explored in the context of the 3GPP Release-12 Study Item UPCON (User Plane Congestion) [78]. The objective of UPCON is to improve resource efficiency in the network and to increase the number of active users while maintaining good user experience (QoE).

The way that the ECN procedures would be supported in a mobile network is illustrated in Figure 3-4.

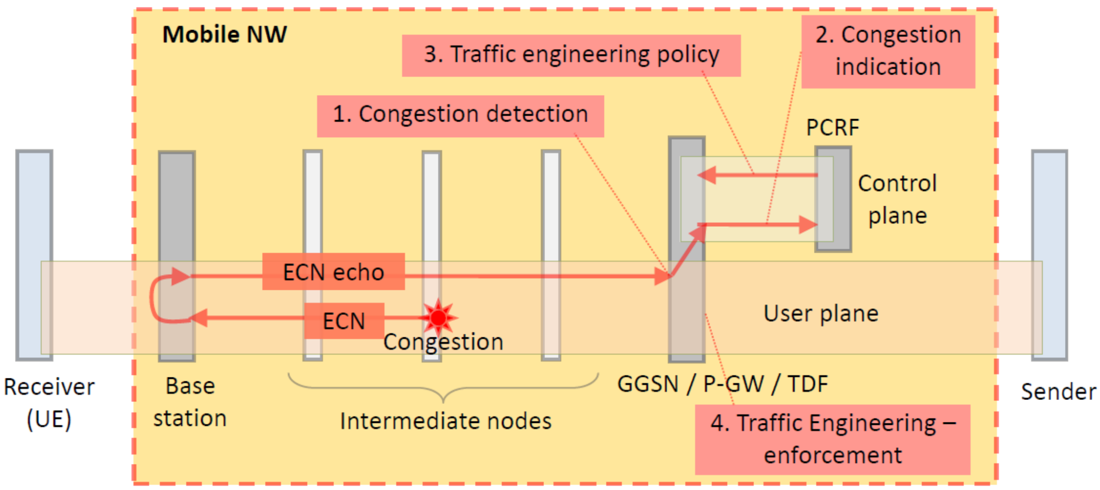

Figure 3-4 Congestion control in a mobile network

This represents a different approach to the standard one, where ECN functionalities are implemented by the edge devices. In the mobile network they are most likely to be implemented in the eNB and the P-GW. On top of this, as the LTE network may use two tunnelling protocols, Proxy-Mobile-IP (PMIP) and GPRS Tunnelling Protocol (GTP), Congestion Indication should be based on GTP/PMIP-level ECN-ECHO.

There are also other differences: whilst the traditional approach in IP networks is to allow TCP congestion control mechanisms to deal with congestion situations, in the approach supported in UPCON congestion is dealt by means of traffic engineering enforcement procedures in the P-GW (based on the policy established by the PCRF). This approach has the advantage of taking into account not only the congestion level, but also the subscriber's profile, when implementing remedial procedures.

As indicated above, different problems should be tackled with depending on the kind of congestion protocol mechanism that is being implemented.

3.5 SON in Backhaul Networks

Self-optimised Networking (SON) refers to the automated operation of networks, in particular mobile networks [79]. The objective of SON is to increase the automation of RAN operation including installation, maintenance, and fail recovery. Among the best known and explored SON functionality are automatic inventory, software download, neighbour relation, cell id assignment, Coverage and Capacity Optimization (CCO), mobility optimisation, Random Access Channel (RACH) optimisation, and load balancing. SON can be divided into different phases:

- Network planning, topology management and documentation,
- Service provisioning and change management,
- Service testing and verification,
- Reporting and network maintenance, and
- Fault and availability management.

So far, SON only played a role to manage the radio access network efficiently but it was mostly disregarded for the management of the backhaul network. SON may be part of the solution to provide the required backhaul capacity for 4G networks, as backhaul networks become a significant bottleneck. SON functions for backhaul networks include [79]:

- Planning: Network configuration and backhaul dimensioning.
- Deployment: Configuration, testing and verification including side-effects, i.e., interaction of neighbouring network elements in RAN and backhaul network.
- Optimisation: Adaptation to traffic characteristics including concepts which may anticipate bottlenecks based on pattern analysis.
- Maintenance: Fast problem identification and recovery due to “avalanche risk,” i.e., failure of one backhaul node or link affects a set of base stations. Also fast integration of new technology is considered.

SON will be part of the solution to implement the convergence towards Ethernet/IP based backhaul networks, which is required for the RAN evolution towards 4G and later technology eras.

3.6 Analysis of Backhaul Networks based on Stochastic Geometry

Stochastic geometry is an area of mathematical research that seeks to provide models and methods to analyse complicated geometric patterns that occur in many areas of science and technology such as geology, biology, material sciences, etc. One of the first instances of applications to wireless communication networks was shown by Baccelli et.al. in 2001[80]. A comprehensive extension to wireless communication networks was seen in their monograph [81] from 2009. A couple of papers from last year, [82] [83], are the first to use stochastic geometry for energy efficiency and energy consumption analysis.

The usefulness of such models to ascertain information about real world deployments was seen in [84], where the theoretic results for the probability of coverage obtained by the authors concurred with the probability of coverage in a real world deployment (information acquired from a US based service provider). Heterogeneous networks can also be described using this method of analysis. Andrews et.al in [85] model a heterogeneous network using a multi-layered Poisson point process, whereas [86] models a heterogeneous network consisting of macro and micro base stations using a Poisson cluster process. These papers show that it is possible to obtain relationships between the probability of coverage and the base station densities in an area. The frameworks illustrated also indicate that user density dependent base station densities can also be used. Figure 3-5 illustrates a realization of a deployment which is modelled by a point process.

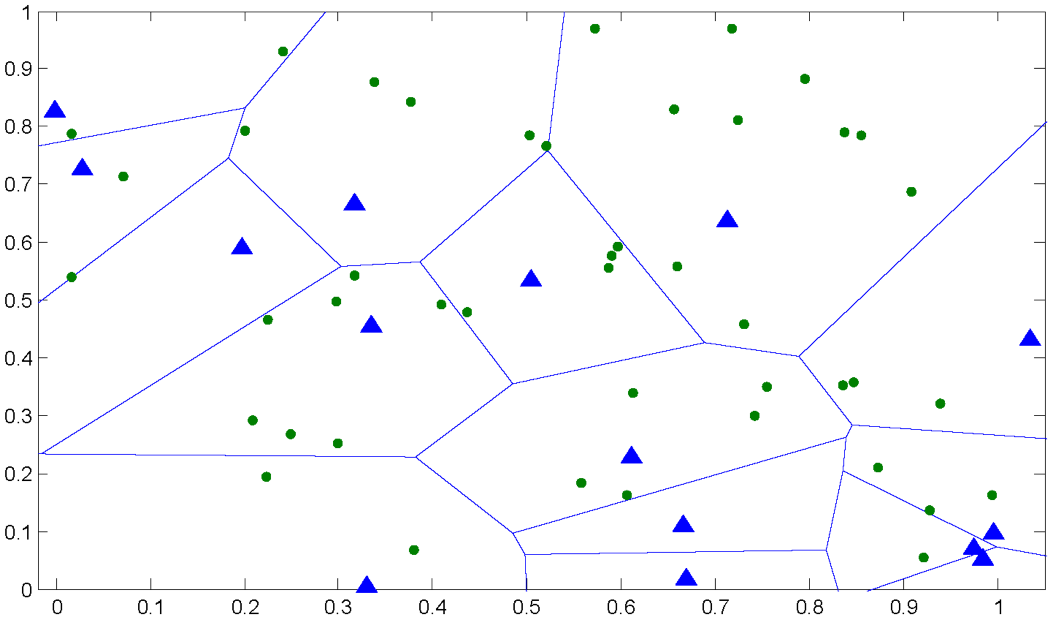

Figure 3-5 Realization of a point process

From all of the works listed above, it is seen that stochastic geometry proves to be a useful tool in providing simple relationships between system parameters of interest and their performance indicators, which can then be used to optimise certain efficiency or effectiveness parameters (such as CAPEX, OPEX, energy efficiency, etc.).

Despite these advantages, to the best of our knowledge to date, there have not been any works that deal with an analysis of the backhaul using stochastic geometry. The investigations carried out in this project aim to fill this gap.

3.7 Load Aware Network Architecture

In real word systems, mobile users (UEs) are not evenly distributed across cells, and different types of data requests traverse the network. Thus, the traffic across the access and backhaul nodes (in iJoin iSC and iTN, respectively), is not evenly distributed as well. So, Load Balancing becomes one of the most active and emerging fields of research in Cellular networks. In order to balance the load among different nodes, it is needed to transfer the over-loaded traffic from “hot” (i.e. nodes with a high amount of traffic) nodes to neighbouring “cooler” ones (i.e. nodes with less traffic).

Works to Improve Load Balancing in Cellular Networks

Different works and algorithms are suggested to improve the load balancing, in cellular networks [87]. Thus, in order to balance the load among different cells, it is needed to transfer the over-loaded traffic from “hot” cells to neighboring cooler ones. Various dynamic load balancing schemes to deal with the unbalanced traffic problem are proposed, and we are going to mention a few of them.

First of all, a dynamic load balancing scheme iCAR employs ad hoc relay stations (ARS's) in the cellular network to balance traffic loads efficiently and to share channels between cells via primary and secondary relaying. These ARS's operate in the unlicensed ISM band and, therefore, do not cause interference to the cellular band. Regardless of the number and position of hot cells, an iCAR based algorithm [88] for a k-tier system is developed.

The authors in [89] suggest that the system performance can be improved for non-uniformly distributed call traffic by controlling cell-size. Using smart base station antennas a scheme is proposed [89] which dynamically changes cellular coverage size and shapes according to geographic traffic distribution.
Another load balancing approach [90] uses online bandwidth management algorithms for multimedia cellular networks which try to minimize the maximum available bandwidth among cells in an adaptive manner based on current network traffic conditions.

Finally, a self-organizing load balancing [91] provides self-optimizing load balancing policies to improve adaptation and robustness of Fixed Relay Station (FRS) based cellular networks. The framework proposes a Self-organizing Cooperative Partner Cluster (SCPC) concept to dynamically select optimal partners of each BS and RS. A novel Comprehensive Load Balancing Policy Stack (CLBPS) is also proposed to utilize merits of various load balancing policies.

Other works for load balancing have been done as well, regarding game-theoretic approaches [92], store-carry and forward relaying [93] and an Integrated Multi-hop Cellular Data Networks (IMCDN) [94].

3.8 Traffic Management in Backhaul Networks

As networks become more complex and traffic diversity increases, there is the apparent need to manage the traffic carried by the network. Traffic management has the goal to maximise the utilization of the network by minimising the maximum link utilisation as well as the objective to avoid congestion in the network. One means is to automate path selection procedures and to optimise traffic utilization because only adding new capacity does not suffice but it is necessary to manage these resources efficiently.

The most widely used protocol for path selection is Open Shortest Path First (OSPF) which assigns weights to links and computes the shortest path across the network (or the path with lowest weight). Even small changes to link weights may have an avalanche effect causing significant impacts to the network [95]. Among others, the selection of the optimal path should take into account QoS as well as dynamic metrics such as available bandwidth, delay, reliability, jitter, and mobility aspects (such as proximity to nodes that can act as anchor/offloading points for certain types of flows).

An alternative to OSPF is Multiprotocol Label Switching (MPLS) which creates explicit paths through an IP network. Using small labels which identify the route through the network, MPLS allows for fast routing and is able to handle heterogeneous networks (which requires a tight integration of MPLS and the underlying physical network). Computing the optimal route is computational complex and may not be done online, e.g. using Genetic Algorithms [96]. Hence, a combination of offline and online algorithms may be used, as presented in [97]. The authors in [97] introduced an algorithm which uses the expected traffic matrix to derive an optimal network topology offline, and an online component applies small, local changes based on dynamic traffic requests. It is further possible to combine OSPF and MPLS as described in [98] where MPLS paths are used to distribute the traffic and OSPF is used locally which implies less frequent and less severe changes of link weights.

In iJOIN, we are going to follow an approach based on SDN for the computation and creation of the paths in the backhaul. We will analyse the advantages brought by the logical centralisation provided by the use of SDN techniques, as well as identify the potential drawbacks (e.g., scalability, speed of reaction) that it could involve.

4 Description of Candidate Technologies

WP4 architecture and candidate technologies are heavily influenced by the SDN approach, which has a strong link with iJOIN Network Controller (iNC). The iNC is a key network entity in charge of configuring, monitoring and driving the operation of the rest of the RAN and backhaul network entities (more details are provided in Section 5). The modules running on iNC and the mapping of the module and candidate technology are summarised in Table 4-1. Please be noted that CT4.6 is not included here as it is not a run-time mechanism. More details of the interaction between the modules will be introduced in section 5.

CT	Topic	Module	Module Abbreviation
4.1	Distributed IP Anchoring and Mobility Management	Anchor and Mobility Management	AMM
4.2	Network Wide Energy Optimisation	Network-wide Energy Optimiser	NEO
4.3	Joint Path Management and Topology Control	Traffic Engineering Enforcement Module	TEEM
		Measurement Module	MM
4.4	Routing and Congestion Control Mechanisms	Routing and Congestion Control	RAC
4.5	Network Wide Scheduling and Load Balancing	Traffic Engineering Enforcement Module	TEEM

Table 4-1Summarisation of WP4 candidate technologies and modules

4.1 CT 4.1: Distributed IP Anchoring and Mobility Management

4.1.1 Motivation

Mobility Management in general is a set of tasks for controlling and supervising UEs in a wireless network to locate them for delivery services, as well as, to maintain their connections while they are on move. Mobility management is concerned with many aspects, such as Quality of Service (QoS), power management, location management, handoff management, and admission control. It is one of the most critical features in wireless communications due to the direct effect on user’s Quality of Experience (QoE), network performance and power consumption. The core components of mobility management are location management, handoff management and the smart selection of network access.

In the past few years we have been witnessing an extraordinary data throughput explosion in cellular networks. Telecom operators have been carefully monitoring the disconnection between the average revenue per user (ARPU) and the associated cash costs per user (CCPU). Despite the remarkable volume increase of broadband data over mobile networks, mobile data revenue is falling fast.

This has some serious impacts on the dimensioning and planning of mobile networks. Specifically, we note that a) spectrum is limited and expensive, so available bandwidth for the access network cannot be easily increased; and b) deployed mobile core networks are highly hierarchical and centralised, which introduces serious scalability and reliability issues. We are tackling both, by looking at more densely deployed cells and enabling selected IP flows not to traverse the backhaul and mobile network operator’s core. This is in line with several on-going 3GPP efforts, namely, the IP Flow Mobility and Seamless Offload (IFOM), the Local IP Access (LIPA), and the Selected IP Traffic Offload (SIPTO).

We also aim at mitigating concern b) above by investigating solutions that allow for distributing the data anchoring and mobility support. With local breakout and offload solutions, the advantages come from the fact of selecting an offloading node for certain flows (unless there is an additional access technology deployed, such as WLAN, which is out of scope of the iJOIN project). This offloading node acts as anchor for the selected traffic, but it may happen that if the user moves, traffic cannot be seamlessly forwarded to the new location. With a true distribution of the IP mobility management, the goal is to fully support user mobility.

Due to the different needs of applications in terms of mobility support and the existence of different IP mobility solutions, there has been recently a trend to enhance the semantics of the IP addresses, and in general also the router discovery. Thus, applications and connection manager running on the mobile terminals are capable of selecting the most appropriate IP address and router on a per application basis. On the presence of local offloading nodes, a given mobile node could be allocated different types of IP addresses (anchored at different routers), and this needs to be taken into account by the mobile terminal when deciding which address should be used.

These goals are related to standardisation activities within the IETF, namely at the NETEXT and DMM working groups.

The use of smaller cells is one of the approaches followed by iJOIN to increase the overall bandwidth capacity available to users. Whereas it is a well-known solution capable of providing significant enhancements, it also raises significant challenges in other areas, such as network selection and handover management. Decisions about which is the best radio access point to connect to are no longer mainly based on received signal strength, but they also need to take into account other multiple disparate aspects, such as: backhaul status, support for local breakout/offload, distributed anchoring, terminal/application specific aspects (e.g., mobility patterns, session lifetime, address continuity requirements, etc). A comprehensive solution is required to optimize the handover procedure, aiming at reducing packet loss, latency and minimising signalling overhead as much as possible. Thus, fast and seamless user experience can be achieved. The mobility management design in iJOIN will go beyond improving network discovery and selection mechanisms, but improves signalling protocols. The SDN-based approach followed in iJOIN will provide the possibility for enhanced mobility management. The available real-time load information and logically centralised control enabled by our SDN-centric view can enable more efficient load-aware handover management schemes, rather than the current 3GPP handover procedure. For example, handover decision can be optimised by taking into account the load information from neighbouring small cells.

4.1.2 Assumptions

Figure 4-1 shows a generic reference scenario (already showing iJOIN network entities). With this regard, we consider the following general assumptions:

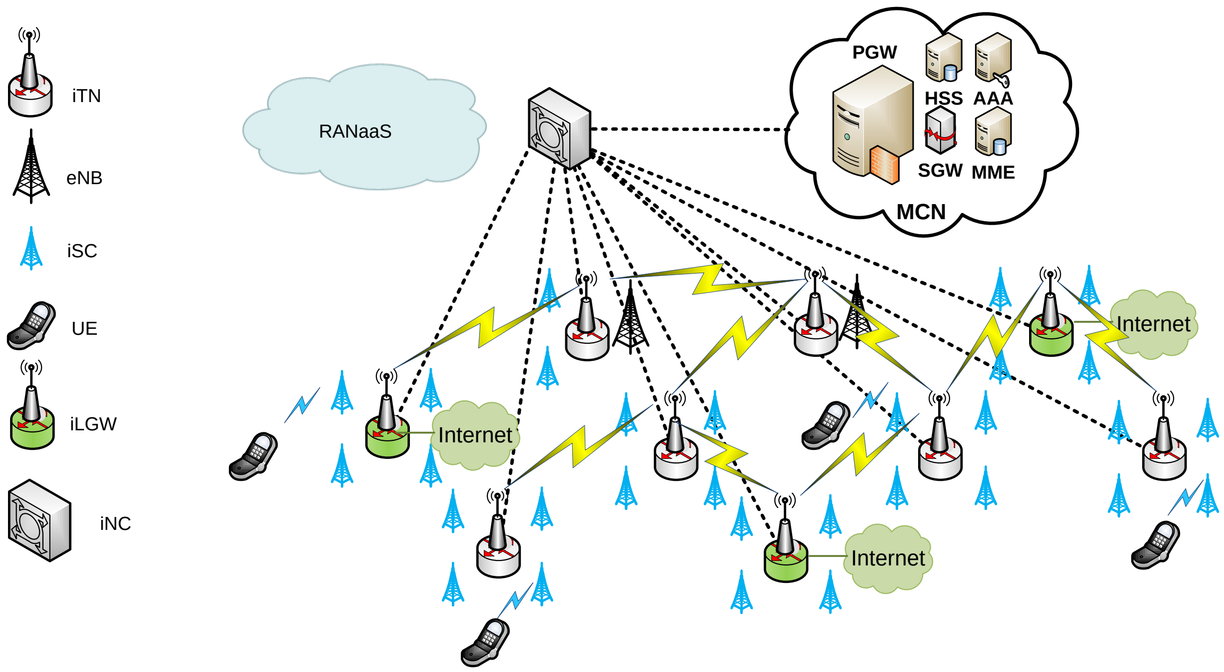

Figure 4-1 Reference scenario for IP mobility related technologies

- The backhaul is a multi-hop IP network, meaning that there might be more than one node in the path between the radio head and the mobile operator's core, and that some (if not all) of these nodes may be different IP hops. This, for instance, enables an easier integration of heterogeneous backhauling technologies. It will also allow for heterogeneous backhaul technologies, such as fibre and in/out-band connection between macro cell and small cells.

- Some nodes in the radio access network or in the backhaul might have local IP connectivity, which can even provide Internet access. These nodes can be used as offloading nodes for some selected flows, alleviating the load of the backhaul and core network.
- Some RAN and transport nodes are also able to provide mobility management (with low or none support from the core network). These nodes have control interfaces with the mobility entities in the mobile operator's core (e.g., MME, HSS, etc.).

4.1.3 Technical Requirements

Although there are various works on IP offloading in the context of a generic network as well as in 3GPP architectures, the iJOIN architecture presents several particularities that require these solutions to be revisited or even reworked from scratch. Among these challenges, we can highlight the following: *i)* the backhaul network is dynamic and can self-configure itself to better adapt to the user traffic demand or to improve energy efficiency; *ii)* an SDN-focused approach is followed to jointly optimise and control the RAN and backhaul. Based on this, we list the main technical requirements of the solutions for this particular use case with respect to the network-layer:

- The solution should be IP based to allow its operation with different wireless backhauling technologies.
- The solution has to both consider the radio access and backhaul, and might need user terminal support. This would allow a certain user terminal for selecting the best radio access point based on whether local breakout is available (for the traffic the UE is sending/receiving).
- The designed mechanisms have to interact with the SDN based backhaul routing and traffic engineering functions, so the path between the wireless radio head where the user is attached to and the offloading node can be dynamically set up and modified. This also allows performing energy optimisations.
- Backhauling capacity has to be taken into account, so the node selected to perform IP anchoring functionality can be selected based on the actual and current network load. Note that since the selected node has to remain playing the function of IP anchor while the user flow is alive, mobility considerations about where the UE might roam, should also be considered.
- The solution has to cooperate with the energy efficiency mechanisms since a decision of switching on/off a node might have an impact on mobility.
- The solution may be completely network-based, completely terminal-based or network-aided, user-based (hybrid approach).
- Not all the traffic might need mobility management (i.e., address continuity). Solutions should try to provide only mobility support to those that require it, and just offload (if possible) traffic that can survive an IP address change. This is actually related to the previous use case, as it might be the case that the networks prefers to switch certain flows to a different anchor/offloading node, because of backhauling capacity reasons, and it might be more efficient to do it for applications that can cope with an IP address change on their own.
- Handover preparation and execution mechanisms should be very fast in order to minimise packet loss.
- Solutions need to be applicable to overlay scenarios where macro and small cells co-exist and small-cells may leverage the support of macro-cells.

4.1.4 Description of Candidate Technology

This technology candidate explores two paths which provide two different solutions which are, nevertheless, harmonised: Distributed Mobility Management and the Virtual Cell Concept. Both are described in further detail next.

4.1.4.1 Distributed Anchoring and Mobility Management

Overview

Figure 4-2 shows an outline of the solution to provide a dynamic IP distributed mobility support with offloading support. The key aspects of the solution are:

- The network has several nodes in the access and in the backhaul that can perform the role of IP mobility anchors or offloading nodes.
- The UE and the network, upon initial start-up, and on an application basis, select the best radio access point of attachment and anchor for its traffic.
- If the UE moves and changes its serving cell (i.e., attaches to a different radio access point), some of the existing flows might need to be provided with mobility support. This is done by dynamically establishing/updating tunnels between the current radio access point and the original anchor of each IP flow. Note that this requires interaction with the backhaul routing and traffic engineering functions, to ensure that each flow is guaranteed the necessary quality of service. In some cases, tunnelling can be avoided if the routing function is capable of performing the required traffic redirection.
- The required control signalling to enable this dynamic and smart IP anchoring functions will benefit from the logically centralised SDN-based infrastructure.

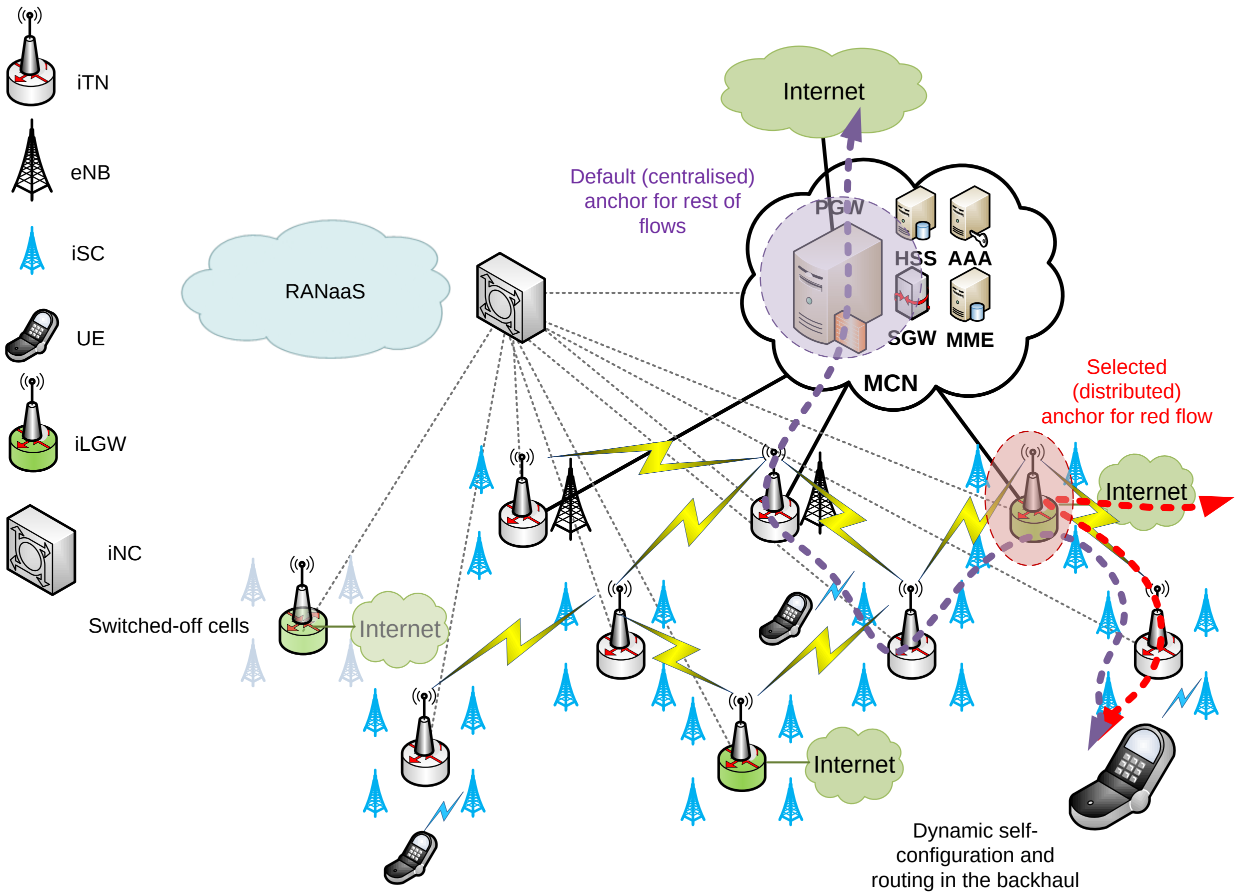

Figure 4-2 Outline of the DMM solution

Mobility procedures

Initial UE attachment

A high level Message Sequence Chart (MSC) of procedures when a user equipment (UE) attaches to the network and requests a PDN connection request is shown in Figure 4-3, focusing only on the anchor and mobility module (AMM) and its interactions.

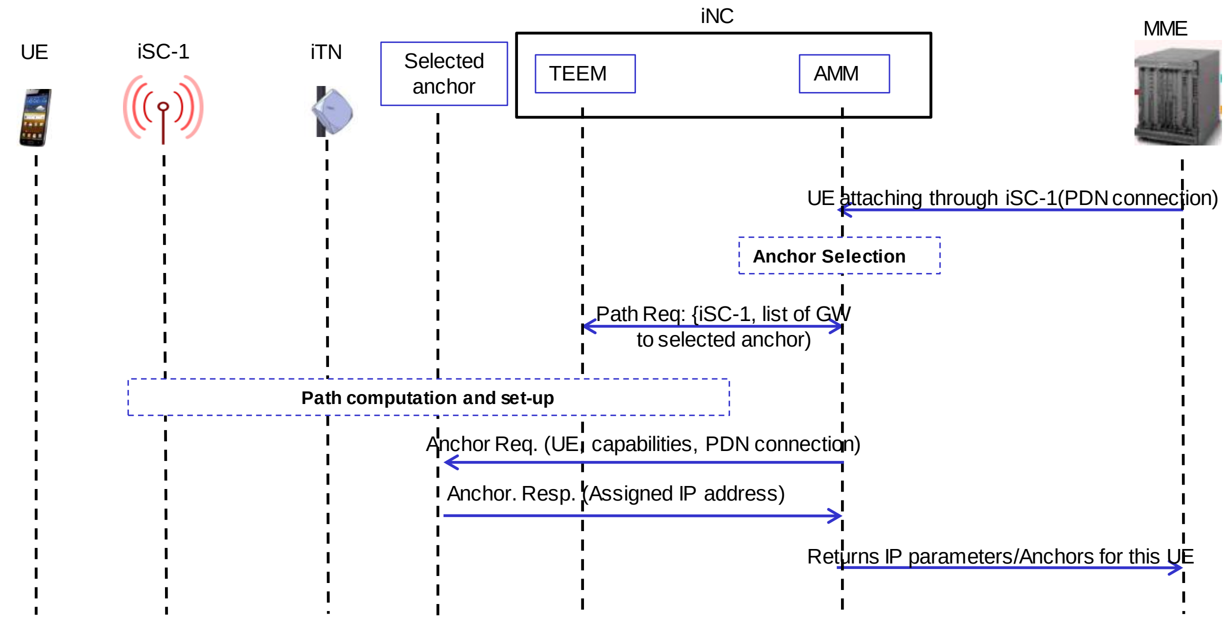

Figure 4-3 MSC of the attachment procedure (anchor and mobility entities)

In case of a complete attachment (not just a new PDN connection request), after LTE attachment and authentication procedures, the MME receives the PDN request, which triggers it to contact the AMM module at the iNC in charge of the RAN where the UE has attached. The AMM proceeds then to compute a prioritised list of anchors, following the algorithm shown in Figure 4-4.

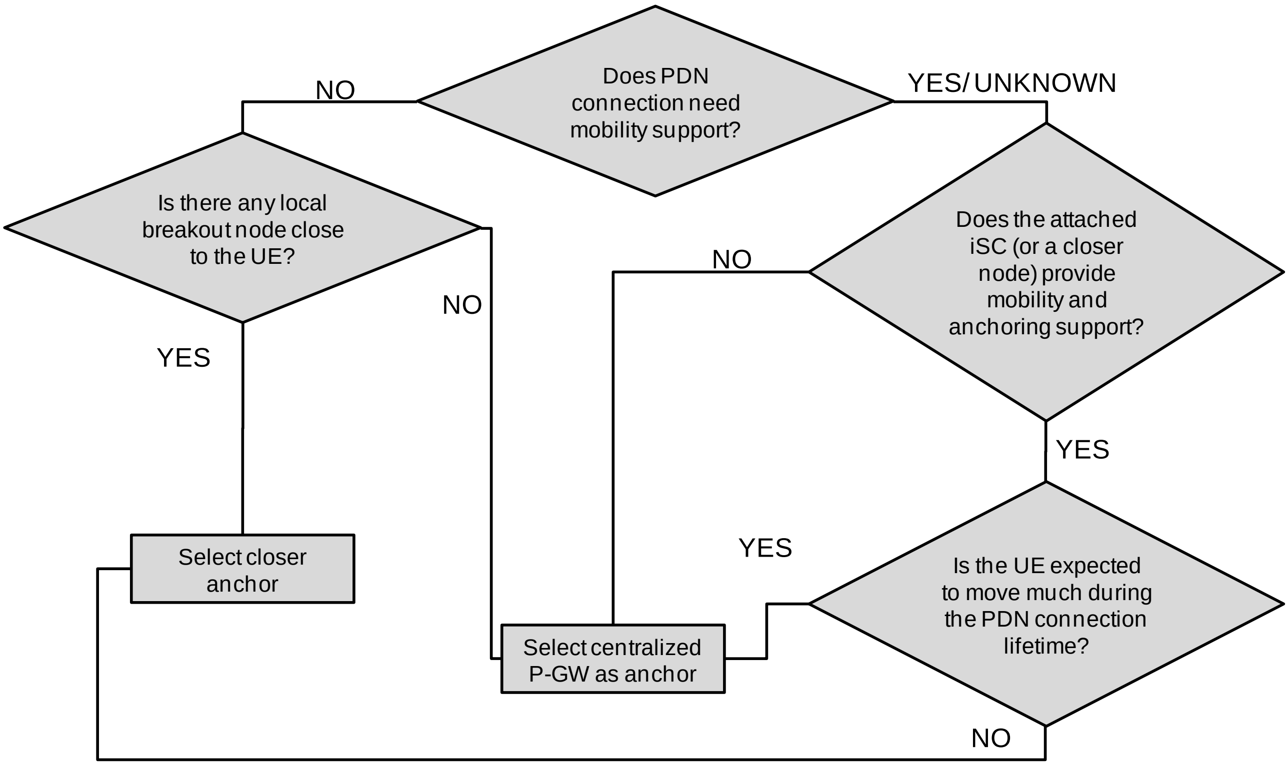

Figure 4-4 Anchor selection algorithm (simplified view)

Once a list of anchors has been selected, the AMM proceeds to request the establishment of the required paths to the TEEM, which needs to check if it would be possible to set-up a path between the iSC and the gateway to the selected anchor, starting for the most preferred anchor. Once the TEEM has set up the path in the backhaul for the selected gateway, the AMM contacts the anchor and requests to provide anchoring services for the UE, providing also information about its capabilities (in terms for example of mobility management) and the characteristics of the requested PDN connection. The selected anchor allocates an IP address/prefix to the UE for the requested PDN connection and informs back the AMM module, which updates the MME.

Intra anchor UE mobility

By intra-anchor handover, we define a UE's change of point of attachment (i.e., connection to a new iSC) which does not involve the new selection of a new anchor. Figure 4-5 shows the simplified signalling operation. As for the attachment case, the MME interacts with the AMM module running on the iNC to inform about a change of point of attachment of a UE. The AMM evaluates whether a new anchor assignment is necessary (just considering mobility and application requirements). In this example, no change is required, which means that the same anchor is provided back as the most preferred one. Since the iSC serving the UE has changed, the AMM needs to contact the TEEM module so the forwarding path is updated in the backhaul and RAN.

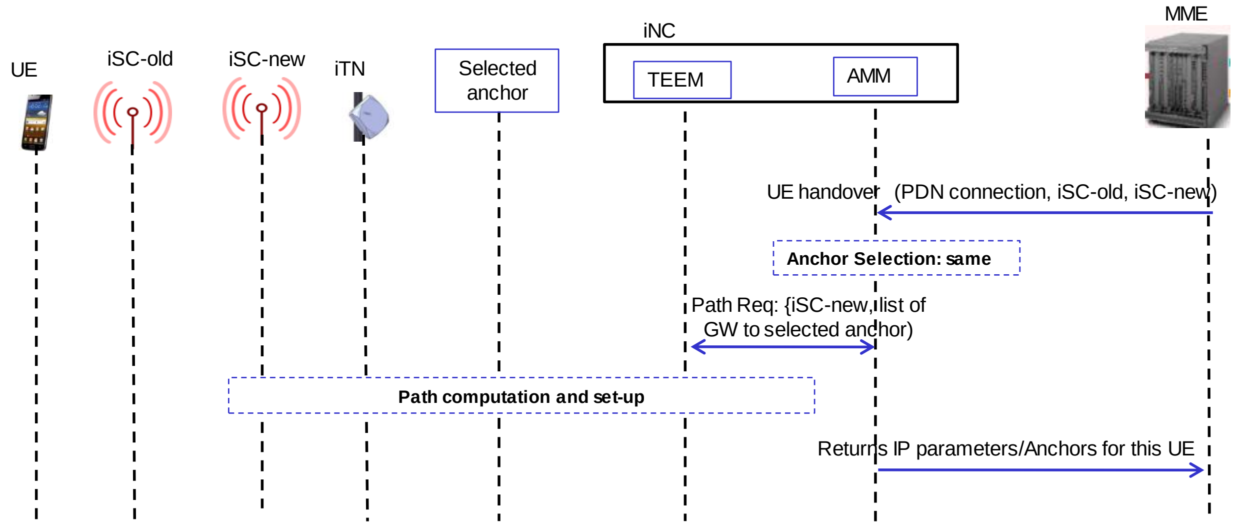

Figure 4-5 MSC of the intra-anchor handover procedure (anchor and mobility entities)

Inter anchor UE mobility

We define by inter-anchor handover, a UE's change of point of attachment (i.e., connection to a new iSC) which involves the new selection of a new anchor. Figure 4-6 shows the simplified signalling operation. Note that the new anchor will only by used by the UE for new communications (ongoing ones that need to be provided with IP address continuity remain anchored at the old anchor).

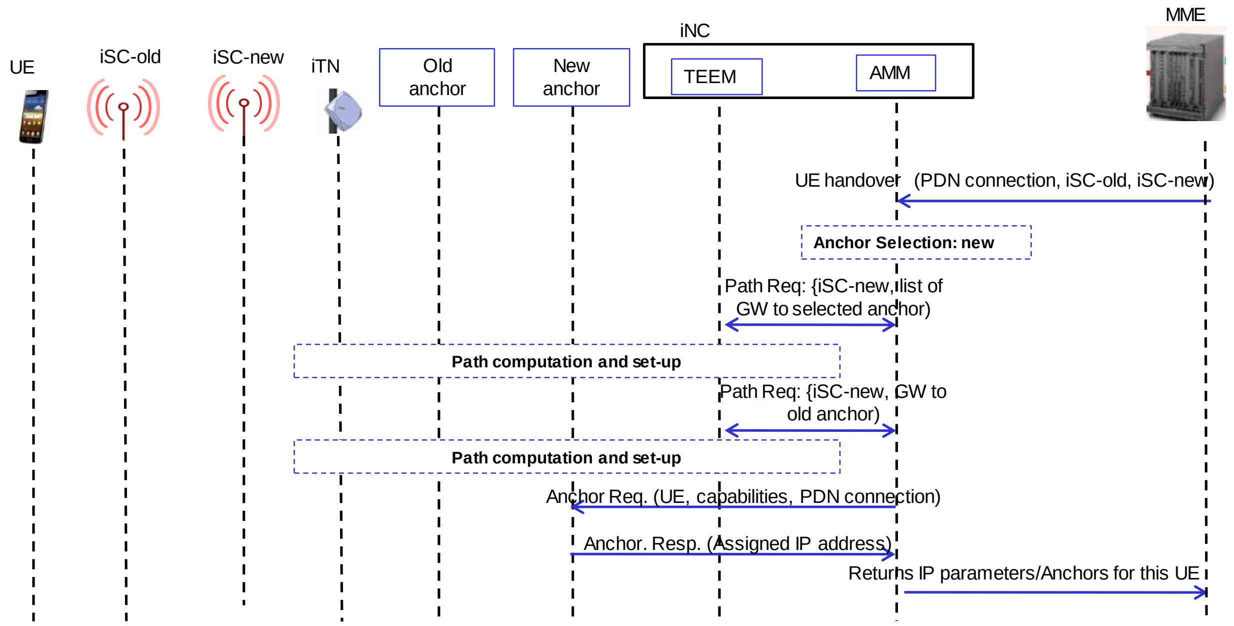

Figure 4-6 MSC of the inter-anchor handover procedure (anchor and mobility entities)

As for the attachment case, the MME interacts with the AMM module running on the iNC to inform about a change of point of attachment of a UE. The AMM evaluates whether a new anchor assignment is necessary (just considering mobility and application requirements). In this example, a different anchor is selected (for example because the new iSC has local-breakout capabilities). The AMM module requests the TEEM module to compute a path for the path between the new iSC and the new selected anchor, but also between the new iSC and the old anchor. This ensures that both new communications (anchored at the new anchor) and old communications (anchored at the old anchor) can take place.

New anchor assignment

Even if the UE does not physically move and change its attachment to a new iSC, the network might decide to assign a new anchor to the UE, so once ongoing communications are finished, the old anchor(s) are no longer used by the UE (this might be needed for example to switch off the old anchors). In this case, the AMM module would receive a request from the NEO or RAC modules to assign a new anchor. The required operations are then very similar to the ones for the inter-anchor handover, though using the same iSC. Figure 4-7 shows the simplified signalling of this case.

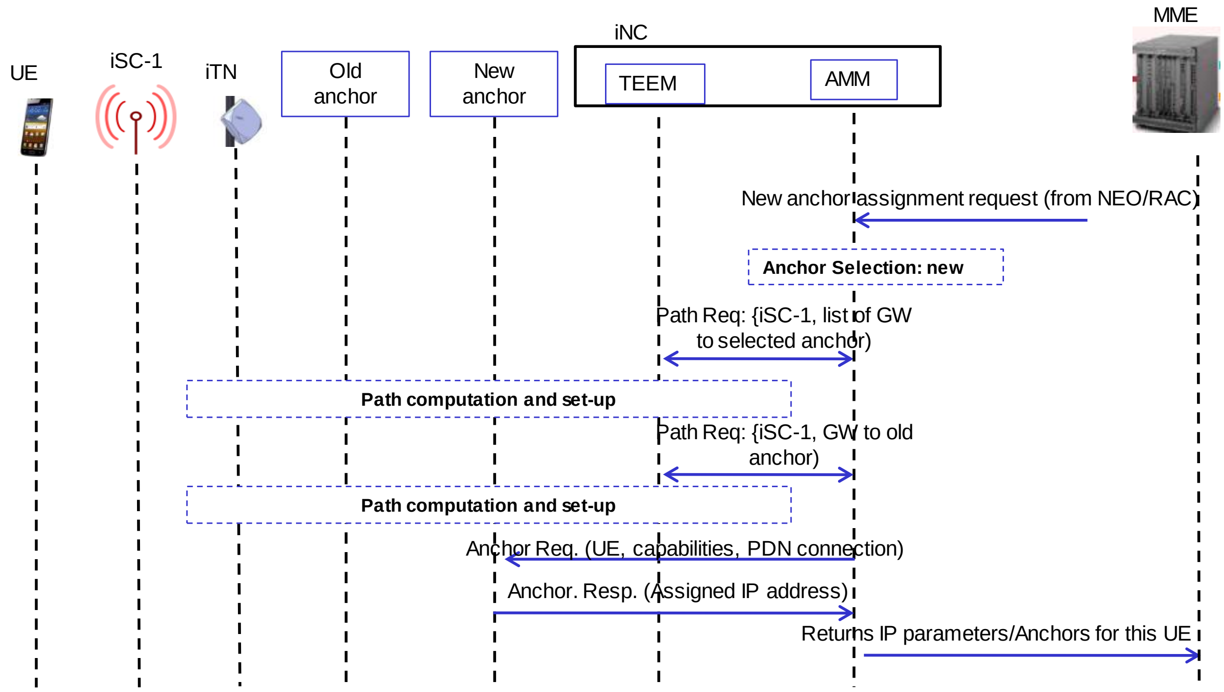

Figure 4-7 MSC of the new anchor assignment procedure (anchor and mobility entities)

Considerations on Energy or Congestion triggered mobility

A change of point of attachment to the network by the UE may not be triggered only because of mobility reasons (i.e., the UE physically moving out of the coverage of current base station), but also from energy or congestion reasons. Figure 4-8 shows an example MSC, in which a UE is triggered to change its point of attachment because of energy considerations (congestion or load balancing can be also valid triggers considered by the iJOIN project). In the example, the NEO module decides that iSC-old can be switched off to save energy, but this requires moving UE to another small cell. In order to do this the NEO module sends a request to the AMM module requesting to move UE. The AMM module selects a new point of attachment (iSC-new in the example) and also analyses if new anchors should be assigned considering existing sessions. The AMM then requests the TEEM to re-compute and set up the new paths for ongoing sessions of the UE (and considering active anchors) and updates the MME with current IP parameters and selected anchors. Note that the AMM also has to trigger the handover of the UE. This can be done by direct interaction between the AMM and the UE (e.g., using new OpenFlow signalling) or via the MME.

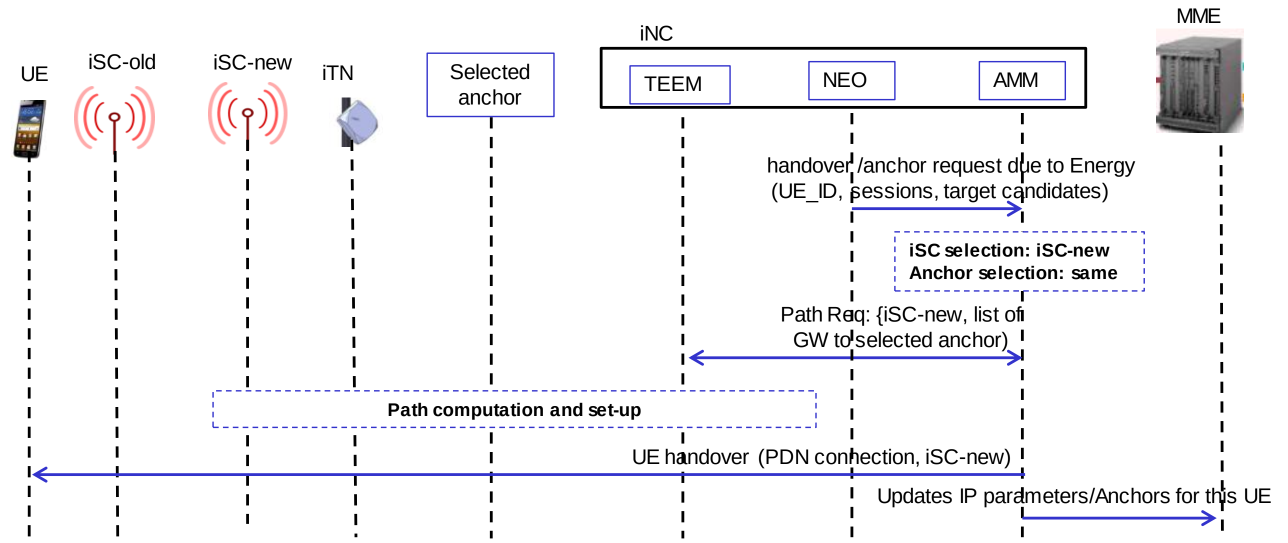

Figure 4-8 MSC of an energy-triggered handover

Considerations on UE involvement

In many of the aforementioned procedures, the active involvement of the UE could aid, in terms of increased reliability, better decision taking and improved performance. Within WP4, the active participation of the UE will be analysed on each of the main procedures. Note that, like in any other engineering decision, there are tradeoffs that need to be considered while analysing this UE involvement. For example, there is a clear disadvantage, which is the need for modified UEs that have to implement the new iJOIN technology. However, we believe that some of the needed changes may be worthwhile, for example in scenarios where the density of small cells is too large and the participation from the UE would help in taking better coordinated decisions. Additionally, we will also seriously consider the implications of bringing down the SDN approach also to the UE, as this might be a good compromise in terms of required support on the UE (i.e., if the UE is SDN enabled, this would facilitate almost any kind of required operation from the UE, without additional updates on the UE's stack), and the provided functionality. Figure 4-9 represents the different types of UEs that the project will consider during the detail CT mechanisms design and implementation phases.

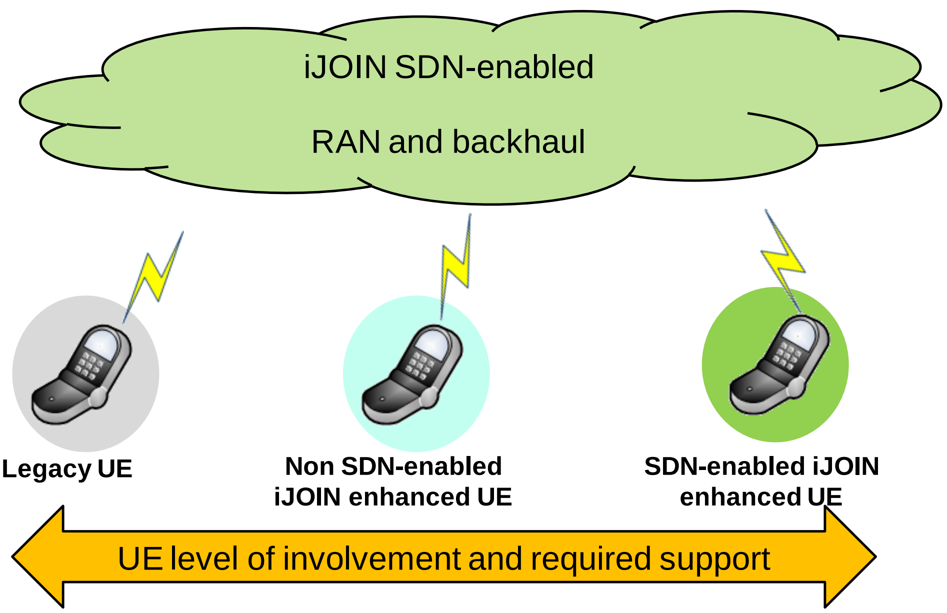

Figure 4-9 UE's involvement in iJOIN WP4

4.1.4.2 Mobility Management using the Virtual Cell Concept

The SDN based centralised solution is proposed in section 4.1.4.1. In the SDN based approach, mobility anchor selection relies on the decision of iNC. Based on the nature of mobility management, SDN solution might not be optimal in some of the scenarios. For example, fast changing UE's points of attachment, mobility domains across heterogeneous networks which may not be SDN-enabled, large network domain which leads to the possible scalability problem of centralised solution. Therefore, it is proposed in this section a complementary approach: 3GPP-based distributed mobility management solution, in which the mobility anchor selection decisions are performed by the network itself.

Overview

The proposed solution requires support of virtual cells as illustrated in Figure 4-10. The formation of a virtual cell, i.e., a cluster of cooperating and logically grouped small cells, appears to the UE as a single cell. In this case, handovers would occur only at the virtual cell boundaries. The handover within a virtual cell could be solved locally by local mobility management scheme. The handover between virtual cells can be addressed using IP mobility mechanisms. And cooperation between virtual cells to enable load-aware handover decision is also going to be considered.

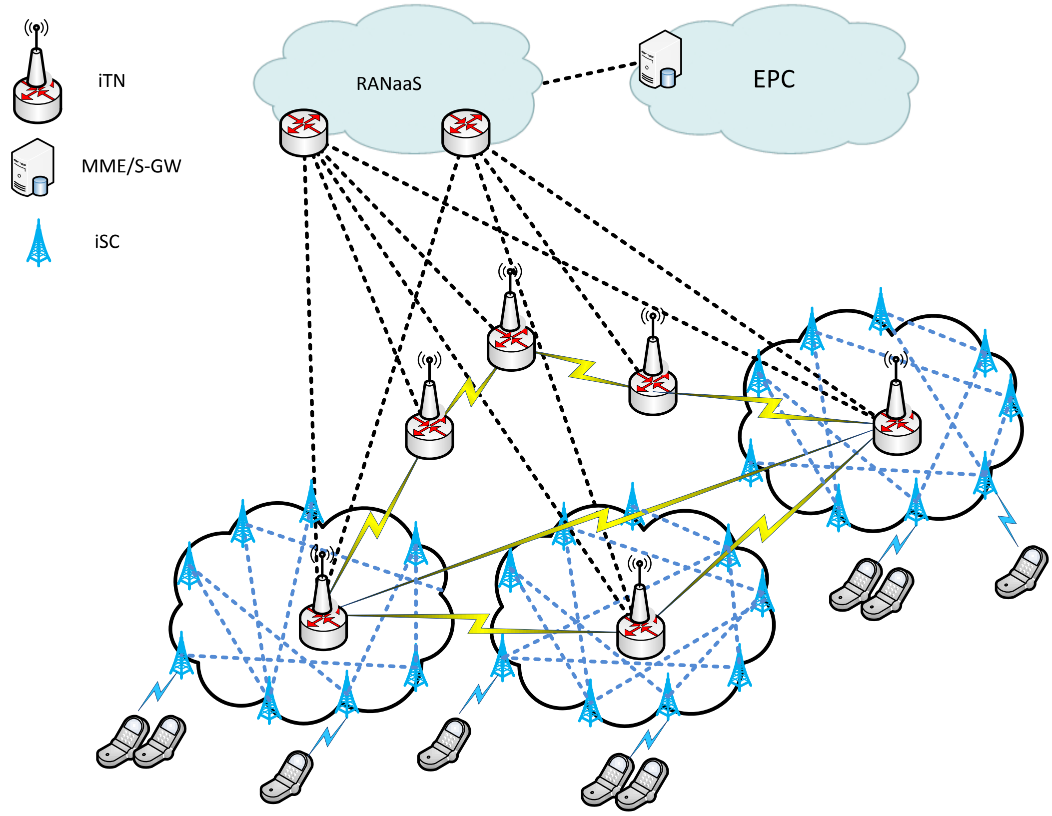

Figure 4-10 Mobility management: concept of virtual cells

Highlight of the solution

Figure 4-11 shows an outline of the mobility management solution which will be further investigated. The key aspects of this solution are:

- Small cells, which are shown as iSC in the figure, are connected to the core network through multiple backhaul transport nodes, which are shown as iTN in the figure. Some of the iTN can provide mobility anchor function and those ITN can be used to handle mobility within one virtual cell domain. The introduction of iTN mobility anchor provides the possibility of localised mobility management for handover between small cells within one virtual cell. The iTN mobility anchor is located between small cells and the mobile core network to act as the mobility anchor point for inter-small cell handover.
- The formation of virtual cells, a cluster of cooperating small cells, that appears to the user as a single cell. In this case, handovers occur at the virtual cell boundaries. The handover within a virtual cell could be solved locally. One virtual cell is managed by the corresponding iTN mobility anchor accordingly.

- The solution investigates the handover signalling for three handover phases: handover preparation, handover execution and handover completion. To minimise the data lost during handover, the traffic forwarding scheme between small cells will be enabled. The traffic forwarding and path switch mechanism will be investigated in the handover execution phase and handover completion phase accordingly.
- The traffic forwarding scheme looks beyond the shortest path traffic forwarding and traffic forwarding with a threshold schemes, by taking into account traffic load conditions along the forwarding chain. Since the local traffic forwarding may increase the end-to-end communication latency and consume local resources, traffic load conditions along the forwarding chain should be considered to balance the trade-off between the path switch cost and traffic forwarding cost.

The solution investigates handover between small cells in two different scenarios: within one virtual cell where the handover can be handled locally by the iTN mobility anchor and between virtual cells where the handover performance is enhanced by cooperation between neighbouring iTN mobility anchor.

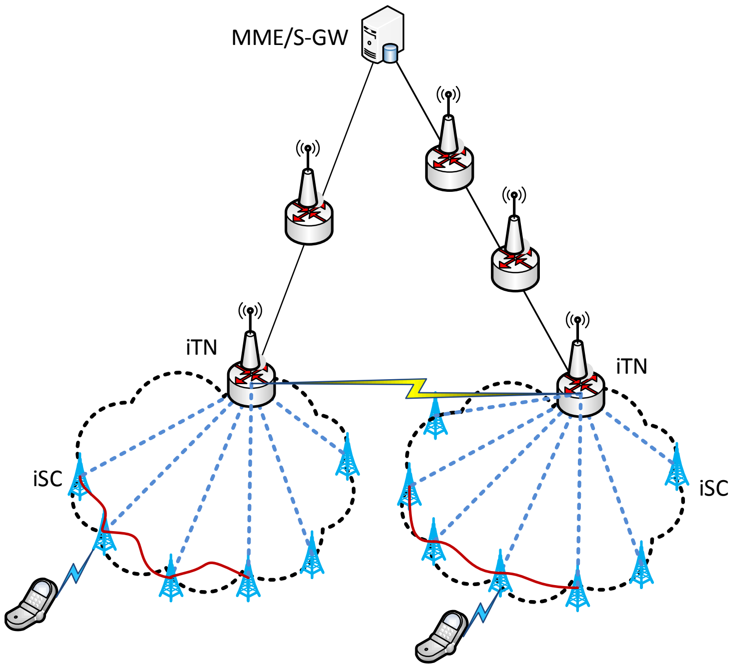

Figure 4-11 Outline of the mobility management solution

Overview of the handover scenarios

It is shown in Figure 4-12 the overview on the handover scenarios. Two main scenarios will be investigated in the following sections: intra anchor mobility and inter anchor mobility. Intra anchor mobility is the scenario in which the handover occurs within one virtual cell domain. In this scenario, the mobility can be handled by local mobility management scheme, with the anchor in the iTN mobility anchor as shown in the figure as tier-2 mobility anchor. The fast handover scheme with the shortest path traffic forwarding algorithm is also used, which means that the traffic forwarding is not following the UE's moving trace but considering the shortest path among the neighbouring small cells. If the local traffic forwarding between small cells using J2 interface is enabled, there will be a temporary mobility anchor in the small cell, as shown in figure the tier-3 iSC temporary mobility anchor. Inter anchor mobility is the scenario in which the handover occurs between virtual cell domains. In this scenario, the mobility can be handled with the cooperation of iTN mobility anchor. The new iTN mobility anchor will be the new anchor point when the UE moves into a new virtual cell domain.

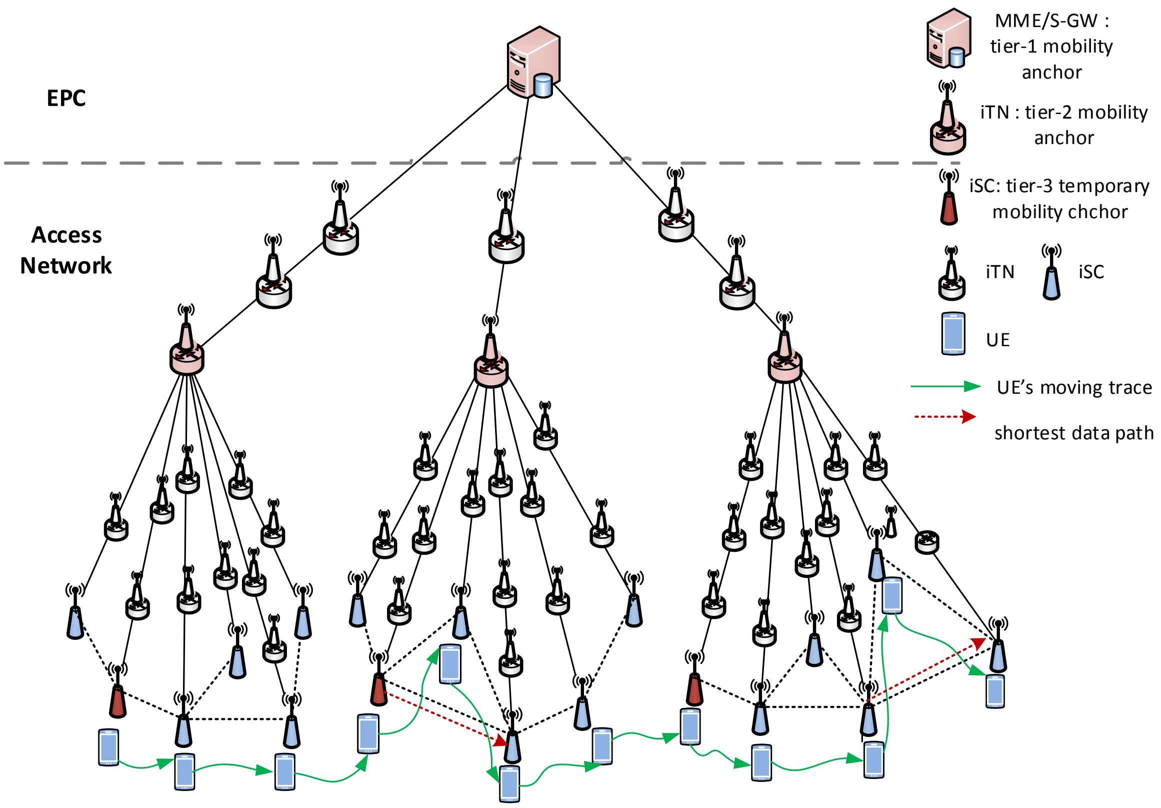

Figure 4-12 Overview of the handover scenarios

Intra anchor mobility

3GPP baseline architecture

A 3GPP LTE-A networked femtocell architecture is considered here as baseline architecture, in which a number of femtocells are deployed becoming a local network. In 3GPP terminology, femto base station is defined as Home Evolved Node B (HeNB). Each HeNB is connected to the Mobility Management Entity (MME) and Serving Gateway (SGW) in the Evolved Packet Core (EPC) via S1-MME and S1-U interfaces, respectively. A HeNB Gateway (HeNB GW) can be optionally deployed to act as a concentrator for the Control-Plane signalling and optionally for the User-Plane data traffic. The purpose of introducing the HeNB GW is to support a large number of HeNBs in a scalable manner without modifying the S1 interfaces between the HeNB and the EPC. Therefore, a HeNB GW appears as a HeNB to an MME, and it appears as an MME to a HeNB. However, a HeNB GW is deployed at the edge of the EPC and the signalling between HeNB GW and HeNBs towards MME/S-GW still need to traverse the backhaul network. And the fully centralised mobility management entity can become a bottleneck if massive signalling from/to HeNBs need to be handled by HeNB GW. The direct X2 interface between HeNB has been supported in 3GPP since release 10 and has been further defined by 3GPP release 11. The direct X2 link is a key requirements for inter-HeNB handover enhancement, as the traffic forwarding between HeNBs can be enabled using X2 interface [99].

The signalling exchanges of 3GPP X2-based handover in Femtocells are illustrated in Figure 4-13. The X2-based handover can be utilised when UE moves between HeNBs, which is similar to the X2-based handover between eNBs. The handover procedure consists of three phases: handover preparation, handover execution and handover completion.

In the handover preparation phase, UE measures the signal strength of the neighbouring HeNBs. If the handover criterion is satisfied, the handover decision will be made and the source HeNB will send a handover request message to the target HeNB using the X2 interface. If the target HeNB admits the request, an acknowledgement message with configuration information will be sent to source HeNB over the X2 interface. The configuration information will be sent to UE via air interface [99].

In the handover execution phase, the UE will detach from the source HeNB and synchronize to the target HeNB. Meanwhile, the source HeNB will send a SN Status Transfer message to the target HeNB notifying the bearer sequence number status. And the data received from the EPC will be forwarded to the target HeNB over the X2 interface [99].

In the handover completion phase, the target HeNB will send a path switch request to the HeNB GW to switch the data path in case that the mobility anchor point is located at the HeNB GW, and the HeNB GW will switch the data path from the source HeNB to the target HeNB and respond with a path switch request Ack to the target HeNB. After the data path is switched, the HeNB GW will send an “end marker” along the old path. Upon receiving the path switch request Ack, the target HeNB will inform success of handover to the source HeNB and triggers the release of radio and control-plane related resources associated to the UE context from the source HeNB by sending a UE context release message to the source HeNB. The release of the data forwarding resource at the source HeNB will be triggered after the “end marker” of the data arrives. In case that the mobility anchor point is located at the MME, the path switch request will be forwarded to the MME and the MME will inform the corresponding S-GW to switch the path. After the MME receives the response from the S-GW, the MME will send a path switch request Ack to the target HeNB, which is then followed by the resource release procedure [99].

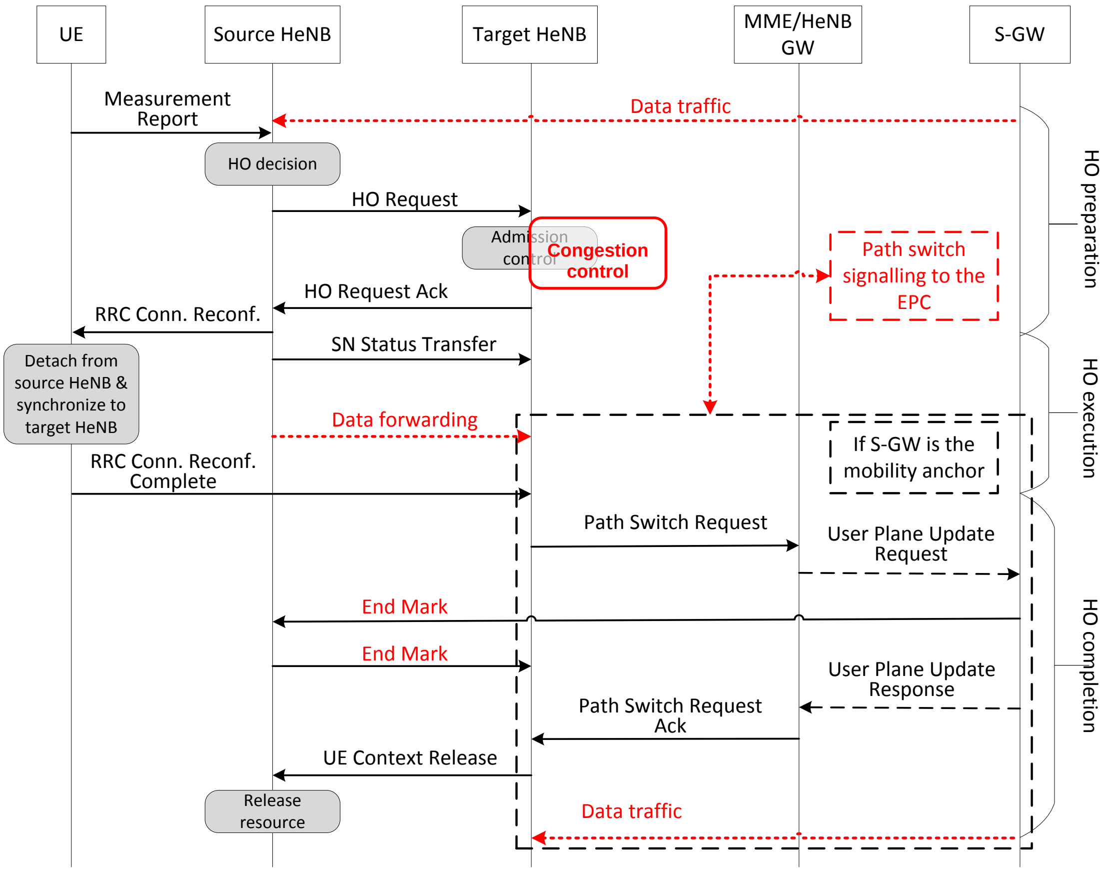

Figure 4-13 MSC of 3GPP X2-based handover in Femtocells [99]

The improvements against 3GPP baseline architecture will be made by the proposed mobility scheme, with focus on:

- To eliminate frequent signalling towards EPC. In the baseline architecture, the mobility anchor HeNB GW is in the edge of EPC, thus the path switch message goes to EPC every time the handover is performed. It causes problems of bottleneck and EPC can be heavily loaded. Therefore, a mobility management scheme with dynamic mobility anchors in the access network will be needed.
- Fast handover scheme with the traffic forwarding over X2 link. The traffic forwarding should be beyond following the UE’s moving trace, but also finding out the shortest local path. The data forwarding mechanism proposed by 3GPP is to be only used during the handover procedure in order to guarantee the lossless session mobility. However, considering the cost of local traffic forwarding chain and the cost of path switch messages, sometimes it is worthy to eliminate the path switch procedures but just implementing the traffic forwarding scheme, if the forwarding chain is lower than a particular threshold.

- The mobility management should look beyond the mobility anchor selection and fast handover scheme, by taking into account traffic load conditions. The congestion control can be taken into consideration in the handover preparation phase, as shown in Figure 4-13. If the target cell is aware that the local network is congested, the local mobility handover request can be denied. In this case, a standard 3GPP mobility management scheme with updates towards EPC will be used.

Proposed mobility management scheme

The MSC of intra anchor handover mobility scheme is shown in Figure 4-14. This MSC applies to the scenario of intra anchor handover with the shortest path traffic forwarding mechanism. Mobility is now temporarily anchored in a temporary mobility anchor point which is an iSC. It can be noticed in Figure 4-14 that the handover preparation and handover execution phase remains the same as in 3GPP baseline system in Figure 4-13. The main improvement is the control plane updates and user plane data traffic in the handover completion phase. After the synchronisation between UE and target small cell is completed, the target small cell will not send the path switch request to EPC. Instead, the target small cell sends a local path switch message to the temporary mobility anchor along the shortest path, as long as the shortest path does not exceed the threshold. The definition of the threshold depends on the topology of local access network and will be explained in the next scenario. In this scenario, the traffic forwarding is triggered, and temporary mobility anchor will switch the data path from the source small cell to the target small cell. The end mark is sent along the old forwarding chain until the end mark reaches the target small cell to release the resources. Upon receiving the local path switch request Ack message, the target small cell will inform success of handover to the source small cell and release the radio and control plane related resources by sending the UE context release message to the source small cell. After the handover procedure is completed, the data traffic now arrives to the target small cell via the temporary mobility anchor.

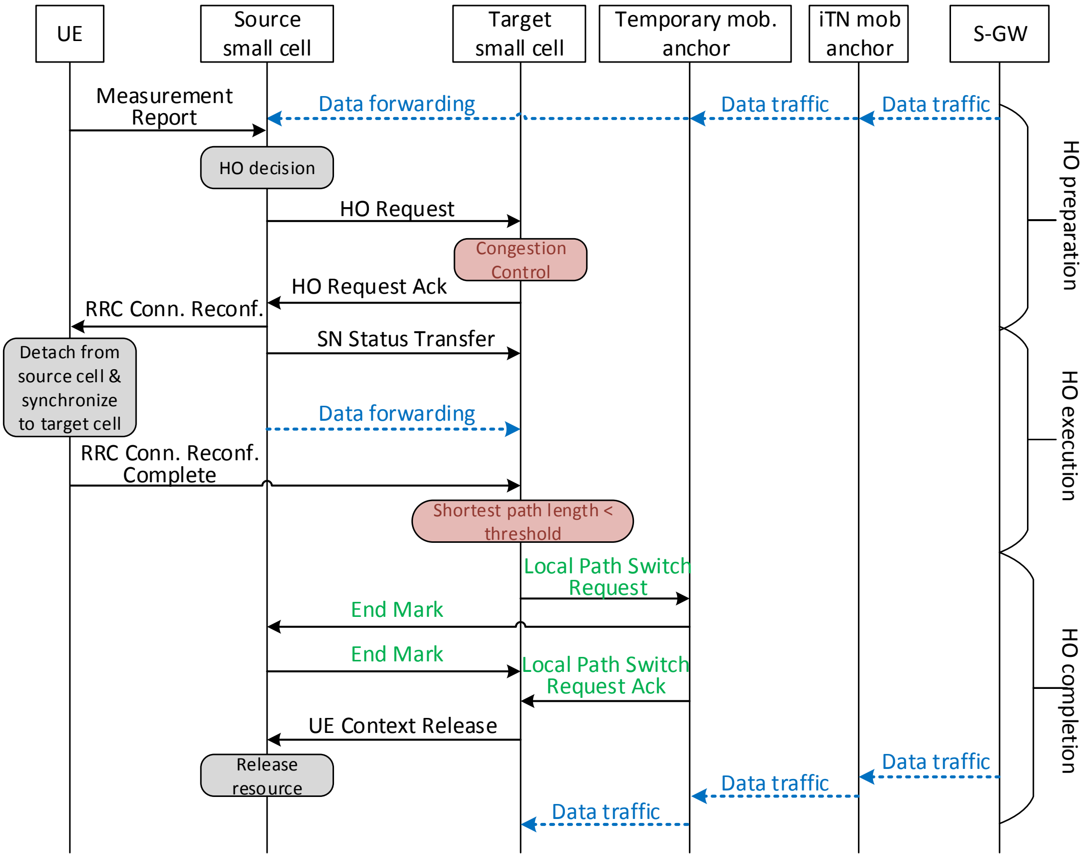

Figure 4-14 MSC of intra anchor handover with temporary mobility anchor in iSC

An example of the handover scenario is shown in Figure 4-15 to illustrate the data path. In the figure, the UE's moving trace is along the green arrows and the temporary mobility anchor is the red iSC which is the start point of UE's movement. However, the data forwarding path as in red arrows is not following the UE's moving trace, but by selecting the available shortest path from the temporary mobility anchor. For example, data path (3) is the shortest path between the temporary anchor and UE's point of attachment. Based on the shortest path principle, the data path from the EPC should be (1)+(2)+(3)+(4)+(5)+(6) as shown in figure, as the UE moves along the green trace arrow.

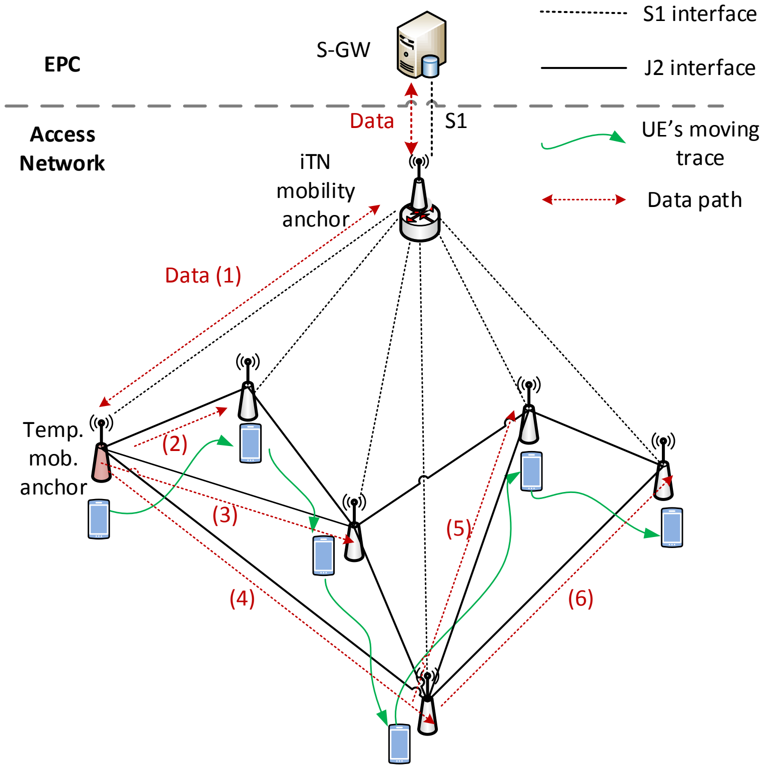

Figure 4-15 Intra anchor handover scenario with temporary mobility anchor in iSC

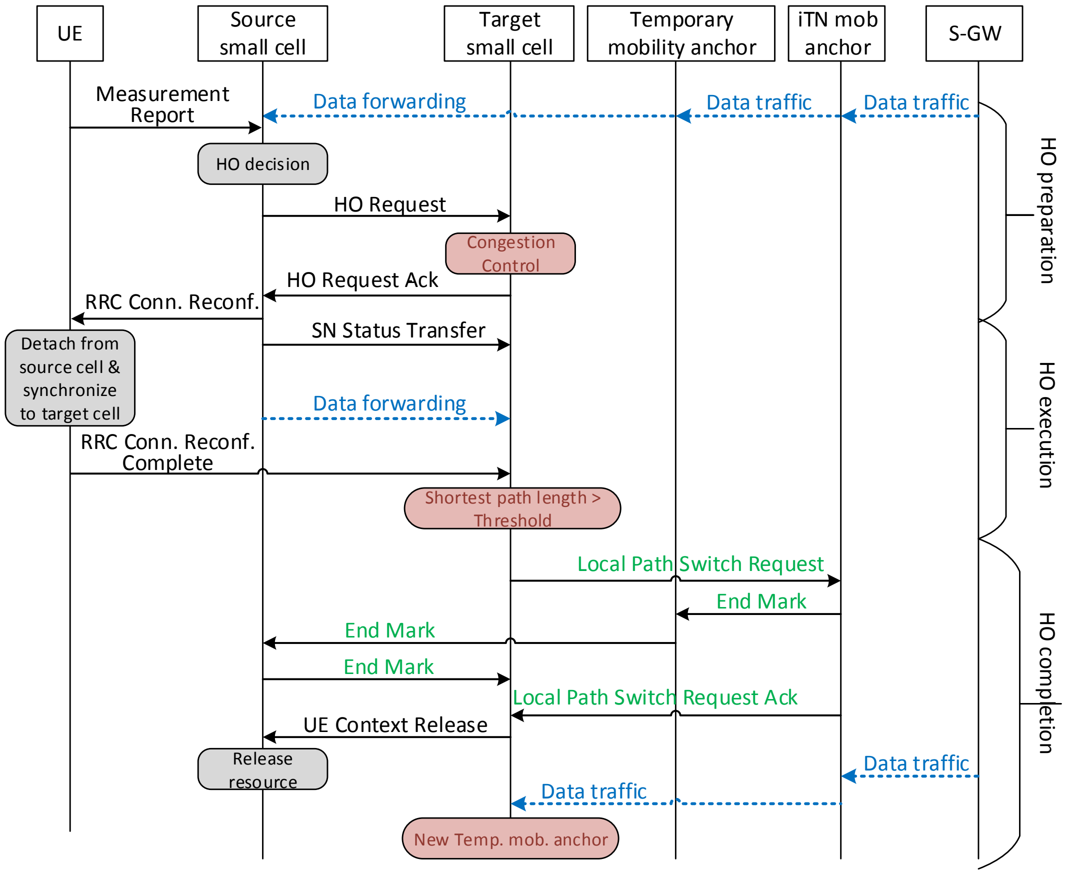

Figure 4-16 MSC of intra anchor handover with iTN mobility anchor

As shown in Figure 4-16, if the shortest path length is larger than the threshold, the local path switch operation towards the iTN mobility anchor point will be enabled. The threshold is calculated as the average number of hops between the iTN mobility anchor and all of the iSCs within its virtual cell domain. If the threshold is exceeded, the traffic forwarding along the iSC is not the best choice anymore. As the cost of forwarding the data traffic from the iTN mobility anchor to the target small cell will be less on average. As shown in the figure, the local path switch request message is sent by the target small cell to the iTN mobility anchor. And the end mark will be send from the iTN mobility anchor to the target small cell along the old forwarding chain to release the resources for data forwarding. When the handover is completed, the new data traffic will be arriving at the target small cell through the iTN mobility anchor. And the target small cell becomes the new temporary mobility anchor, when the UE moves on within the virtual cell domain.

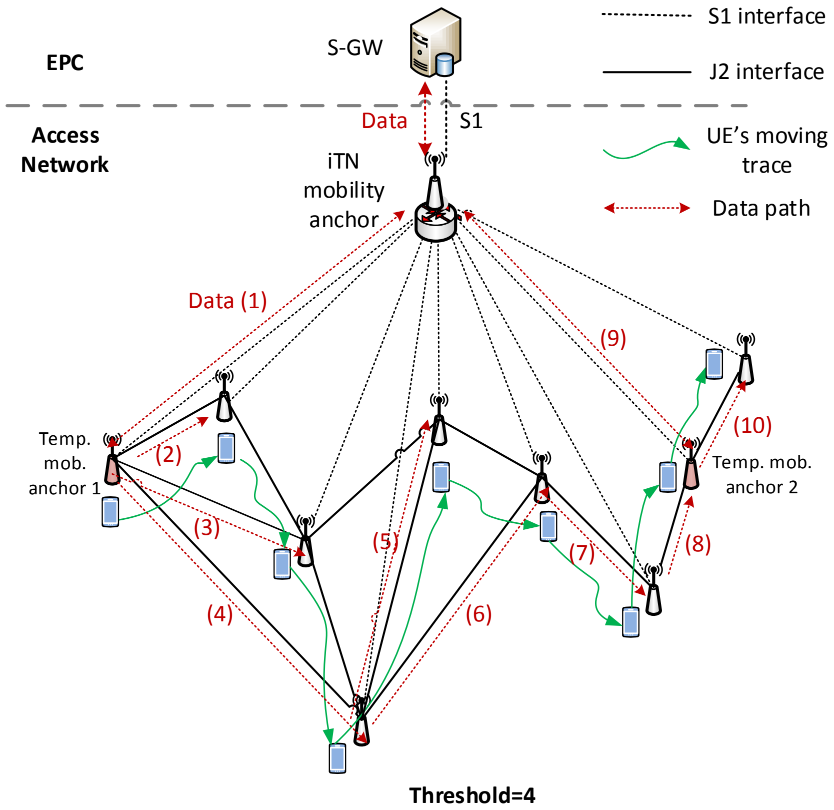

Figure 4-17 Intra anchor handover scenario with iTN mobility anchor

An example of the handover scenario is shown in Figure 4-17 to illustrate the data path. The threshold is 4 in this example. In the figure, the UE's moving trace is along the green arrows and the temporary mobility anchor is the red iSC 1 which is the start point of UE's movement. Based on the shortest path, the data path of (1)+(2)+(3)+(4)+(5)+(6)+(7)+(8) is selected as the UE moves along the green moving trace. However, if UE moves on from (8), the shortest path (4)+(6)+(7)+(8) will exceed the threshold of 4. Therefore, the local path switch to the iTN mobility anchor is triggered. The temporary anchor will be moved to temporary mobility anchor 2 as shown in the figure. And the data path from the EPC should be (9)+(10) as the UE moves on.

Inter anchor mobility

The MSC of inter anchor handover mobility scheme is shown in Figure 4-18. The inter anchor handover occurs when UE moves across the virtual cell boundary. Thus, the handover procedure involves the change of iTN mobility anchor, from source iTN mobility anchor to the target iTN mobility anchor. The path switch request message is sent by the target small cell to the target iTN mobility anchor, and is forwarded to the MME/S-GW in the EPC. The EPC will switch the data path from the source iTN mobility anchor to the target iTN mobility anchor. The end mark is sent by EPC along the UE's old forwarding chain until it reaches the target small cell to release the resources. When the handover is completed, the data path from the EPC will arrive at the target small cell via the target iTN mobility anchor. And the target small cell will become the new temporary mobility anchor when the UE moves in the new virtual cell domain.

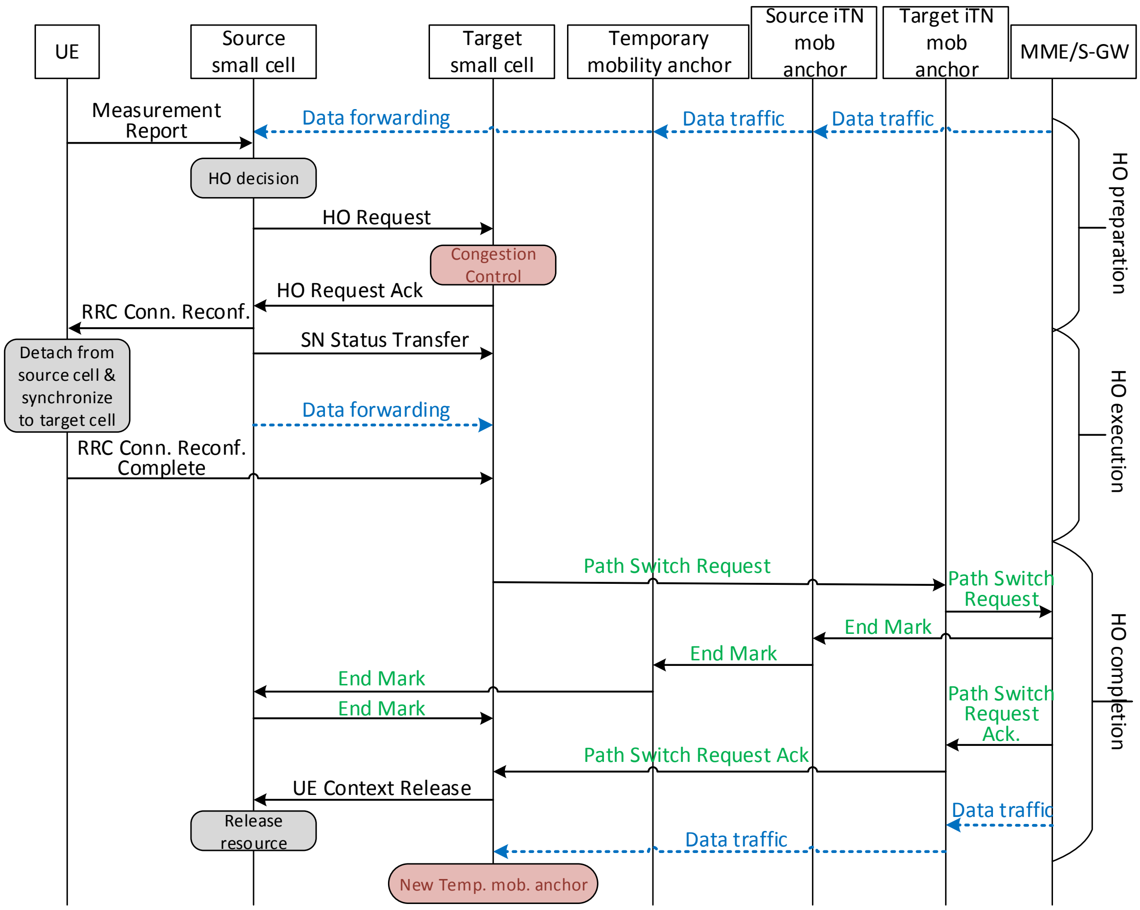

Figure 4-18 MSC of inter anchor handover

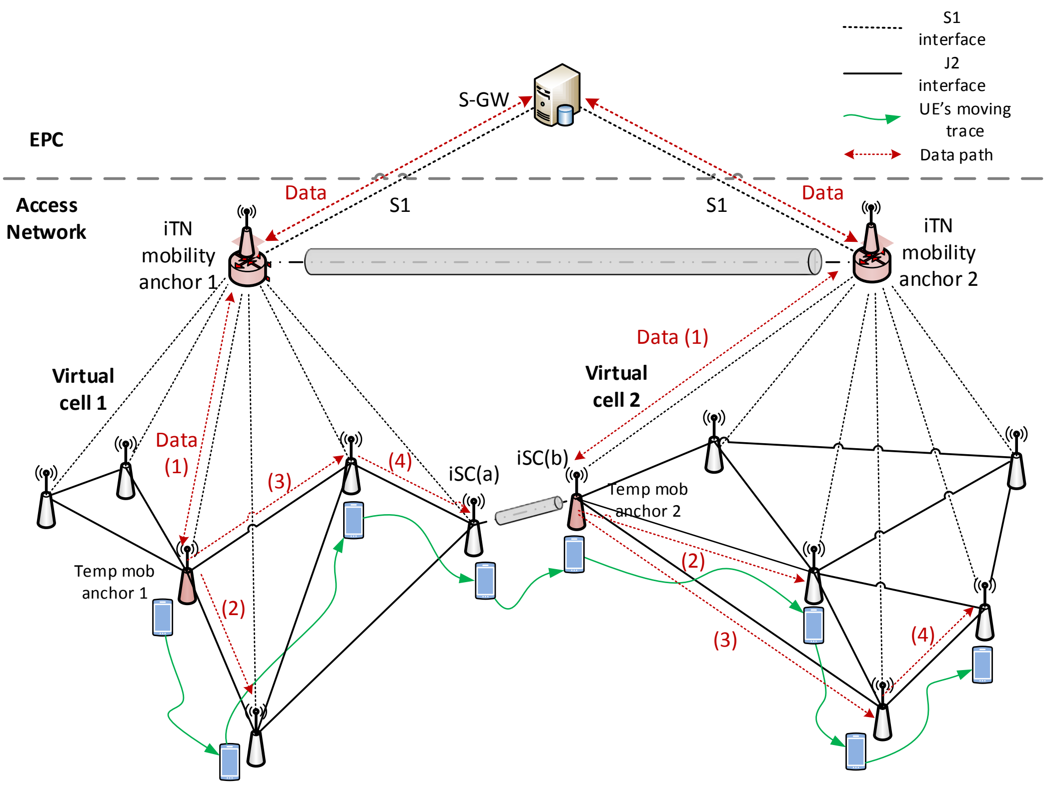

Figure 4-19 Inter anchor handover scenario

An example of the inter anchor handover scenario is shown in Figure 4-19. In virtual cell 1, the handover procedure follows the scheme explained before in the intra anchor mobility section, with the data path of

(1)+(2)+(3)+(4). When UE crosses the virtual cell domain from iSC(a) to iSC (b), the inter anchor handover procedure will be triggered. Within virtual cell 2, iSC(b) will become the new temporary mobility anchor. And the tier-2 anchor will be the iTN mobility anchor 2. The new data traffic in virtual cell 2 will be arriving from iTN mobility anchor 2 following the data path of (1)+(2)+(3)+(4). Please be noted that before the inter anchor handover procedure is completed, existing data traffic will be forwarded from iSC(a) to iSC(b) to reduce the data loss rate. Optionally, the temporary data traffic forwarding from iTN mobility anchor 1 to iTN mobility anchor 2 can be enabled to reduce the cost of temporary data traffic forwarding.

4.2 CT 4.2: Network Wide Energy Optimisation

4.2.1 Motivation

Energy efficiency is both ecologically and commercially important to Information and Communication Technologies. Over 0.5% of the global energy consumption comes from wireless communication systems, mainly by outdoor cellular network BSs. A key challenge is to significantly reduce the energy consumption level whilst maintaining and even enhancing network capacity. Moreover, in order to improve competitiveness and the average revenue per UE, operators have to reduce OPEX of cellular networks.

Existing research on reducing the energy consumption of cellular networks has mainly focused on capacity improving transmission and RRM techniques, such as multi-user MIMO. Considering the total energy consumption of the Radio Access Network, the amount of energy saved by transmission and RRM techniques alone is fundamentally limited, while the energy saved by re-deployment can be much more significant.

According to a recent survey, nearly 80% of the energy consumption of a typical cellular network comes from the BSs [100]. Furthermore, 70% of the BS energy consumption is caused by power amplifiers and air conditioning, which are used to keep the BS active even when there is no traffic in the cell. Hence, the optimisation of BSs should have a large impact on the overall cellular energy efficiency. Energy efficiency can be improved from the following aspects:

- Offloading traffic from macro-cells to small-cells in order to be able to turn off macro-cells.
- Mechanisms that intelligently switch small-cells on/off for energy saving purpose. Energy saving for green networking is mainly realised by preventing cells from emitting at full power when there is no UE to serve.

The backhaul links will consume energy and some forwarding nodes on the backhaul links may be switched off based on the network utilization. SDN techniques can be used to provide network-wide energy optimisation. Each time the controller in SDN gathers fresh data, it can perform a new optimisation if necessary.

4.2.2 Assumptions

Based on the previous introduction, we can state the following key assumptions for the technology candidate:

- Backhaul is an IP-based network
- Interfaces between small-cells, between small-cells and backhaul transport nodes, and between small cells/transport nodes and iNC
- Availability of network topology and real-time energy consumption information for small cells and transport nodes
- Network topology and energy consumption (as a function of load) per node is available.
- Current utilisation level of backhaul links and radio resource utilization at small cells are known.
- Spatio-temporal traffic profiles are known.
- Service Level Agreement (SLA) and users QoS requirements are known.

4.2.3 Technical Requirements

From the previous description, we can derive the key technical requirements which need to be fulfilled in order to implement the technology candidate:

- The small cells and transport nodes should have the ability to invoke a low-power sleep mode when not required to serve any data traffic. The low power mode can be either driven by small cells/transport node or iNC.
- When a small cell/transport node is switched off or suspended to a low power sleep mode, the attached UEs need to be evenly assigned to neighbouring small cells/transport nodes. Both access and backhaul links should be able to support the newly assigned UEs.
- SON technologies for energy optimisation, should take user mobility and traffic pattern into consideration. Small cell deployments and hierarchical deployments with overlay macro-cells may lead to a situation where many cells are barely loaded. In particular, this applies to situations where the load varies over different times of the day. In high load situations the best solution may be to provide coverage using many small cells, whereas in low load situations cells with only few users can be turned off by the network management. Self-organising mechanisms and signalling protocols are required to detect traffic situations in order to redirect UEs and to adjust the network coverage.
- The solution needs to provide better energy efficiency performance and can therefore be possibly coupled with a mobility management approach. For example, when making a handover decision, the small cells which are on low-power mode should be exempted from the neighbouring cell lists, thus they will not take any more UEs unless the small-cell is asked to return from low-power mode.
- SDN technique should enable iNC to collect real time network information from the access and backhaul network and to perform energy-optimising configuration accordingly.
- A traffic profiling mechanism should be available at small cells.
- The solution should interact with fast and reliable handover mechanisms which ensure non-disruptive behaviour when small cells are turned off.
- Traffic should be able to handover from one transport node to neighbouring transport nodes in the case that the energy saving algorithm indicates that a particular transport node may be turned off.
- Traffic profiling mechanism needs be improved with the support from UEs, i.e., UEs need to be capable of gathering statistics of their usage.
- Multi-tier energy saving algorithm needs be considered, i.e. a transport node may decide to turn a small cell off in the case of low utilisation.

4.2.4 Description of Candidate Technology

4.2.4.1 Network Wide Energy Optimisation using SDN Approach

Overview

Figure 4-20 shows an outline of the energy efficiency solution to be further developed within the iJOIN project. In the following, we summarise the key features of the solution.

The energy efficiency solution for small cell deployment scenarios are based on cell/network load situation. The solution should guarantee user accessibility when a cell/backhaul node is transferred to low-power sleeping mode. The solution should not have negative impact on the UE power consumption.

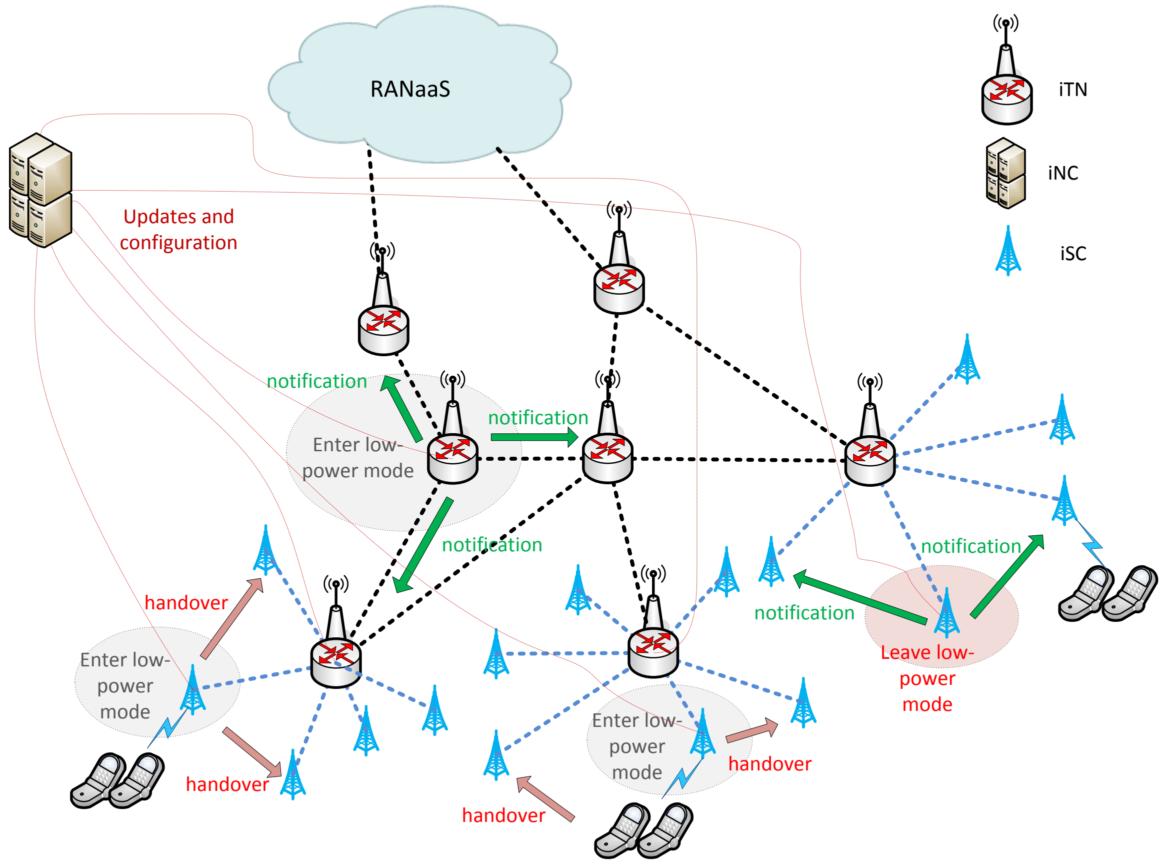

Figure 4-20 Outline of the network-wide energy optimisation solution

General Energy Saving Approaches

The energy saving procedures can be executed by following different approaches: centralised approach, decentralised signalling approach and hybrid approach. In the centralised approach, small cells/transport nodes enter or leave low-power sleeping mode based on centralised decisions, which are made based on the real-time information obtained from the network, e.g., load information. The decisions can be either pre-configured or directly commuted to the small cells/transport nodes. If a small cell/transport node enters or leaves its low-power sleeping mode, the neighbouring small cells/transport nodes should be informed through signalling. In this case, the neighbouring cell list can be updated and routing decisions can be made correctly.

In the decentralised signalling approach, small cells/transport nodes may decide to enter low-power mode autonomously or based on information exchanged with neighbouring small cells/backhaul nodes. The small cells/transport nodes are aware of whether they are energy saving capable or not based on their own information, e.g., load information. When a small-cell/transport node decides to enter the low-power mode, it will initialise communication with the corresponding small cells/transport nodes, and related information may be included in the request message. The final decision is made after the signalling exchange. The enter/leave low-power mode decisions/requests will be based on information locally available in the small cells/transport nodes, including load information of the neighbouring nodes. Leaving low-power mode can be invoked based upon requests from the neighbouring cells/transport nodes, or the local policy available in the node, such as a pre-defined max switch off time. The neighbouring small cells/transport nodes should be informed after each on/off decision. And in order to perform energy optimisation in a more efficient way, some energy efficiency parameters might be exchanged between small cells/transport nodes if it is required, e.g., power consumption, traffic threshold, etc.

In the hybrid solution, the small cells/transport nodes are pre-configured by a centralised network entity, such as iNC. Also the controller communicates to all small cells/transport nodes the values of some parameters that determine the behaviour of entering/leaving low-power mode.

Overview on Energy Optimisation Solution using SDN approach

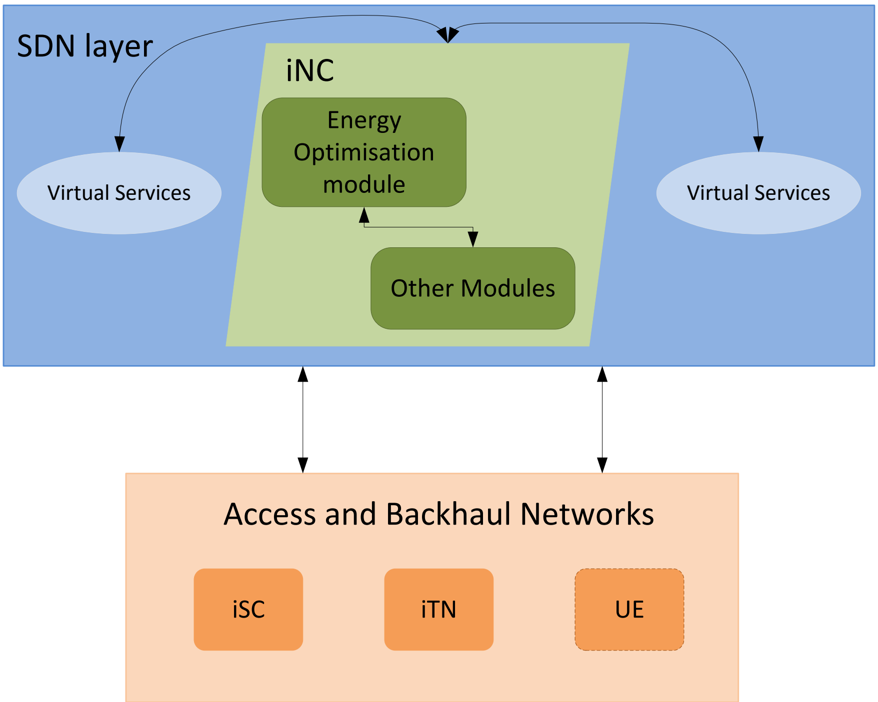

Figure 4-21 System overview of the SDN-based energy optimisation solution

The SDN technique decouples the control plane and the data plane, by implementing a central controller which can manage the network without requiring physical access to the network entities. The central controller runs on the iNC here. Several network layer functions can be implemented in SDN layer, such as routing, mobility management and energy optimisation. Implementing energy optimisation on the SDN platform implicates that the energy efficiency solution is mainly a centralised approach as described in previous section. In the SDN based centralised solution, energy optimisation decisions are made by iNC of the SDN platform and the decision can be commuted to the relevant network entities. The benefit of using SDN platform is that iNC has a global knowledge of the network and this knowledge is periodically updated to ensure it is spontaneous. Therefore, running energy optimisation module in iNC is a good solution to provide network-wide scalable energy optimisation.

OpenFlow is the key protocol in SDN architecture which allows the communication between central controller and network switches. Since the OpenFlow architecture removes the control software from the switches and moves it to a central location, it is expected a reduction in power consumption in the switches at the cost of the power consumption of the controller [101]. Moreover, with the ability to switch on/off network elements, it can lead to further reduced network-wide power consumption. However, to enable energy efficient networking applications to run on the OpenFlow controller, some extra messages might need to be added to the current OpenFlow specification. Those messages should not only indicate the status of a network element, but also allows the controller to instruct the network element to perform any energy-related actions.

Figure 4-21 shows the system overview of the SDN-based energy optimisation solution. There are many modules running in iNC on the SDN platform. The energy optimisation module is responsible of running energy efficiency algorithms to decide which network elements will be going to low-power sleep mode and which network elements will need to be waked up from the sleep mode. To guarantee the coverage and QoS, the energy optimisation module will also decide whether there will be some compensation cells in some scenarios. The energy optimisation module needs to interact with other network management related modules running in iNC in order to obtain the most up-to-date network utilisation information, such as load information of the network. The energy optimisation module also needs to update the other module in iNC which is in charge of the network topology, if any changes have been made in the network elements. The iNC communicates with the physical network including the access and backhaul networks to obtain updates on the network utilisation knowledge and to configure the networks according to the decision of energy optimisation module. The network elements communicating with iNC include iJOIN transport node (iTN), iJOIN small cell (iSC) and optionally UE. The signalling exchanges between SDN platform and the network elements will be described in the next section.

As highlighted in Figure 4-20, the SDN-based energy optimisation solution can be summarised as below:

- iNC interacts with the network elements (iTN and iSC) to get network updates and configure the network.
- The network elements take actions including entering low-power mode and leaving low-power mode.
- The neighbouring network elements are informed about the change and to take over the handover of UE/data traffic. iSC/iTN can inform its neighbouring nodes by signalling them directly. Alternatively, this can be done in a centralised manner by iNC informing all the relevant nodes.

Proposed SDN-based energy optimisation solution

MSC of the proposed energy optimisation solution

As shown in Figure 4-22, the signalling exchanges between iNC and network elements can be summarised as below:

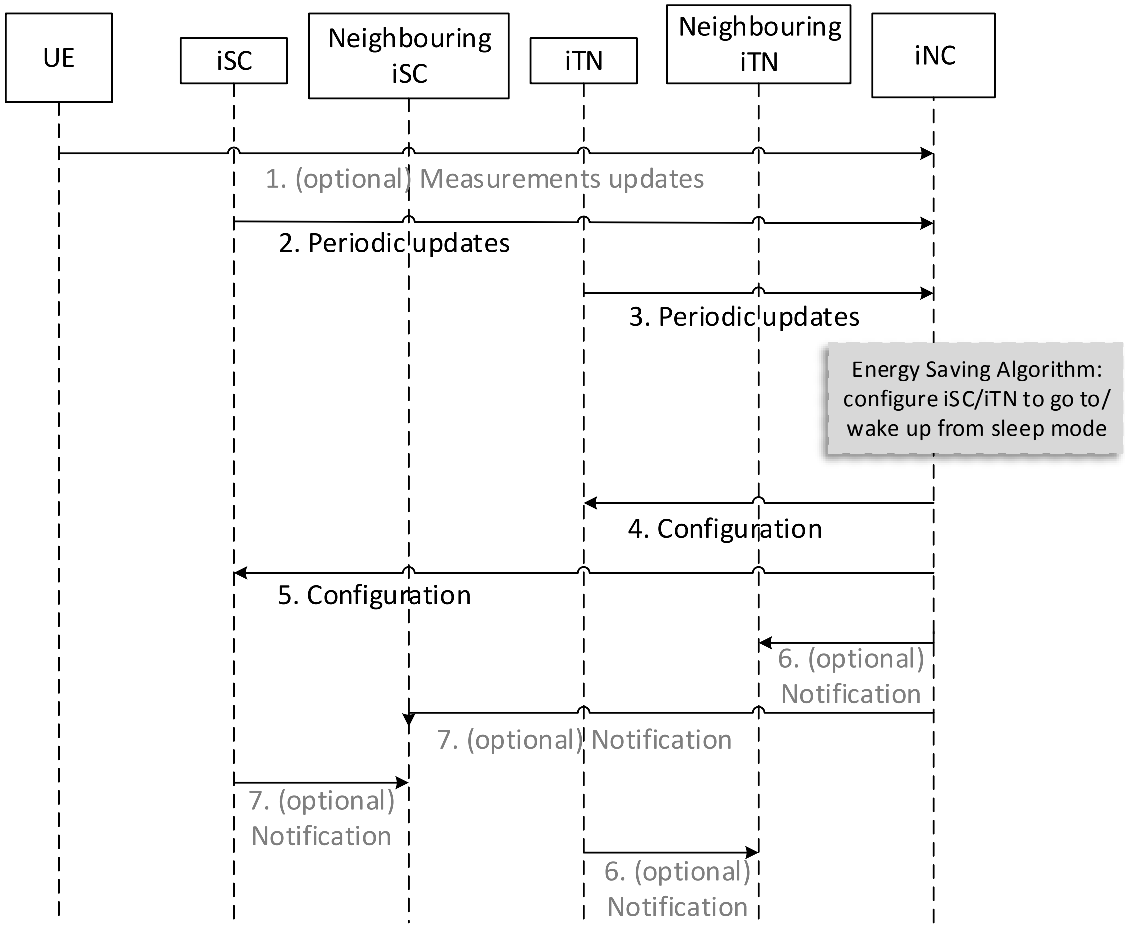

Figure 4-22 MSC of SDN-based energy optimisation solution

1. Measurements updates sent from UE to iNC. This is an optional message, although the measurements report from UE can assist the iNC to decide which iSC to the best to be in sleep mode/waked up. One example is that when there are a few candidate iSCs which can be waked up, the iSC which can serve the UEs with strong signal strength can be selected.
2. Periodic updates sent from iSC to iNC. By the updates message, the iNC obtain the knowledge of network utilisation information, such as how many UEs are served by each iSC, the load information of the iSC. The interval of the updates depends not only on the requirement of the energy optimisation module, but also on other criteria such as location of iNC, physical connection between the iSCs and the iNC.

3. Periodic updates sent from iTN to iNC. This message is similar to message 2. In the update message, the iNC obtain the load information of each iTN and its backhaul links.
4. Configuration message sent from iNC to iTN. Based on the network utilisation information obtained from the update messages, iNC runs the energy saving algorithm to decide which iSC/iTN to go to/wake up from sleep mode. This decision is commuted to the iTN in the configuration message.
5. Configuration message sent from iNC to iSC. This message is similar to message 4. The energy saving decision is commuted to the iSC by the configuration message.
6. Notification message to the neighbouring iTNs. This can be done in two ways. The iTN can either inform its neighbouring nodes by signalling them directly. Alternatively, iNC can send notification message to the neighbouring iTN to notify the change.
7. Notification message to the neighbouring iSCs. This message is similar to message 6. It is used to notify the neighbouring iSCs the change of status.

Energy Saving Procedures and Considerations on Energy Saving Algorithm

The energy saving activation/deactivation procedures in the iNC can be summarised as below:

- Step 1: the iNC makes a decision on which network elements should enter/leave energy saving mode based on the input from the network
- Step 2: Based on the output from step 1, the iNC initiates energy saving activation/deactivation on the selected network elements
- Step 3: After completion of the energy saving activation/deactivation process, the iNC is informed
- Step 4: Inform the neighbouring network elements about the change in energy saving status, either by the iNC or network element itself.
- (optional) Step 5: Energy compensation state to be enabled in some network element to guarantee coverage and QoS.

The energy saving algorithm running in the iNC may need to take into consideration the following criteria:

- Access and backhaul networks condition. This include:
    a) Network topology: geographic positions of iTN/iSC, coverage area of iSC.
    b) Whether iSC/iTN is capable of entering low-power sleep mode.
    c) Up-to-date load information.
    d) Historical statistics of network element’s traffic load to determine the peak/off-peak time.
- Predefined parameters, such as max switch off time.
- (optional) UE measurements. The UE may assist in deciding which iSC is the optimal to be waked up by providing measurement reports. iNC can request some sleeping iSCs to transmit the pilot signal for a short time interval. After this interval, the iSCs will still remain in sleeping mode. UE will perform reference signal measurements from the iSCs during this interval and send feedback. Based on the measurement results, optimal iSC can be selected.

Discussions on Centralised and Distributed Energy Optimisation Solutions

The key features to distinguish centralised SDN approach and distributed approach are listed in Table 4-2. The major difference is what entity makes decision to initiate the energy related actions: iNC or the network element. The decision is made by the iNC in a centralised SDN solution while it is made by the network element itself in a distributed approach [60].

Comparison Feature	Centralised SDN approach	Distributed approach
Energy saving related measurements	iNC	Network
Decision making to trigger energy saving actions	iNC	Network
Energy saving action execution	Network	Network

Table 4-2 Key features to distinguish centralised SDN approach and distributed approach [60]

The performance comparisons between centralised SDN approach and distributed approach are listed in Table 4-3. The two solutions are compared in many aspects in the table to show the advantages and disadvantages of each one.

Criteria	Centralised SDN approach	Distributed approach
Feasibility	Yes	Yes
Backward compatibility	Yes	Actions need to be taken to avoid coverage holes. Impact on handover parameter setting
Complexity	High, because SDN platform needs to coordinate the network elements switch on/off	High, coordination between network elements is needed to avoid coverage holes or excessive interference levels
Potential energy saving gain	Depends on the update frequency. Long term statistics may lead to a conservative solution	More flexible compensation schemes
Specification impact on 3GPP	OpenFlow need to be enabled	Signalling between network elements is needed, as well as definition of energy compensation scheme
Network management impact	High, iNC needs to define how to switch on/off network elements, the compensation scheme, and etc.	Low
iSC/iTN impact	Low, iSC/iTN needs to support OpenFlow	High, network element must be able to adapt coverage
UE impact	Not foreseen	Not foreseen

Table 4-3 Comparisons between centralised SDN approach and distributed approach [102]

4.2.4.2 Energy Optimisation Solution Considering QoS Constrains and Traffic Profile

Overview

Figure 4-23 shows an outline of the energy efficiency solution to be further developed within the iJOIN project. In the following, we summarise the key features of each one of the three different scenarios depicted in the following figure.

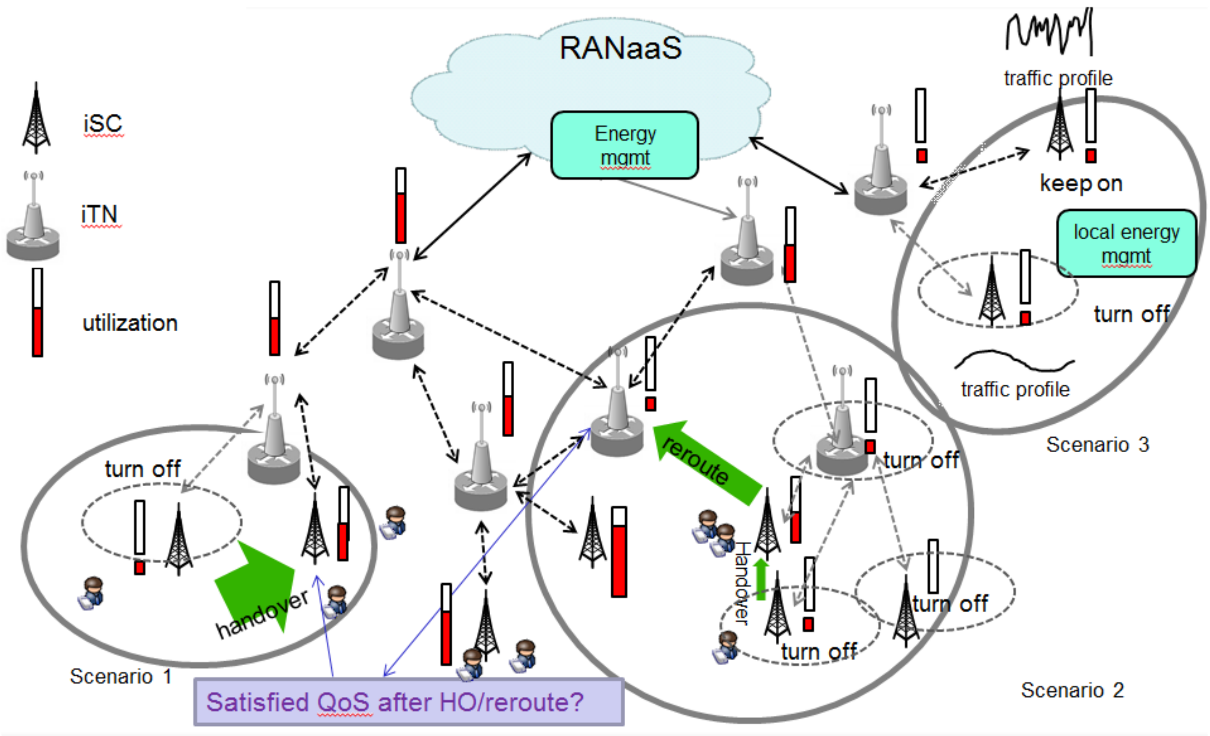

Figure 4-23 Network Wide Energy Optimisation

Scenario 1: In the simplest scenario (depicted in the left circle), we can select one (or more) “low-utilised” iSC and switch them off. Also, the users (if any) that were being served by this iSC, must be handed over to neighbour iSCs. In our approach, before proceeding to switch off nodes, we should check if some QoS constraints are still satisfied after the proposed action(s) (details in next section); if not, it might be advisable not to turn off the selected iSC).

Scenario 2: In this scenario, we jointly consider access and backhaul nodes, in order to decide which nodes (iSC, iTN) to turn off. For example in the middle circle, we can turn off the iTN and the two iSCs mentioned in the figure, in order to save energy, since they are “underutilised”. Also, the users of the iSC that is going to sleep (depicted in the circle) are being handed over to the iSC depicted (the iSC that remains ON) in the figure. Finally the iSC that remains ON, and is connected to the switched–off iTN must be rerouted to another iTN. This scenario is related to energy optimisation considering **jointly** access and backhaul nodes, and requires some topology control to be “approved” by the TEEM module (for TEEM module, see section 5.1).

Scenario 3: In this scenario (right circle) we also consider the traffic profiles of the iSCs (and so, the traffic profile of the users). Thus, if the traffic profile exhibits high variability over time, as the above iSC in the right circle, there might be no benefits (or even a loss) in switching off. For example maybe the load increase a lot, as soon as the iSC is turned off, requiring the iSC to be turned back on. This oscillatory behaviour not only has energy costs (due to the energy spent in turning on and off), but also introduces delays, altogether costs that will not be amortised by the amount of savings during the (short) off period. Instead, a more stable traffic profile like the bottom one is a better candidate when it enters a low utilization period. A “smooth” traffic profile is further considered in Figure 4-24 in an example taken from [103] (**t** depicts the time, whereas **f(t)** the traffic profile). For such a profile, we can find more easily a “suitable” time slot (in this example this time slot, can be a slot centralised at the T/2 time instance) that the traffic is low and finally to switch this node off. The horizontal white boxes show the energy saved, if we switch off an iSC (at τ_1 we switch off one BS, at τ_2 another etc).

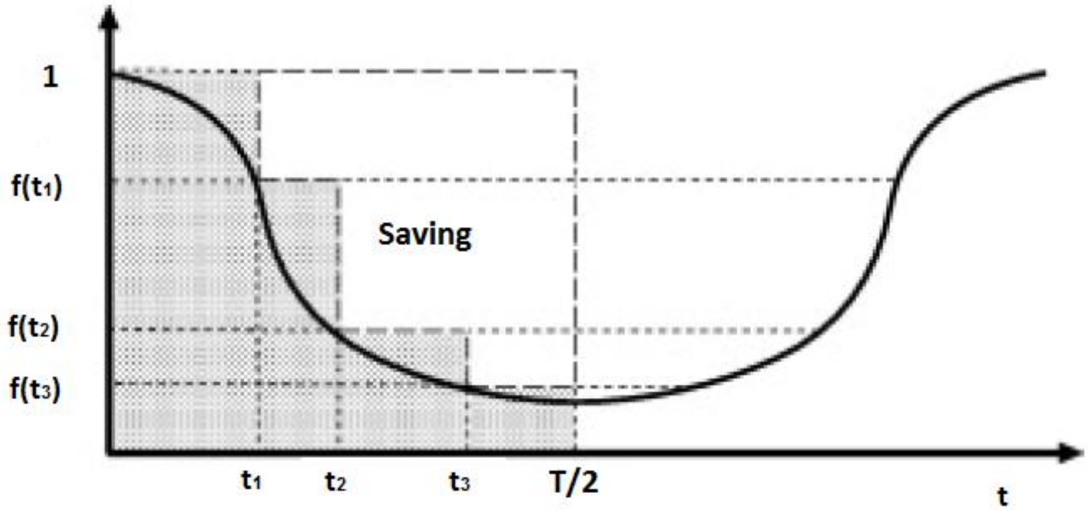

Figure 4-24 Traffic Profile of an iSC

Before we consider the various QoS constraints that will limit us from switching off elements too aggressively, it is important to discuss here briefly the possible power profiles of different iJOIN elements, which affect the objective function and thus the optimal actions. Figure 4-25 illustrates two possible curves for the power consumed by an iSC, as a function of its load. The power required for just keeping the iSC ON (*P[load=0]*), corresponds to the point where load=0. This is an essential characteristic that must be considered in order to formulate effective algorithms for energy savings that decide when and how long we should switch off a base station. For macrocells, this "idle" power is usually very high (e.g. as in the top curve), implying that the only efficient way to save power is to turn the cell off. However, small cells might have different profiles with much lower idle power consumption, and a more "load-proportional" case. In the latter case, NEO actions might employ power control and "cell breathing" methods to better allocate users between iSCs and iTNs, so as to minimise the overall energy consumption.

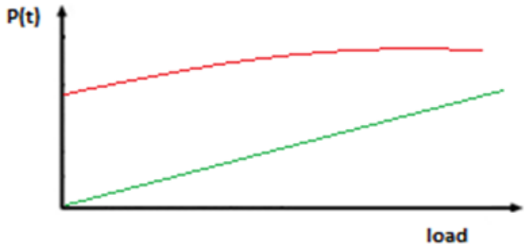

Figure 4-25 Power-consumed profile of two different iSC

Considering QoS Constraints

In this final part, we turn our focus on the various QoS constraints that the NEO CT must not violate when taking power-related (or topology-related) decisions, towards minimising the energy consumption of the network. Below, we discuss each of these constraints briefly. Finally, we remind the reader here that we plan to explore two different scenarios related to iSC power management: *i)* completely switching off an iSC, and, *ii)* "cell-breathing", namely increasing/decreasing the power transmission of an iSC, thus enlarging or shrinking the cell, respectively.

Thus, our optimisation problem has as its objective to **minimise the total power consumption of network elements**, subject to QoS constraints along the following three directions:

- **Coverage constraint**: Shutting down one (or more) iSCs, may lead to coverage "gaps" in areas where active users are located. Thus, it is crucial to ensure that:

  $$S(I)NR_j > \text{threshold}$$

  j may account for all, the "average", or an "edge" user. The actual choice is a design decision relating to different contexts and different QoE perspectives, and multiple such settings will be explored. For example, we can see an example in which the iNC has decided to switch off some iSC, and this leads to important coverage gaps because of the switched-off iSC depicted in the Figure

4-26. The colours red/ orange/ yellow/ green and blue depicted in the areas, symbolise the signal strength of the nearest iSC (the red is the highest, the blue the lowest). In this example, we can see that two users, depicted with an arrow, are not well-covered, so their QoE when they talk/use data may be quite low. The S(I)NR is a good general indicator as it relates both to the chance of having "reception" to a base station (e.g. for a call), as indicated by the simple "reception bar" indicator, as well as for the data rate than can be supported (e.g. through Shannon's formula). We intend to combine theoretical models for SINR as a function of distance, as well as position and call/data session initiation models to derive probabilistic metrics for the influence of different switch-off decisions.

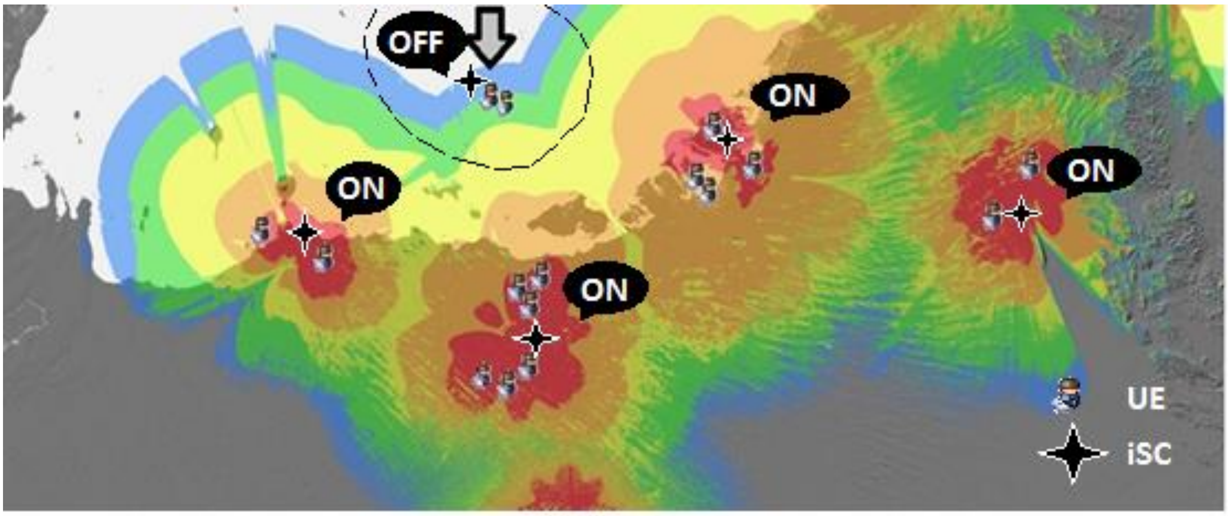

Figure 4-26 Coverage gaps due to turned-off iSC

- **Blocking Probability**: Moving one (or more) users to a new iSC leads to an increase of the "associated" users of this iSC. Thus, the probability of not being able to obtain a radio bearer for a user, i.e., the probability of a dropped call, increases. We are going to use standard queuing models (e.g. M/M/k/k or M/G/k/k systems) in order to compute the blocking probability and how it is affected by considered power management and handover decisions, with a goal to keep it below a desire threshold:

Pblock < pthr

- **Data Performance**: Moving one (or more) users to a new iSC, even if there are radio resources available (i.e. constraint 2 is satisfied) lead to an increase of the load that this iSC (and potential path on backhaul) has to handle. Thus, the delay of flows belonging both to the existing *and* new users, will increase. We are going to use advanced queuing models in order to compute the prospective delay in the data path and how this gets affected by different NEO-related decisions. The goal again is to maintain this delay lower than a chosen threshold. As before, we will consider various metrics such as average delay, worst case/user delay, priorities and different QoS, fairness issues, etc:

E[T] < Tthr

As a final note, we stress here that these constraints and their interplay will eventually be considered for both iSC and iTN elements, as well as jointly. Our exploration however will begin from the simpler, single element scenarios, in order better elucidate the impact of key parameters on performance, and will culminate in the complex, all-including scenario.

4.3 CT 4.3: Joint Path Management and Topology Control

4.3.1 Motivation

Dense small cell deployments are going to be subject to load demands that vary both in space and time. In addition, given the declining revenues experienced by mobile operators and the high number of small cells to be deployed, the small cell backhaul infrastructure should be as cost efficient as possible. Hence, dimensioning the small cell backhaul for peak traffic demands is neither a scalable nor a cost efficient approach. Instead, in iJOIN we envision that the infrastructure used to backhaul small cells should be able to adapt to varying capacity demands and to allocate backhaul resources where and when they are needed.

The iJOIN project has identified wireless backhaul technologies as a critical enabler for small cell backhauling. Several wireless technologies are currently considered as viable candidates for small cell backhaul, both high capacity LOS technologies (e.g. 60GHz and E-Band) and lower capacity NLOS technologies (< 6GHz). Thus, the future small cell backhaul will likely be composed of a mix of different technologies.

4.3.2 Assumptions

The following assumptions are imposed by this candidate technology:

- The leftover capacity in the current backhaul infrastructure used for the macro-network may be used to provide backhaul resources for small cells. A typical backhaul infrastructure for the macro network is composed of two or three levels of microwave aggregation trees connected to an optical ring.
- Additional fibre connection points might be deployed to backhaul small cells, i.e. some small cells may have a direct fibre connection.
- A centralised controller is available which is capable to manage the backhaul network and to modify the traffic flow on a per-node basis,
- In order to connect street-level small cells to the macro cell backhaul (rooftop), an heterogeneous set of technologies may be used:
    - 60GHz LOS P2P links: 60GHz radios are installed at the rooftop macro-site and on street level together with small cells. This allows small cells with LOS to the macro cell, to connect directly to the macro-site. Other small cells may use multi-hop to the macro site. In order to achieve path diversity, a single small cell can be backhauled with more than one 60GHz P2P backhaul unit.
    - E-Band (70-90GHz) LOS P2P links which are licensed but achieve longer range than 60GHz links at similar data rates. E-Band links can be installed at rooftop level to connect macro-sites with each other.
    - NLOS technologies (< 6GHz) which offer lower bandwidth and some of them may be subject to interference, but may offer a solution to backhaul small cells with no LOS to other small cells and a macro-site.

4.3.3 Technical Requirements

The following technical requirements are set in order to implement the candidate technology:

- The backhaul network needs to provide sufficient path diversity in order to allow for traffic engineering solutions.
- The backhaul network allows for low configuration granularity, i.e., backhaul nodes may be configured to route traffic through different connections,
- Provision of long and short time statistics by the eNB towards the iNC, which requires an appropriate interface (possibly even on flow-level),
- Existing backhaul infrastructure may be reused and existing equipment may be upgraded to support the proposed candidate technology.

4.3.4 Description of Candidate Technology

In order to describe this technology candidate, Figure 4-27 depicts an exemplary radio access and backhaul network deployment envisioned for a typical European city.

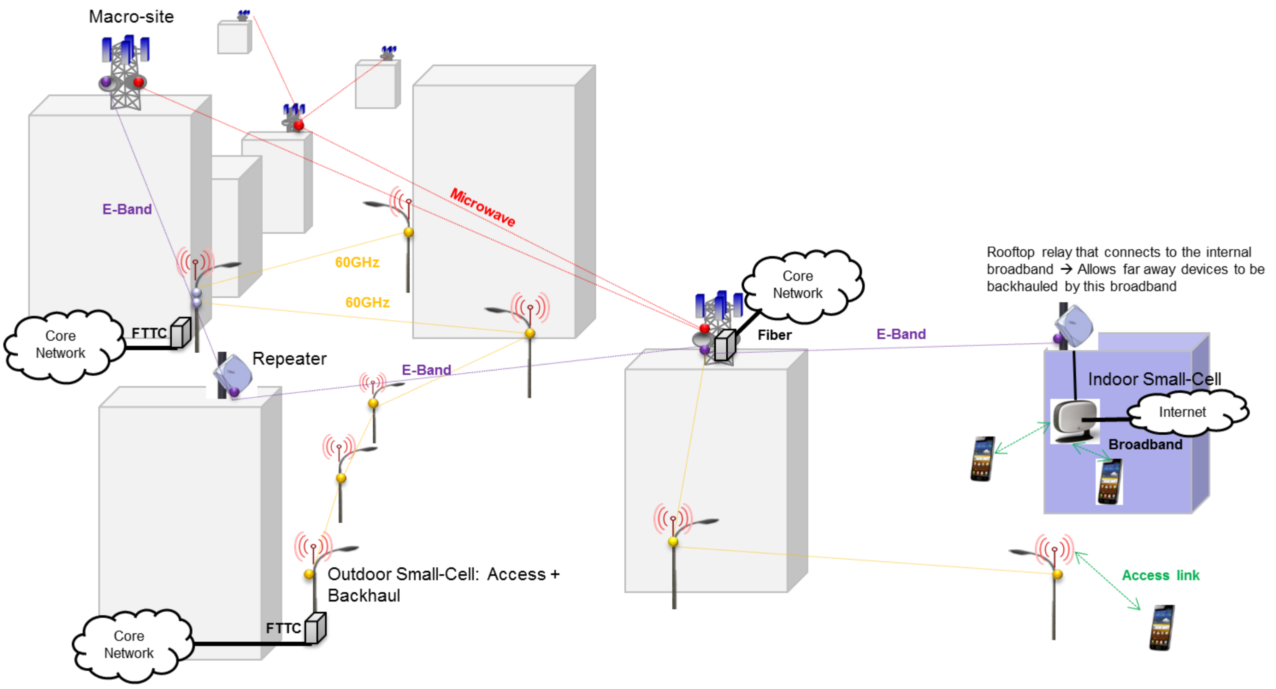

Figure 4-27 Exemplary radio access and backhaul network deployment with heterogeneous backhaul technologies

In Figure 4-27, we can see an heterogeneous backhaul network composed of three layers:

- Rooftop layer: Formed by macro-cells and potentially also repeater nodes. This layer provides macro coverage and might be backhauled by means of optical fibre links, microwave or a combination of both. In addition, E-Band links [104] connecting different sites will also be considered as a suitable backhaul technology in this layer.
- Street-level layer: Formed by small-cells installed at street level, for instance mounted on a lamp post, or attached to a wall. We consider these small-cells to be backhauled with multi-hop 60GHz links and occasionally also with fibre connections.
- Offload layer: Formed by, typically indoors, small-cells that are backhauled using DSL or optical fibre links, and that can be used to offload the cellular traffic in the case of network congestion.

In addition, these three backhaul layers are connected such that traffic can flow to the core network using the best path at any moment. Therefore, the goal of this candidate technology is the design of a system that manages the resources of this heterogeneous backhaul while deciding the best path for each flow. The design of the system we envision will be based on the following principles:

- Centralisation: We envision a coordinator node, typically located in the operator's core network, which will have a global view of the state of the access and backhaul networks in order to allocate resources in an optimal way.
- Continuous Monitoring and Actuation: The coordinator node should be able to continuously monitor the state of the network, which requires the network nodes to be able to measure and report information. In addition, the coordinator node should be able to configure remotely the network nodes.
- Fast reaction: Different network management actions will require actuation at different time scales. We envision the coordinator node to be able to react quickly upon disruptive events such as congestion or node failures, which need to be handled in the order of seconds. Other network optimisation can operate at larger timescales, e.g., minutes.

Another aspect to be considered in the design of this candidate technology is the way how to define flows that need to be engineered across the backhaul. Of particular interest is the granularity at which flows are defined, because too coarsely defined flows might result in a lack of flexibility for the resource allocation, and too finely defined flows could result in an excess of overhead, e.g. in terms of signalling load, switch memory occupancy or computation capabilities of the coordinator node. Therefore, a critical aspect of the design is to provide a flexible flow granularity that can be adjusted depending on the network conditions.

In order to define the scope of backhaul flows, it is useful to illustrate how a typical LTE deployment operates. For this purpose, Figure 4-28 depicts a typical integration of an LTE and backhaul networks [105] [106] from a protocol point of view:

- A UE attaches with an eNB through an established Data Radio Bearer (DRB) that is identified with a QoS Class Identifier (QCI). Thus, the UE sends encrypted user level packets over the radio access.
- The bearer continues between the eNB and the S-GW with the S1-U bearer, where packets are transmitted by means of GTP (GPRS Tunnelling Protocol).
- In LTE, the user data encryption terminates at the eNB. Therefore, another secure channel is required between the eNB and another node in the secured operator premises. In this case, we assume that an IPSEC tunnel is used between the eNB and a security gateway inside the operator premises. Therefore, the backhaul nodes are oblivious to the user data traffic transported within the IPsec tunnel.

Despite the IPsec tunnel, QoS information can still be made visible to the backhaul nodes. In particular, the eNB can encode the DSCP of the GTP-IP header according to the QCI of a particular bearer. Further, the DSCP code of the GTP-IP header can be copied to the outer IPsec-IP header, hence becoming visible to the backhaul nodes. Finally, in the case of an Ethernet backhaul, the DSCP code of the IPsec-IP header could also be reflected in the VLAN tag inserted in each packet.

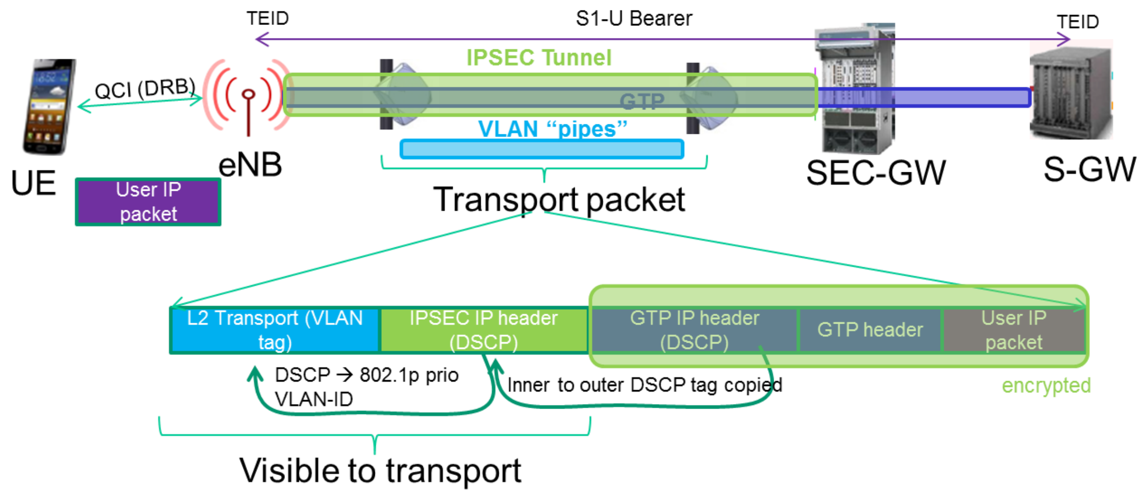

Figure 4-28 Protocol view on radio access and backhaul network

Motivated by the previous description, we will assume an Ethernet based backhaul and consider the definition of a backhaul flow to be encoded in the Ethernet header. In particular, in order to achieve a flexible flow definition we will make use of the QinQ approach defined in [107]. Figure 4-29 illustrates the structure of a QinQ tag with an inner and an outer VLAN tag, each composed of a 12 bit long VLAN-ID and 4 priority bits.

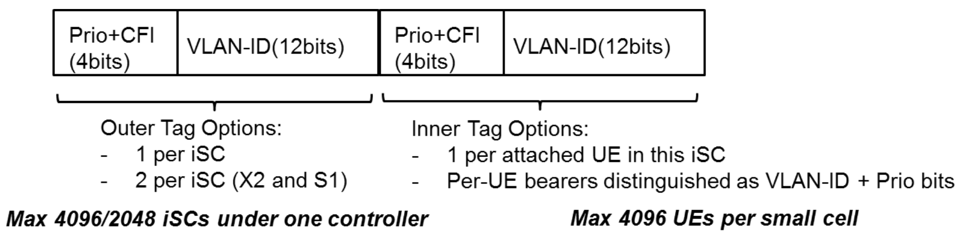

Figure 4-29 Q in Q

Different flow definitions are possible with a QinQ tag structure. We will study as initial definition the following:

- One or two VLAN-IDs in the outer Q tag can be reserved for each iSC, e.g. in the case of two VLAN-IDs one could be used to forward the S1-U/J1 traffic and the other one can be used to forward the X2/J2 traffic. Hence, looking only at the outer Q tag, all the traffic coming from a given iSC would be considered to be a single flow.
- If the coordination requires further granularity the inner Q tag can be used. An example would be to use the VLAN-ID in the inner Q tag to represent each UE attached to a given iSC. In addition, per-bearer granularity could be achieved by considering together the VLAN-ID and the priority bits (set according to the QCI) of the inner Q tag.

Figure 4-30 depicts an example how the QinQ tag can be used to achieve a flexible granularity in flow definitions. It shows the forwarding table of a given iTN specifying that the traffic with outer VLAN 3 and inner VLAN 2, i.e., the traffic coming from a specific UE within the small-cell represented by outer VLAN 3 should be forwarded along the red path, while the rest of the traffic coming from the same small-cell should be forwarded along the black path.

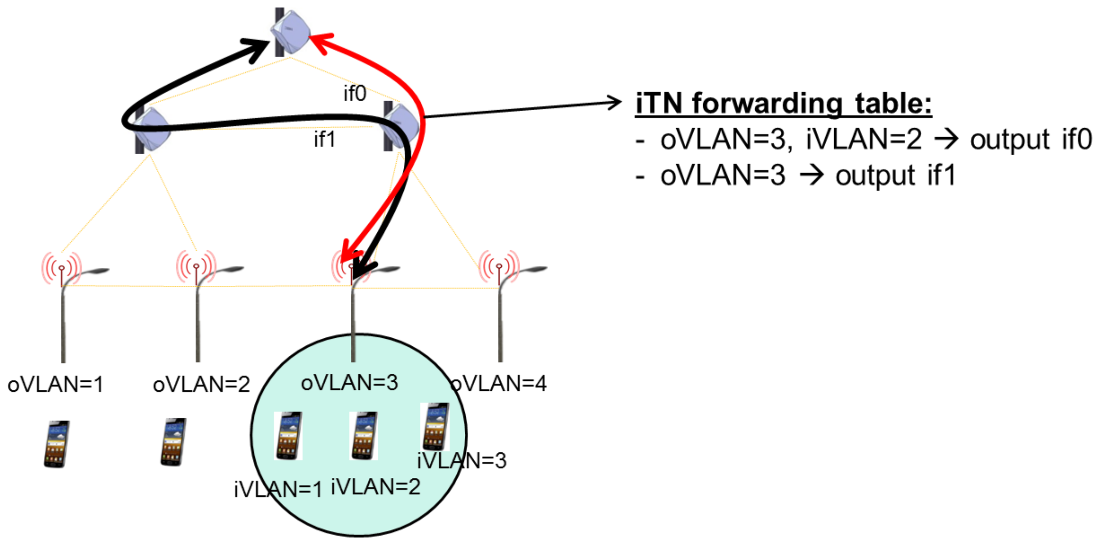

Figure 4-30 Example for QinQ tag used in iJOIN

4.4 CT 4.4: Routing and Congestion Control Mechanisms

4.4.1 Motivation

In this section, the specific solutions for iJOIN congestion control in the access network are described, for the two lines indicated in section 3.4. On one hand, AQM/ECN mechanisms for avoiding the congestion in the backhaul network, and, on the other, solutions for avoiding the interference of these mechanisms with those implemented at the edge of the network, and more specifically, LPCC mechanisms.

The necessity for a specific iJOIN congestion control technology derives from the need to adapt the congestion control mechanisms used in the backhaul network to the flexible functional split supported by RANaaS. In this sense, the iJOIN congestion control functionality should be prepared to handle situations like the one represented in Figure 4-31.

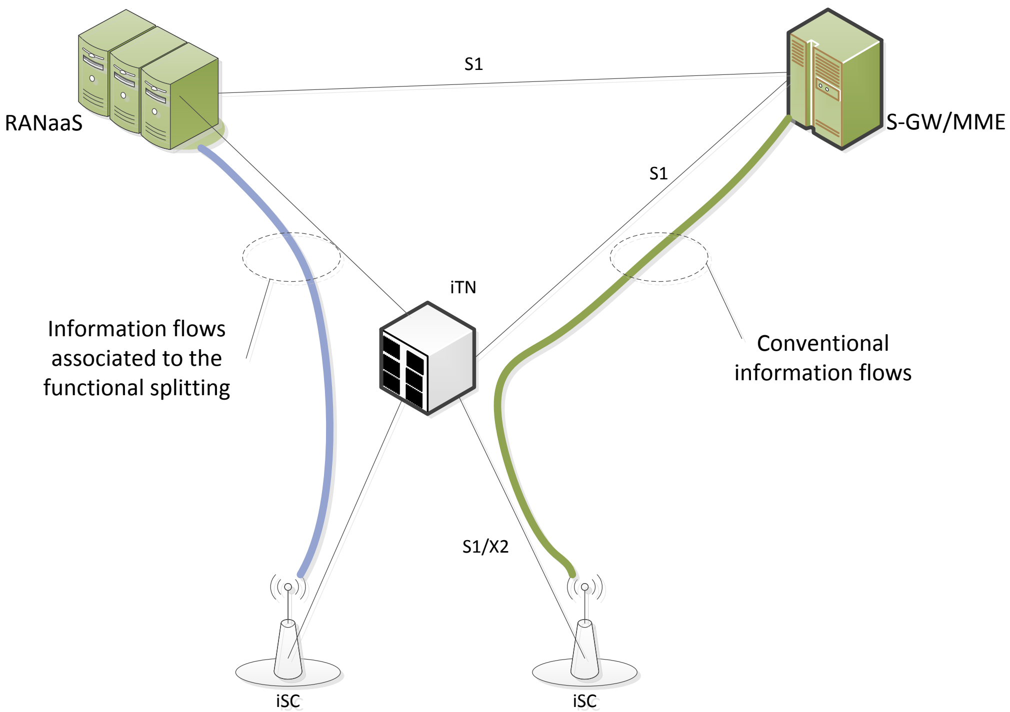

Figure 4-31 iJOIN congestion control use case (example)

In this context, it is visible that congestion control becomes more complex, as actions should take into account, for conventional S1/X2 flows, the characteristics of the QoS associated to the user/application (e.g., the QCI), and for non-conventional flows, the specific functional split that is being supported. But both types of requirements should be managed jointly.

The deployment of small cells poses challenges on routing in the network layer, such as UEs performing handovers between small cells very frequently, traffic-aware portioning of small cells into either static or dynamic clusters causing congestions towards one central entity. Therefore, cooperation of small cells for an enhanced routing algorithm is required. Within this technology candidate, a novel routing algorithm will be investigated beyond the classic centralised/distributed routing algorithms. It is important to investigate the routing and admission/congestion control issues jointly to optimise the load distribution between small cells/backhaul nodes and provide seamless connectivity considering user's mobility between small cells. Routing algorithms should address the congestion control issues by avoiding traffic to be routed through the backhaul nodes which are already overloaded.

4.4.2 Assumptions

- AQM will be implemented by the operator as part of its backhaul infrastructure.
- Backhaul is an IP-based network.
- Ethernet is be used as transport container for all the different functional splits supported. QinQ protocol (or equivalent, like MinM) is assumed to be used to identify either individual or groups of information flows.
- Small-cells are connected to the RANaaS through iJOIN Transport Nodes (iTN).
- An interface is required between small cells for signalling exchange (J2 interface).
- Network topology and network utilisation information for small cells and backhaul nodes are available upon request.

4.4.3 Technical Requirements

- Delay measurements are performed in order to estimate the queuing delay. While the time synchronisation itself is not relevant, the clock skew is of particular interest as it may imply that measurements are biased.
- Delays that are implied by scheduling need to be measured and taken into account.

- Delay measurements need to consider route changes and incorporate them.
- Mobile network congestion management should be configured in such a way that policy enforcement to overcome congestion is activated before end-to-end TCP congestion mechanisms begin to operate.
- ECN mechanisms should be enhanced to deal with the case of several operators sharing the backhaul network.
- Novel routing approach for small-cell deployments. The approach should exploit a cooperative routing algorithm. The classic centralised routing algorithm without any intelligence can easily cause congestion towards EPC. Improvements have been made in distributed routing algorithms to distribute resource consumption amongst network nodes whenever it is required. By proposing the novel routing approach, the transport nodes should cooperate with each other to optimise the routing decisions, by selecting the less congested routes. The routing algorithm should be designed to avoid the heavily loaded transport nodes to improve congestion control.

4.4.4 Description of Candidate Technology

This candidate technology considers two parts which regard iJOIN congestion control as well as joint routing and congestion control. Both parts are separately explained but are meant to interwork, which will be detailed in the course of the project.

4.4.4.1 iJOIN Congestion Control

The implementation of the iJOIN congestion control is described in terms of the different operational phases foreseen:

- Configuration phase

  For each new information flow entering the backhaul network, it should be determined the requirements that would determine the congestion control to be applied: required capacity, QoS, latency, ...

  Under the assumptions indicated above the iTN should not necessarily be aware of the kind of information flow that ingress the node. So in order to identify the CC configuration it may use two methods:

  - The first option is the use of signalling, e.g., some QinQ tags may be reserved for specific kinds of flows. For these purposes, a template for each kind of flow should be defined beforehand.
  - A second option is that the congestion control policy to be applied to the new flow is determined by the iNC. This option provides increased flexibility (e.g., the congestion policy may take into account the user profile).

  It is considered that both options should be supported, the second one as preferential, the first one when the connection to the iNC is not available.

- Measurement phase

  Once the congestion control entity has been configured, the functional entity should initiate a measurement phase in order to be able to detect situations that may require some action. The main measurements are the ingress and egress queues levels (possibly distinguishing between different priorities), but other parameters can also be measured, like the processing time per packet.

  The measurements can also be virtual. For example, for supporting the very low latency requirements that may be associated with some functional splits (e.g., when baseband processing is carried out at RANaaS), it is required that congestion signalling happens before any queuing occurs. In this sense, a mechanism called Phantom Queue (PQ) for creating bandwidth headroom has been proposed for congestion control in data centres. The PQ simulates queue buildup for a virtual egress link of a configurable speed, slower than the actual physical link (e.g., running at 95%of the line rate). The PQ is not really a queue since it does not store packets. It is simply a counter that is updated while packets exit the link at line rate to determine the queuing that would have been present on the slower virtual link (the idea is that it would mark packets that pass through it when the simulated queue is above a fixed threshold.

- Action phase

This phase activates when it is detected that, according to the policy configured, some actions should be carried out. The kind of actions that can be performed are: marking of packets to signal congestion, changes in the queues size, introduction or elimination of queues, selective dropping of packets/frames, explicit signalling to other network elements (e.g. PCRF) of congestion situation,...

For conventional information flows (X2/S1 interfaces), the action that can be performed by the iTN would be either modifying DiffServ field in the IPv4 or IPv6 header in selected flows to announce congestion, or dropping packets in selected flows (or both). Other actions are possible, such as the use of alternative routes to offload the congested node, using, for example, multipath TCP/IP as illustrated in Figure 4-32, the use of SLAs to prioritise traffic flows, and the activation of mechanisms that may reduce overhead, for example, header compression, use of different security mechanisms (MACsec instead of IPsec), packet aggregation of no delay sensitive flows, etc.

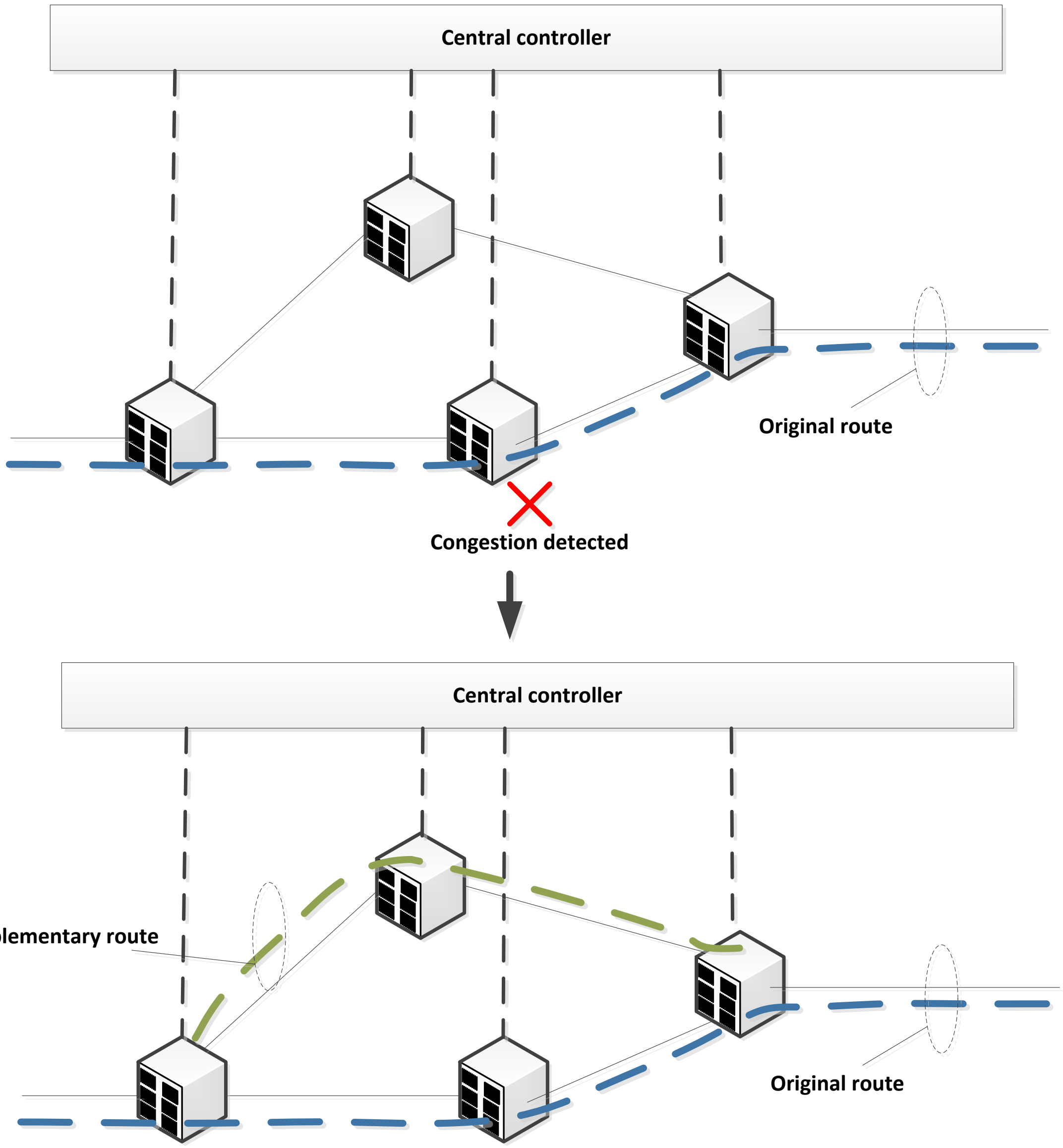

Figure 4-32 Use of alternative routes for off-loading congested nodes

- Reconfiguration phase

  When the congestion control action phase is activated, it may be necessary to carry out a reconfiguration of the congestion control parameters of the congested node and/or other nodes in the network. Also the RAC should interact with TEEM in case the actions may require changes in the traffic engineering policies.

Figure 4-33 represents a simplified vision of the messages that may be exchanged in each phase.

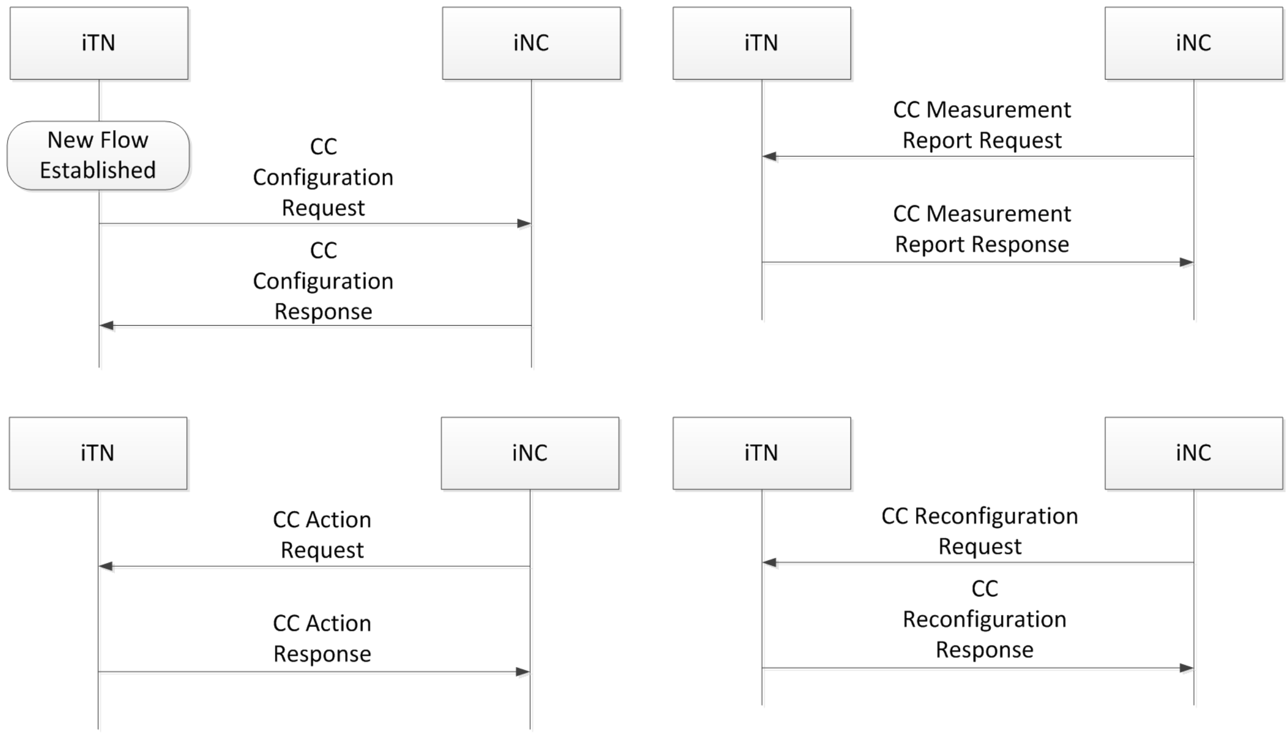

Figure 4-33 Congestion Control process message exchange

In addition, an LPCC based solution is investigated which builds upon LEDBAT. The proposed solution will be implemented in the iJOIN Transport Nodes (iTNs). The basic functional blocks of this solution are:

- Detection: mechanisms for detecting which flows are considered as low priority and may be affected by LPCC mechanisms. It should be based on the relationship between the variable delay and the bit rate of the flow.
- Baseline delay equalisation: mechanisms to keep the baseline delay as constant as possible. This can be done by packet inspection, looking at timestamps, or by means of leaky bucket type of mechanism.
- Baseline delay modification, in such a way that it can compensate events like handovers or activate congestion control mechanisms in order to prevent congestion situations. This can be achieved by properly modifying the parameters of the baseline delay equalization mechanism.

In iJOIN, a solution based on AQM is investigated which is based on traffic policing and assuming that the proposed iJOIN architecture enables more sophisticated congestion control in the backhaul network. Upon detection of congestion there are a number of alternative actions that may be pursued,

4.4.4.2 Joint Routing and Congestion Control

Figure 4-34 shows an outline of the joint routing and admission/congestion control solution which is further investigated within the iJOIN project. We summarise the key features of the solution as below:

- It can be considered a mesh base approach, indicating that multiple transport nodes and paths in the backhaul network are available between small cells and the EPC. However, the proposed solution goes beyond existing mesh-based approaches exploiting queue length and geographic information.
- The transport nodes cooperate with each other. The corporation can be either achieved by a central controller (iNC) or by exchanging of signalling information between transport nodes. In the SDN based approach, the iNC has the up-to-date global knowledge of RAN and backhaul network and can make decisions on routing on behalf of the network. The information of current network utilisation conditions of transport nodes should be available in iNC. There are two ways for collecting network utilisation statistics: either by centralised approach provided by iNC or by initiating requests to neighbouring transport nodes. In the centralised approach, the iNC should be able to learn the network topology and load information for transport nodes and commute decision to transport nodes

accordingly. Otherwise, every transport node should request the information by signalling its neighbouring nodes.

- The solution selects the neighbour based on the principle of minimising the congestion towards the EPC. When taking forwarding decisions at a given node, instead of just considering the shortest distance vector as in traditional routing algorithm, the load information in the neighbouring transport nodes is also taken into consideration. As shown in Figure 4-34, the optimised route is selected so that the congested transport node is avoided.
- The solution might be a proactive routing approach. All transport nodes maintain a routing table that contains separate entries for all the possible destinations, which need to be periodically updated. There might be a scalability problem if there are a lot of transport nodes towards the EPC. Therefore, it is important to find a trade-off between the cooperative routing approach and the traditional IP routing/Mobile IP routing.

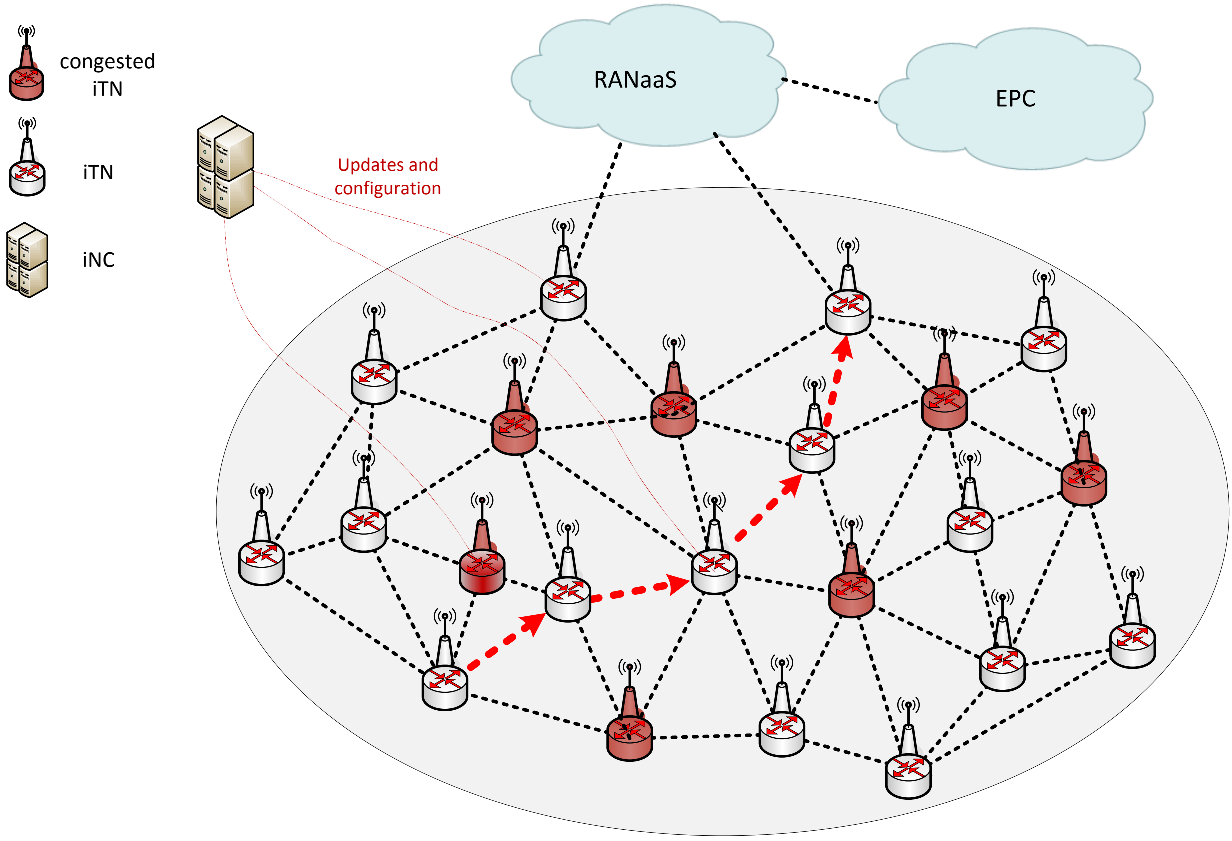

Figure 4-34 Outline of joint routing and admission/congestion control

4.5 CT 4.5: Network Wide Scheduling and Load Balancing

4.5.1 Motivation

Figure 4-35 illustrates a small cell network that potentially utilises heterogeneous backhaul links. In this scenario, packets might need to traverse multiple hops with different capacities and nodes with different queue handling capabilities. This situation can severely impact end-to-end user experience when one or multiple links are congested. It would cause the QoE to degrade significantly. Such cases might require packet multiplexing for various traffic classes with smart priority queuing. Better user experience can be realised by *i)* intelligent load balancing over all alternative resources that can serve the same function (e.g., iSCs, backhaul links, middleboxes), and *ii)* with fine grain queue management (e.g., priority queuing) and scheduling of backhaul and radio resources.

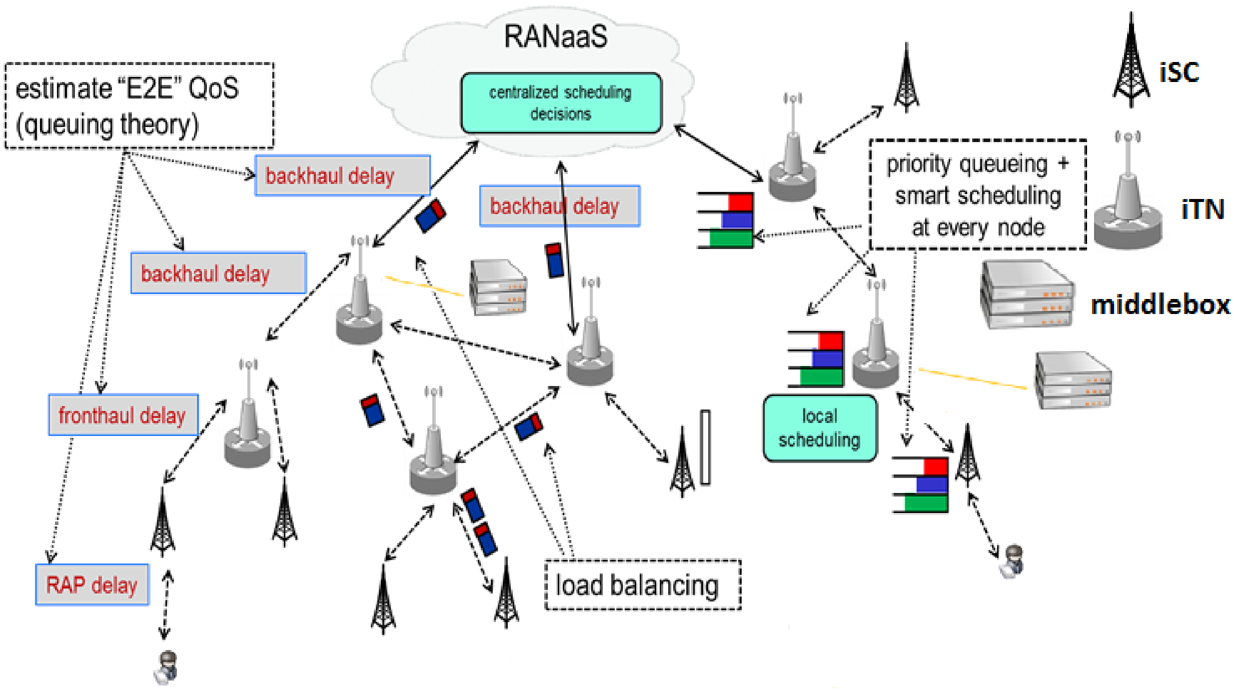

Figure 4-35 Overview of Load Balancing and Scheduling

Thus, as we can see in the above figure (left side), the main aim of Load Balancing is to distribute evenly the traffic, through the (iSC and iTN) nodes, whereas the main objective of Scheduling (on the right side of the figure) is to allocate optimally the resources of each one of the iJOIN elements, based on queuing and smart priority algorithms. When no alternative resources exist to balance, scheduling and priority queuing are the main tools. Otherwise, load balancing and scheduling will be used jointly.

In addition to the optimal allocation of the iSC and iTN resources, "middleboxes" (e.g. transcoders, firewall, etc.) are key elements in current cellular infrastructures, serving large traffic loads, and can thus be virtualised as well as load-balanced and scheduled (see also the SoftCell project [108]).

4.5.2 Assumptions

This candidate technology imposes the following general assumptions:

- Network topology is known.
- Capacity of various backhaul links, node buffer capacity, and node buffer (queue) occupancies are known.
- Long(er) term statistics (e.g., average load, variability etc., as a function of time-of-day) per element (iSC, iTN, middlebox) are known.
- The current availability of radio resources for iSC is known.
- Application requirements and constraints of different traffic types are known.

4.5.3 Technical Requirements

The technical requirements for this candidate technology are listed below:

- Introduction of the "middlebox" concept (physical or virtualised) in the architecture.
- Users' demands for QoS are known.
- Fast and reliable handover mechanism which ensure non-disruptive behaviour when flow are assigned to different iSCs or iTNs for load balancing.
- Fast and stable path rerouting/topology update schemes that can be relied upon when topology (e.g. backhaul) should be modified to better balance load.

- Virtual queue mechanisms introduced in most elements of interest.

4.5.4 Description of Candidate Technology

This CT considers two parts to iJOIN load balancing control as well as scheduling control. We explain first each part separately, although in many cases the two mechanisms will be applied (and optimised) jointly.

Load Balancing

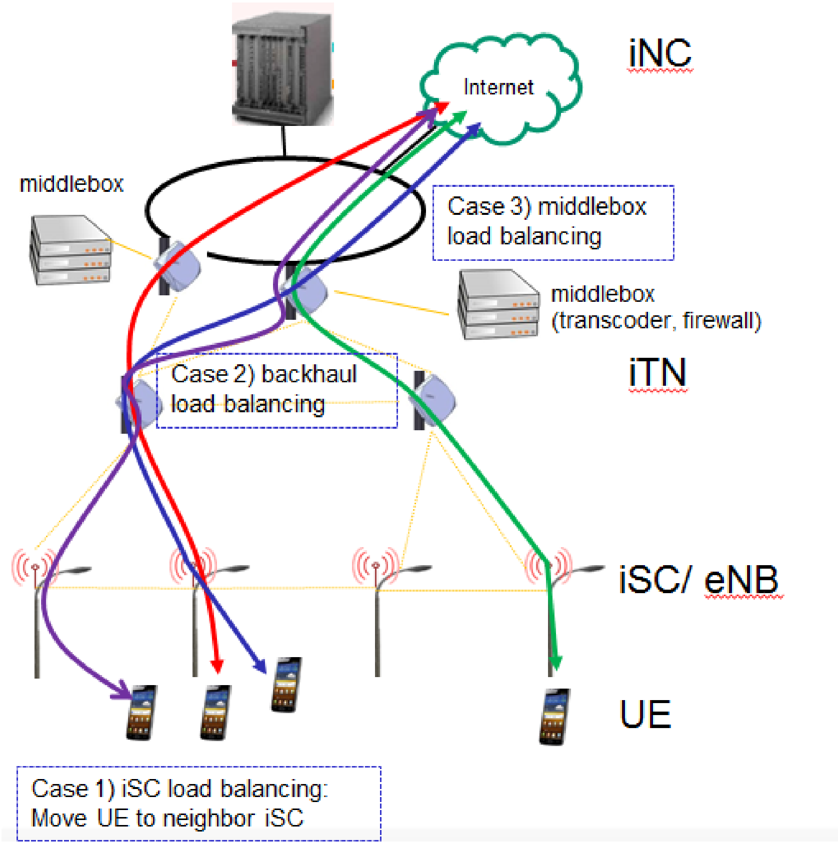

Figure 4-36 Load Balancing example

Figure 4-36 shows an outline of the load balancing CT to be further developed within the iJOIN project. Its goal is to distribute evenly the traffic over iJOIN elements, improve performance and pre-empt congestion by taking into account the complete end-to-end (E2E) path. The key point here is that the iNC decides how the overall (traffic) load should be distributed along different (alternative) elements, taking into consideration iSC and iTN nodes jointly. In the following, we summarise the key features of the solution. For simplicity, we “break” the load balancing functionality into three general use cases as follows:

i. *iSC load-balancing*: if the traffic that an iSC serves is relatively high or on the verge of congestion (e.g., due to a high number of serving users or a few “heavy” users), load-balancing should decide to move a few users to (one or more) neighbouring iSCs. Figure 4-36(case 1), indicates this use case in which we can transfer the UE associated with the blue line, to the neighbour iSC (to the iSC located on the right of the current attached) to achieve load-balancing between iSC. This can be explained, if we notice that the iSC attached in the considered UE serves two users (in more complex examples could be a higher number of attached users), and the iSC on the right serves no one; so we can balance the load between the iSCs with this transfer.

ii. *iTN load-balancing*: The iNC should also select the best end-to-end path over different iTNs, for each new flow or user, in order to balance the traffic optimally. For example, we consider the following scenario in Figure 4-36(case 2): the UEs associated with the green, blue and purple line are being served from the iSC/iTN illustrated in the figure. At that point, the user associated with the red line turns on its UE and starts generating new flows; the iNC should select the “best” path for this UE. In this scenario, load measurements on the iTN serving the UEs associated with the green/blue/purple lines reveal a utilisation of 80%; thus, assigning additional traffic through it would lead to high delays for both the new red flows, as well as the existing ones. As a result, the best path is the one illustrated in the figure with the red line.

iii. *Middlebox load balancing:* middleboxes are necessary and ubiquitous components in recent cellular architectures with many if not most flows requiring passing by one or more boxes along their path. Hence, the iNC should select the proper middleboxes for each user requests, in order to distribute the traffic optimally. Figure 4-36(case 3), indicates a load balancing example for middleboxes. Contrary to use case 2, the iTN serving the 3 existing UEs is not congested and could handle the new (red) user. However, it is the middlebox serving the 3 UEs that is highly utilised, which forces the new UE to be routed through the middlebox on the left.

In addition, we are going to consider different task assignment policies. One promising task assignment policy seems to be the Size Interval Task Assignment (SITA) policy [109] (illustrated in Figure 4-37), since we have to deal with "high-variability" flows (each user asks a variety of different requests; so, each one of the different requests needs a different amount of resources).

For example, SITA splits up user requests based on size, so that requests smaller than a given cutoff are sent to one iSC, and requests larger than the cutoff are sent to another iSC (this premises to ensure, that a set of users/flows can "speak" and be attached to 2-3 different iSC at the same time). This is typically used under high variability in the job sizes (service requirements), because it allows short requests (e.g. pressing the "like" option in a photo, on facebook or writing a static comment in a blog) to have their own freeway and not get stuck behind long requests (uploading an album of 150 HD photos, on a server).

Also, SITA can be considered when an iSC tries to transfer requests to iTNs. Thus, if an iSC can be attached to a set of 2-3 different iTNs, then if you have highly variable data flows you could use SITA to send small flows to one iTN and large to another.

An example of SITA in the every-day life is in the supermarket: where there is an express-line for short jobs.

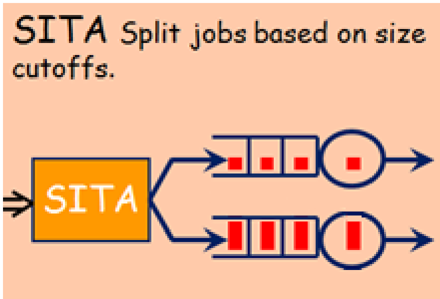

Figure 4-37 SITA overview [109]

Although we have discussed each use case separately, it is important to stress that the iNC takes decisions about the optimal distribution of the traffic flows, by taking into consideration jointly the iSCs, the iTNs and the middleboxes.

The load balancing procedure that the iNC follows can be summarised as follows (depicted in Figure 4-38):

- Step 1: The iNC collects measurements from iSCs, iTNs, and middleboxes (e.g., about their load, utility).
- Step 2: The iNC runs one (or more?) algorithms in order to decide the best combination of (iSC, iTN) nodes, for a given flow such that the load is balanced optimally (and such that the probability of congestion is minimised). If at least one path is found (and no conflicts are noticed from other components), then the chosen route is assigned.
- Step 3: We configure the current flow template (QinQ-oriented approach) to the iSC.
- Step 4: Finally, we update the forwarding tables for this QinQ in iTN nodes.

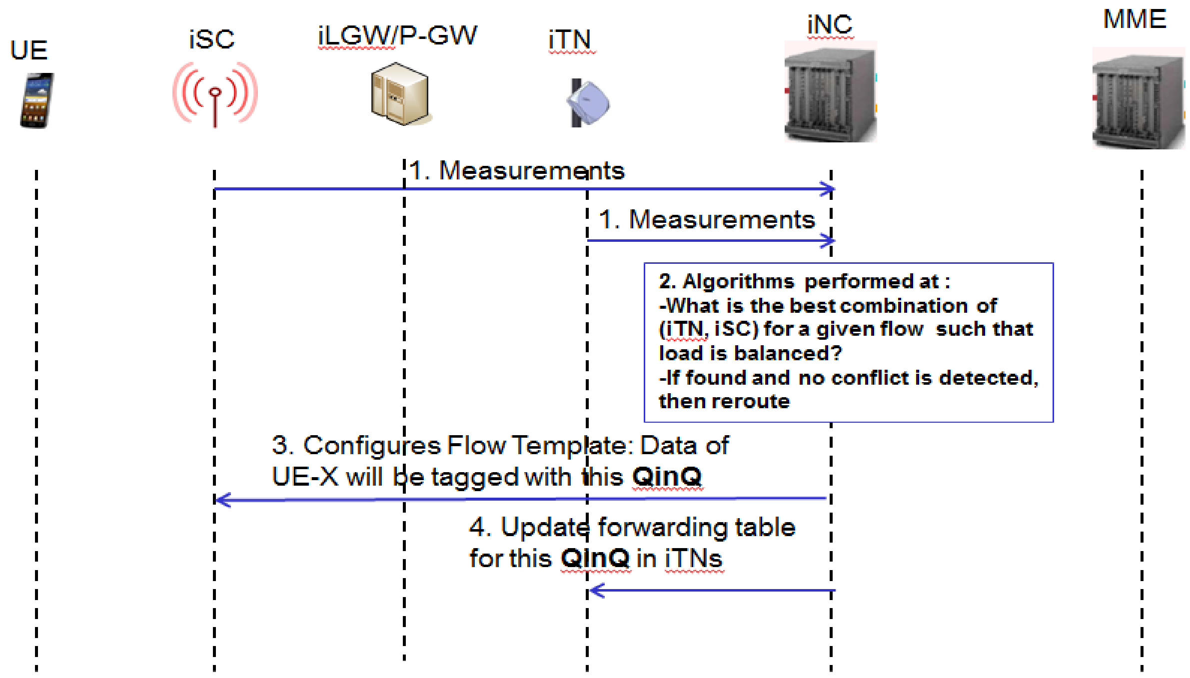

Figure 4-38 MSC of Load Balancing

Scheduling

Figure 4-39 shows an outline of the Scheduling CT to be further developed within the iJOIN project. As illustrated in Figure 4-39, we introduce virtual queues, traffic classes and related priorities in order to achieve optimal scheduling.

Thus, we assume that each one of the iJoin elements (iSC/ iTN/ middleboxes) will queue flows from/to UEs separately. Specifically, it implements a *virtual queuing system* based on traffic classes and priorities for each flow/user (e.g. video traffic get higher priority than file). This system targets the maximum utility of flows subject to the available resources of iJoin elements along the path of a flow. This is achieved by (i) assigning the right flow to the right (priority) queue, and (ii) by choosing the scheduling policy for each queue (e.g. First Come First Serve, Processor Sharing, Shortest Job First, etc.). The scheduling parameters (priorities/traffic classes) change dynamically, if we consider different spatio-temporal policies.

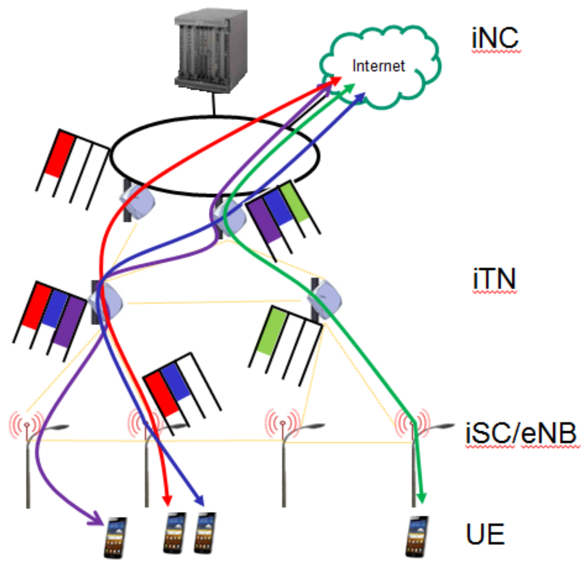

Figure 4-39 Scheduling example

The Scheduling procedure that iNC follows is summarised below (depicted in Figure 4-40):

- Step 1: The iNC collects measurements from iSCs and from iTNs (e.g. about their load, queues).

- Step 2: The iNC decides on a (virtual) queueing discipline for each network element, choosing priorities and classifying the traffic accordingly.
- Step 3: The iNC applies a new configuration to the current flow template (QinQ-oriented approach) to the iSC/iTN, in order to implement the desired queueing discipline(s)
- Step 4: Finally, the iSC and iTN serve their queues, obeying the instructions of the iNC.

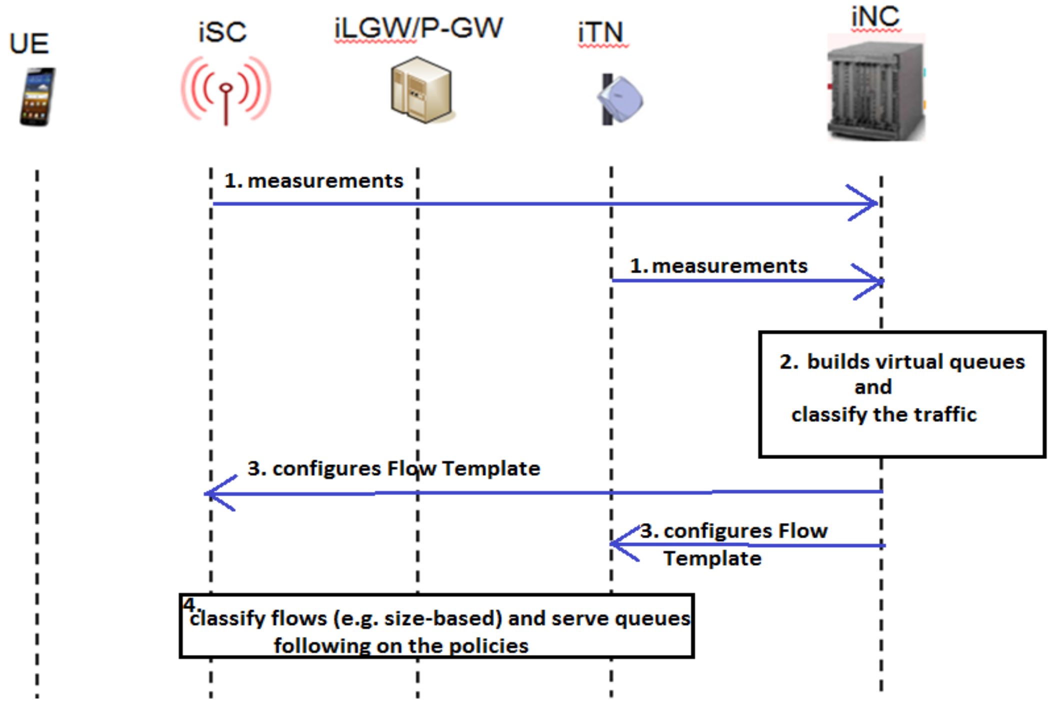

Figure 4-40 MSC of Scheduling

Interaction of Load Balancing with NEO and other modules

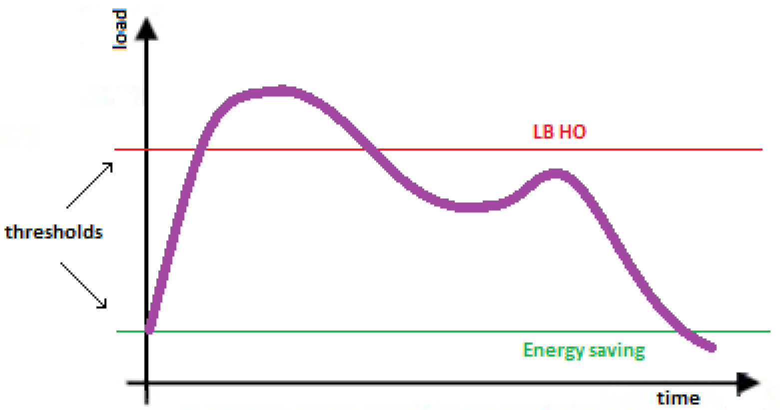

Figure 4-41 LB and Energy saving procedures

On the one hand, when an iJOIN element (e.g., iSC or iTN) is “under-utilised”, the NEO component, tries to push more flow/users to neighbour nodes in order to switch off the low-utilised node. On the other hand, when an iJOIN element is “highly–utilised”, *Load balancing* tries to assign or reroute flows to alternative nodes that are underutilised, in order to balance the load and improve user QoE. Thus, we can see that there is a direct interaction between these two CTs. The former attempts to reduce the usage of network resources, to improve (energy) performance from the perspective of the infrastructure provider, while the latter attempts to increase the usage of some network resources in order to improve the performance of users. Evidently, these two goals can be conflicting, and interaction between the two is needed, either directly, or indirectly (e.g., in the TEEM) to allocate a “cost” or “value” on the suggested decisions of each policy, and decide on the best trade-off. Nevertheless, Load Balancing and NEO can be seen as two policies applied on the same problem setup: the utilization of each element is monitored, and two thresholds are introduced: *i)* if utilisation drops below a “low” threshold it triggers NEO; *ii)* if utilization exceeds a “high” threshold it triggers LB, as depicted in Figure 4-41 [110]. Also, we can assume a "medium" threshold that triggers reports for a prospective "load balancing" action.

Finally, there is also an interplay between the load balancing and topology and congestion control modules. The load balancing module is triggered to avoid congestion at each network element, while the topology and congestion control model is triggered typically when the maximum capacity is reached. Once more, the TEEM module is in charge of conflict resolution among such requests received from different modules.

4.6 CT 4.6: Backhaul Analysis based on Viable Metrics and "Cost" Functions using Stochastic Geometry

4.6.1 Motivation

Heterogeneity in wireless networks implies increased randomness in base station deployment. This in turn leads to a scenario where the backhaul for such networks has to be highly adaptable. Design and deployment of such backhaul infrastructures depends on our ability to analyse networks of today and assess various methods of improvement using a common metric or standard. Stochastic geometry provides one such method of obtaining a metric that can be used as a benchmark for comparison. This approach deals with base stations as points of a point process, wherein system parameters such as transmit power, fading, path-loss, etc. are treated as functionals (or attributes) of each of these points. This can then be used to analyse the probabilities of coverage, spatially averaged rate or spectral efficiency. In such models, the backhaul can be described as a higher layer of points (distributed according to another point process) which is superimposed upon the base station layer and imposes certain restrictions on the layer below (i.e. base station layer), e.g., a throughput limit. The major advantage of using such a model is the fact that such an analysis observes the "average" behaviour of various system parameters by taking an expectation over infinitely many realizations of the point process. This implies that every network topology that can exist has been implicitly included in the observations. This is also the reason why such models can prove to be effective benchmarks against which other models can be compared.

4.6.2 Assumptions

- The backhaul network is considered to be the topmost layer of a multi-layered network with a point process describing the components (BSs or users) of each layer.
- This layer imposes constraints on the "cost functions" of interest that need to be evaluated. These cost functions could be energy consumption parameters or CAPEX/OPEX.

4.6.3 Technical Requirements

The following technical requirements are imposed by the above described analysis:

- The mathematical model requires that users are not allowed to connect to the backhaul layer.
- For energy efficiency analysis, there exists a mechanism which enables an effective turn on and off of base stations.
- Users are served by the base station closest to it. If the base station is turned off, users are automatically served by the next nearest base station.

4.6.4 Description of Candidate Technology

This approach treats base stations and UEs as points of a point process in the Euclidean plane. Each point of the base station process has some functionals attributed to it. These functionals are typical system parameters such as transmit power, path loss, and fading. They further determine how the given area is divided or tessellated around each base station. The UEs (points of another point process) are assumed to connect to the base station from which the highest power is received (usually the nearest one). Based on such a framework, various performance indicators such as coverage probabilities and spectral efficiency can be observed over many such (theoretically infinite) realisations of the point processes. An expectation over these realizations can result in a description of the average behaviour of the network for a given number of UEs and base stations. With this framework, the backhaul can now be considered to be another layer containing points of a point process which is superimposed on the layer consisting of base stations and UEs. The backhaul layer imposes certain restrictions on the base station point process in the layer below, thereby limiting the performance and altering the performance indicators mentioned above. The interactions between the points of the backhaul layer can be modelled as functionals which vary based on the type of backhaul considered, i.e. wired or wireless backhaul. The model can be visualised as follows in Figure 4-42.

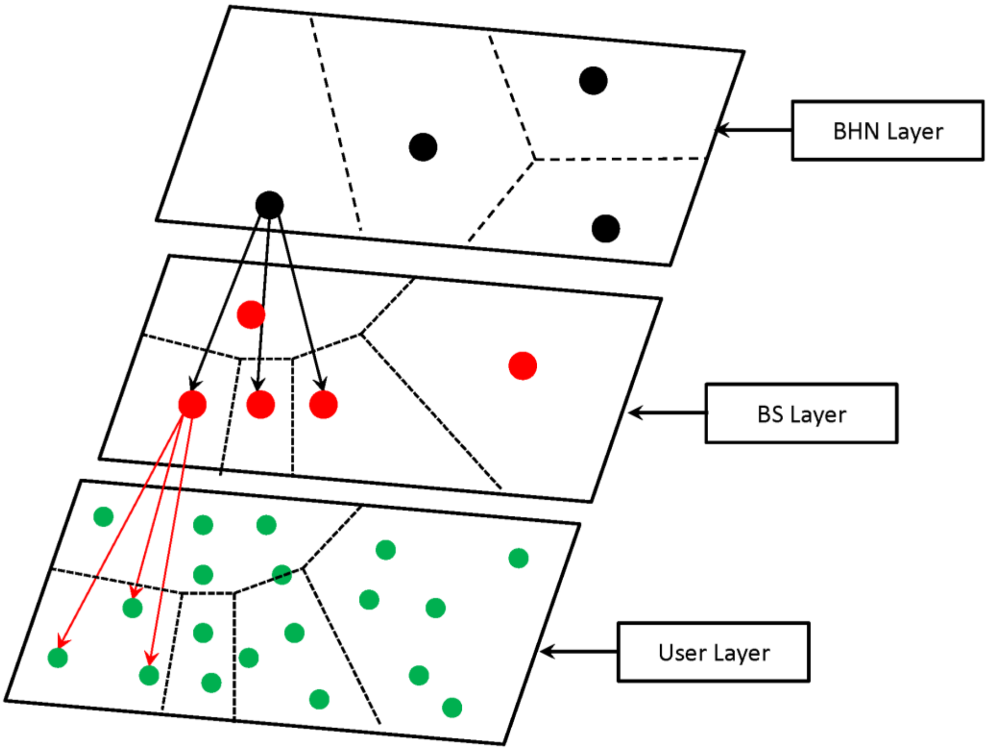

Figure 4-42 Network Model considered for Analysis

The topmost layer consists of back-haul nodes, with the middle layer containing base stations, and the bottommost layer containing UEs. The connection between the UEs and the base stations is based on the maximum SINR criterion, whereas the connections between the base stations and the back-haul nodes could be based on either maximum SINR criterion (for microwave back-haul) or nearest distance (for fibre optic back-haul). This framework now results in a relationship between the performance indicators, base stations, backhaul, and the UEs, which can then be used as a constraint in an optimisation problem to evaluate the effectiveness of the backhaul in terms of CAPEX, OPEX, and energy consumption. The optimisation problems of evaluating the effectiveness of the backhaul can be made based on the average "cost" of deploy an additional back-haul node. This "cost" could be any of the parameters (CAPEX, OPEX, or energy consumption) mentioned above.

5 WP4 Functional Architecture and Interactions of CTs

5.1 Overview

Figure 5-1 summarises the WP4 functional architecture. This architecture is heavily influenced by the Software Defined Architecture (SDN) approach that has been adopted, and the figure reflects so by strongly focusing on the iJOIN Network Controller (iNC) internal architecture. The iNC is a key network entity in charge of configuring, monitoring and driving the operation of the rest of the RAN and backhaul network entities. The extended OpenFlow controller (called *iOpenFlow controller*) is a critical entity located at the iNC, taking care of all the protocol interactions with the rest of the WP4 network entities which only need to support OpenFlow, using the so-called Southbound protocol interface.

Inside the iNC there is basically one module per WP4 CT (except for CT4.6 which is not a run-time mechanism):

- Anchor and Mobility Management (AMM). This module implements most of the functionality of CT4.1, namely the selection of the proper anchors on a per application and UE basis, as well as ensuring that those sessions needing mobility support are provided with it. Note that CT 4.1 is in charge not only of providing mobility support on an address (application) basis, but also to ensure that resources are optimally exploited, both in the backhaul and in the access. This is achieved by selecting and using an anchor closer to the UE. This does not mean that for some flows legacy EPS Rel-10 mobility mechanisms (and anchors, i.e., the PGW) are not used but that they are actually complemented by the iJOIN solutions. Although it is not shown in the figure, there might also be some AMM functionality on the UE for the case of non SDN-enabled iJOIN enhanced UEs, so the overall performance is improved. Analogously, the UE might also be an SDN-enabled entity. This will subject of part of the WP4 research to be conducted in the following months.
- Network-wide Energy Optimiser (NEO). This module, defined by the CT 4.2, monitors the network status and load, and runs different algorithms to optimise the overall energy consumption while ensuring that the network wide performance is not compromised. The module is in charge of taking network wide decisions about switching on/off physical nodes, as well as ensuring that UE traffic is still properly routed by the nodes that are running at each time.
- Routing and Congestion (RAC). This module, defined by the CT 4.4 is in charge of avoiding network congestion in the RAN/backhaul, by properly configuring the network and requesting changes on the paths used by active data flows. In order to do so, both the status of the RAN and the UE traffic requirements are considered.
- Traffic Engineering Enforcement Module (TEEM). This module, defined by CT4.3, is a key WP4 module that hosts all the intelligence required to compute the best path within the backhaul to support the different traffic and network-wide requirements, providing the necessary conflict resolution functions (e.g., to ensure that a request from the NEO module does not introduce congestion or contradicts a previous request from the RAC module).
- Measurement Module (MM). This module, defined by CT4.3, is in charge of configuring the iOpenFlow controller to perform the required measurements. These measurements might be dynamically requested by both the WP4 modules as well as WP2/WP3, and therefore has intra- and inter-WP interfaces. Envisioned metrics to be reported by the MM comprise: locally experienced congestion, available neighbours, available data rates, and number of connected UEs. The MM will support several reporting modes, e.g., asynchronous, periodic or event based
- Network Model Module (NMM). This module also plays a very important role as it is in charge of initially acquiring a topological and functional view of the network (i.e., which nodes are up and running and how they are interconnected, as well as which capabilities they have, for example, in terms of energy configurability of IP anchoring support), building a model of the network that reflects the status and that is kept up-to-date by continuously monitoring the status and load.

It is important to note that given the SDN approach followed, the functionality associated to each CT is not actually provided just by the respective module, but often requires the participation of several of them. A relevant example is the TEEM module, because it is in charge of computing and setting up the data forwarding path in the RAN and backhaul, and therefore, many other CTs rely on it to perform the different reconfiguration actions that might be required. For example, the AMM module may require some traffic to be redirected from one active anchor to the new point of attachment of a given UE after a handover, and this has to be executed by the TEEM module.

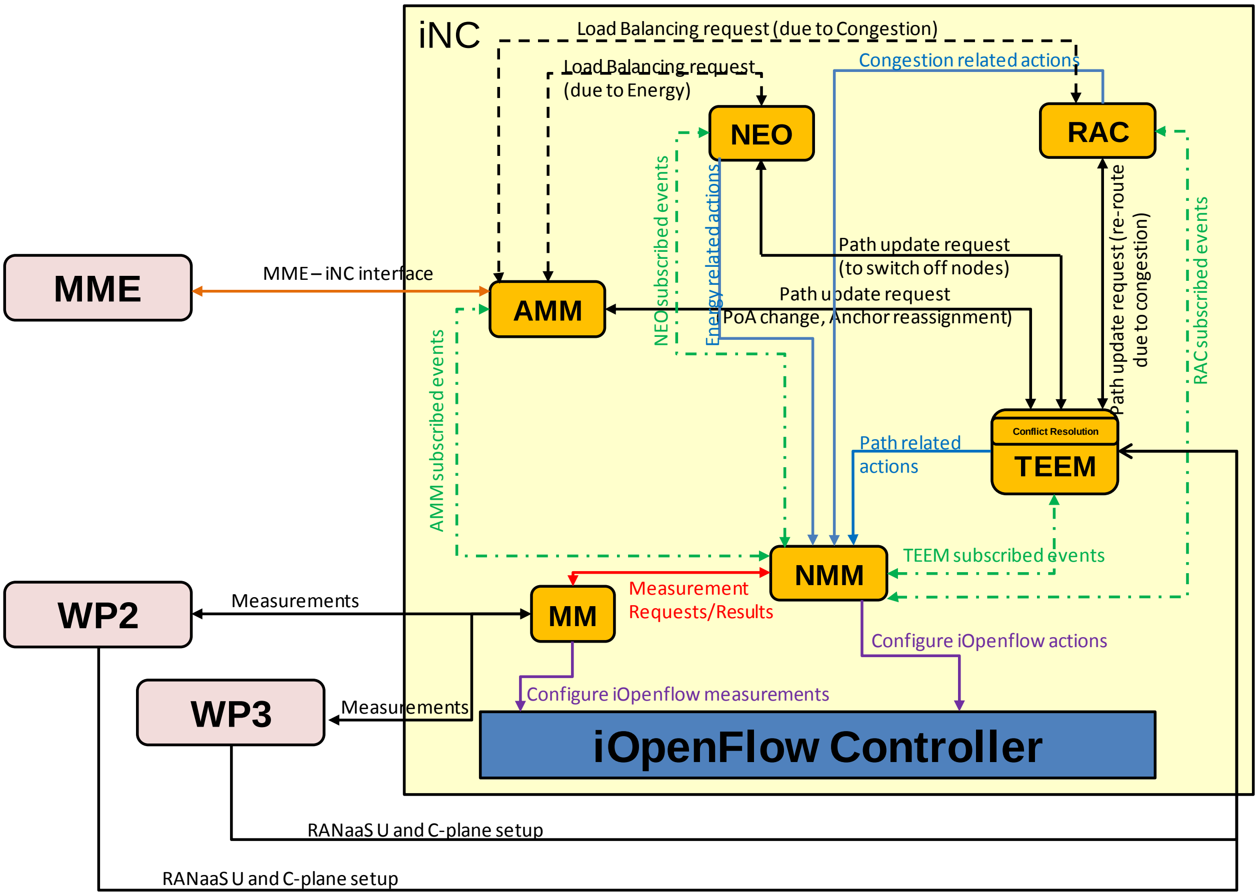

Figure 5-1 WP4 functional architecture

Next, we summarise the different types of interfaces that are shown in the Figure 5-1(grouped by colours):

1. Interface between NMM and WP4 module X:
   - Used by module X to tell NMM what he wants to be informed about. (e.g., RAC tells NMM: tell me when queue in node 1 is above 10 packets).
   - Used by NMM to notify module X about changes on its subscribed information.
2. Interface between NMM and MM:
   - Used by NMM to forward to MM the measurement requests from the modules.
   - Used by MM to update NMM with measurement results.
   - Note: MM takes all the measurement requests and has the intelligence to combine them in order to minimise network overhead.
3. Interface between AMM/NEO/RAC and TEEM:
   - Used by AMM/NEO/RAC to request a path change to TEEM. Note that the TEEM resolves path related conflicts.
4. Interface between WP2/3/4 module X and NMM:
   - Used by module X to commit new network state changes to NMM. Example: Forwarding rules, scheduling parameters.

- TEEM is in charge of any path related changes. RAC is in charge of congestion control parameters that do not involve path changes (e.g., RED thresholds, scheduling weights, etc).

5. Interface between MM/NMM and iOpenFlow Controller:
    - Used by MM and NMM to tell the iOpenFlow controller to write some new configuration/measurement state to the network.

Additionally, there is another key interaction between WP4 and WP2/WP3 currently under active discussion, which is related to the setup of the User and Control plane forwarding for the RANaaS functional split. This interface, critical for the whole iJOIN architecture, will be discussed and specified within WP5 and described in future WP5 deliverables.

5.2 Interface Specification

5.2.1 Consolidated List of Required Input of WP4 CTs with Source Information

IP	CT	Requested Input	Source of Information		
			CT or system function/module	Logical network entity	Parameters
I4.1	4.1, 4.3	Network topology map: points of attachment (cells with L2 associated information), anchor nodes and its capabilities	NMM	iNC	<listof(iSC_ID, iSC_L2_info, listof(available anchors)), listof(anchorID, breakout_capabilities) >
I4.2	4.1, 4.4	Request to move a UE to a different cell because of congestion reason. An order (by priority) list of target PoAs is provided.	RAC	iNC	<UE_ID, listof(iSC_ID, iSC_L2_info>
I4.3	4.1, 4.2	Request to move a UE to a different cell because of energy optimisation reasons. An order (by priority) list of target PoAs is provided.	NEO	iNC	<UE_ID, listof(iSC_ID, iSC_L2_info>
I4.4	4.1, 4.2, 4.4	Request to change default anchor for new UE communications (a filter so it only applies for certain types of new communications can also be provided)	RAC/NEO	iNC	<UE_ID, anchor_ID, [flow filter]>
I4.5	4.1, 4.3	Reply from the TEEM that includes the selected anchor (after filtering considering network constraints)	TEEM	iNC	(UE_ID, PDN connection, selected anchor)
I4.6	4.1	UE attachment / new PDN connection request		MME	(UE ID, iSC ID, PDN connection information)
I4.7	4.1	Reply from selected anchor including assigned IP prefix	AMM	iLGW/PGW	(UE ID, assigned IP prefix)
I4.8	4.3	A notion of "flow" or packet aggregate that iTN nodes can engineer	TEEM	iTN, iNC	<flow identifier/filter, QoS parameters, Source and Destination endpoints>
I4.9	4.3 4.4	Instantaneous (TBD) load requirements of each iTN/iSC.	TEEM	iTN, iNC	<Load requirements (Mbps), Reported time interval, QoS class (optional)>
I4.10	4.3	Instantaneous (TBD) data/error rates	TEEM	iTN	<Data/Error rate, Reported time

		experienced by each transport link.			interval, endpoint>
I4.11	4.3, 4.2	End points of attachment for each flow. Connectivity options for each iTN (e.g. how many other iTN nodes each iTN node can connect to).	MM, TEEM, NEO	iTN, iNC	<list of (one hop reachable iTNs, iSCs) >
I4.12	4.2 4.4	Network topology map	NMM	iNC	<list of (iTN_ID and its one hop reachable iTN_ID)> <list of (iSC_ID and its neighbouring iSC_ID)>
I4.13	4.2 4.4	Instantaneous BH node state information including: -max load capacity per iTN/iSC -remaining load capacity per iTN/iSC -energy utilization status per iTN/iSC	NMM	iNC (optional) measurement report from UE	<iTN/iSC_Max_Load > <iTN/iSC_Rem_Load > <iTN/iSC_Energy_Status> <reported time interval>
I4.14	4.2 4.4	(optional) QoS related parameters	NMM	iNC	<Max bit rate> <Guaranteed bit rate> <Packet delay budget>
I4.15	4.2	iSC ID (UE's current point of attachment)	AMM	iNC	<UE_ID, iSC_ID>
I4.16	4.4	Request to change congestion control for specific flows	RAC	iTN	<flow identifiers, Congestion Control parameters' values>
I4.17	4.1 4.5	Request to move a UE to a different cell because of LB. An order (by priority) list of target PoAs is provided.	TEEM	iNC	<UE_ID, list of (iSC_ID, iSC_L2_info)>
I4.18	4.5	Request to change paths that some flows follow, due to LB or scheduling reasons	TEEM	iNC	<flow identifiers, iSC_L2_info , GW_ID>

5.2.2 Consolidated List of Provided Output of WP4 CTs with Sink of Information

OP	CT	Provided Output	Sink of Information		
			CT or system function	Logical network entity	Parameter
O4.1	4.1, 4.2, 4.4	PoA Attachment/handover result (cell)	NEO/RAC (depending on who triggered the handover)	iNC	<UE_ID, iSC_ID, result>
O4.2	4.1, 4.2, 4.4	Anchor change result	NEO/RAC (depending on who triggered the handover)	iNC	<UE_ID, anchor_ID, flow_filter, result>
O4.3	4.1, 4.3	Path Request: request to compute and install the path between the provided iSC and the first GW to the provided ordered list of anchors	TEEM	iNC	(UE_ID, iSC_ID, listof(GW_ID, anchor_ID))
O4.4	4.1	Anchor Request: request to provide anchor and mobility functions to a given UE	AMM	iLGW/PGW	(UE_ID, UE_capabilities, PDN connection)

O4.5	4.1	Returns IP parameters and active anchors to the MME (which keeps a centralised database, maybe by storing it on HSS)		MME	(UE, listof(anchor_ID, assigned_IP_prefix))
O4.6	4.3	Path followed by each transport flow. Additionally, configuration parameters in each iTN (e.g., the MCS to be used).	TEEM	iTN	<Flow identifier, next hop destination, Transmission parameters (e.g. rate, power, MCS)>
O4.7	4.2	List of switched on/off iSC/iTN	NEO	iNC	<list of iSC_ID> <list of iTN_ID>
O4.8	4.2	Path update to update the active network topology map, taking into consideration NEO and LB results	NEO, TEEM	iNC	<list of (iSC_ID/iTN_ID to be switched off/on)>
O4.9	4.2	Assigning UE/flow to neighbouring iSC/iTN	NEO	iNC	<List of (flow_identifier, iTN_ID)> <list of (UE_ID, iSC_ID)>
O4.10	4.4	Updated routing table: recomputed next hop as per iTN	RAC	iNC	<iTN_ID, routing info>
O4.11	4.5	PoA handover result (cell) due to Load Balancing	TEEM	iNC	<UE_ID, iSC_ID, result>
O4.12	4.5	Path update to update the active network topology map, due to LB or scheduling	TEEM	iNC	<list of (iSC_ID/iTN_ID to be switched off/on)>

List of Abbreviations

Abbrev	Full Name (including explanation if necessary)	LTE or CT specific
Identifier		
NEO	Network-wide Energy Optimiser	iJOIN
AMM	Anchor and Mobility Management	iJOIN
RAC	Routing and Congestion	iJOIN
TEEM	Traffic Engineering Enforcement Module	iJOIN
MM	Measurement Module	iJOIN
TC	Topology Control	iJOIN
iSC_ID	iJOIN Small Cell Identifier	iJOIN
iSC_L2_info	iJOIN Small Cell Layer 2 information	iJOIN
iTN_ID	iJOIN Transport Node Identifier	iJOIN
iTN/iSC_Max_Load	Maximum load capacity as per iTN/iSC	iJOIN
iTN/iSC_Rem_Load	Remaining load capacity as per iTN/iSC	iJOIN
iTN/iSC_Energy_Status	Energy Utilization Status as per iTN/iSC	iJOIN
anchorID	Identifier of the anchor	iJOIN
UE_ID	User Equipment Identifier	LTE
GW_ID	Gateway (router to the selected anchor) Identifier	iJOIN

5.3 Module Specification: Interaction of WP4 Modules

This section provides a high level description of the envisioned interactions between the identified modules in the iNC during major network procedures. Notice that our goal at this stage is not to be exhaustive on the defined procedures, but rather illustrative on the approach that will be followed in the design of the iNC. In particular, we illustrate our envisioned operation of the following network procedures:

- Network Bootstrap (configuration of a new iTN)
- UE attachment
- Congestion Management
- Energy Optimisation

5.3.1 Network Bootstrap Procedure

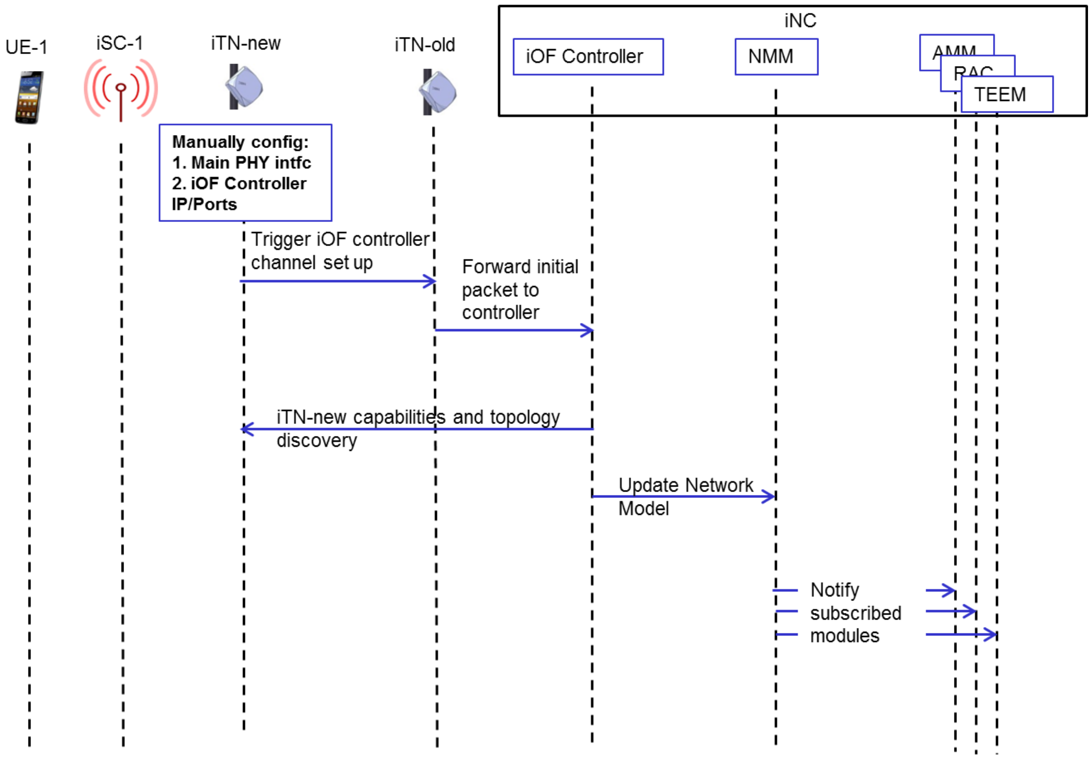

Figure 5-2 Interaction during Network Bootstrap

This procedure is depicted in Figure 5-2. It describes the operation followed by the iNC when a new iTN is added to the transport network. The envisioned sequence of operations is the following:

1. When deploying a new iTN, an operator will first perform the physical site installation of the device. In addition, the operator will provision an initial physical interface for this iTN with any other iTN already connected to the network, and the operator will also provision the initial IP address and port of the iOF Controller. For an easier configuration of the latter step an operator may choose to deploy the OF-CONFIG protocol [111].
2. After the initial manual configuration is completed, the new iTN will try to set up a connection with the provisioned iOF Controller through the defined main interface.
3. When the already existent iTN receives the initial packet from the new iTN, it will automatically tunnel this packet back to the iOF Controller for which a route must already exist.
4. Upon receiving this initial packet, the iOF Controller knows that a new iTN is present in the network and knows that this new iTN can be reached through the iTN that forwarded the initial packet. Having this knowledge, the iOF Controller proceeds with the connection set-up of the new iTN while discovering further capabilities and network interfaces available in the new iTN.

5. Once the iOF Controller finished gathering knowledge about the new iTN, it communicates the new network state to the NMM, which is the module maintaining a representation of the network state.
6. Once the network state is updated, the NMM triggers a "Topology Change" event to convey the new state of the network to the modules (AMM, TEMM, and RAC) that subscribed to that information.

5.3.2 UE Attachment Procedure

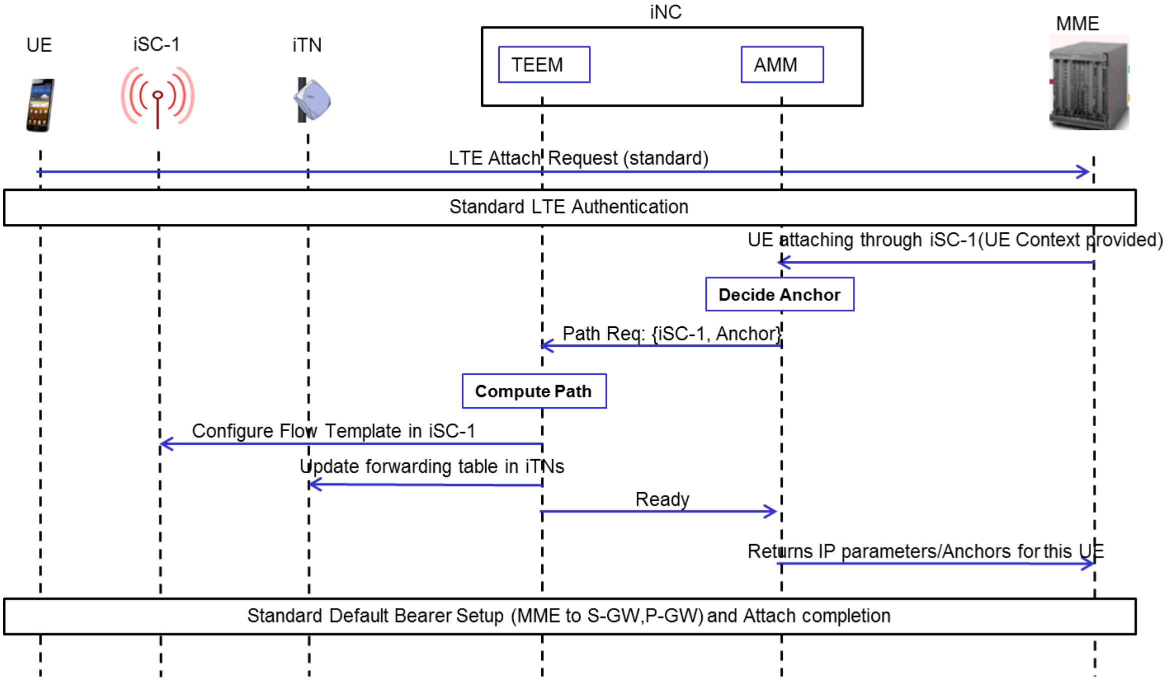

Figure 5-3 Interaction during UE Attachment

The UE Attachment procedure is depicted in Figure 5-3 and describes the individual steps which are executed when a UE attaches to the network. The envisioned operations are the following:

1. An unmodified UE discovers an iSC or eNB, and it sends an Attach Request message to the MME in order to get authenticated using LTE standard procedures.
2. Upon a new attachment request, the MME communicates to the AMM in the iNC that a new UE is requesting to attach through a particular iSC. In addition, the MME provides the AMM with UE Context information that can be useful in order to dimension backhaul resources.
3. Upon receiving the message from the MME, the AMM determines the gateway where this UE should anchor its bearers. Upon deciding on the anchor, the AMM sends a request to the TEMM to provide a backhaul path between the given iSC and its anchor.
4. The TEEM computes the backhaul path that should be allocated to the new UE. If required, it updates the configuration of the flow templates in the iSC to assign a proper QinQ tag to the packets originating from this UE. Furthermore, it updates the forwarding tables of the affected iTNs. Once the backhaul is effectively provisioned to accommodate the new UE, the TEEM sends an acknowledgement back to the AMM.
5. Upon receiving notice that the backhaul path is provisioned, the AMM returns to the MME the IP parameters and anchor that should be used for this UE.
6. Finally, the MME continues with the standard LTE attachment set up.

5.3.3 Congestion Management Procedures

We now describe the procedures related with the way how the iNC manages congestion in the radio access and backhaul network. In particular, we describe two different procedures depending on whether a UE needs to be relocated or not. The first procedure is depicted in Figure 5-4 and corresponds to the case where a UE does not need to be relocated. The envisioned sequence of operations is the following:

1. The RAC module in the iNC detects congestion in the network, i.e. in a particular iTN. In a first instance, the RAC module tries to resolve congestion locally by modifying the local parameters of the affected iTN, as described in detail in Section 4.4.4.
2. After a pre-defined period of time, the RAC module detects that congestion persists and therefore triggers the TEEM for path relocation. Specifically, the RAC module requests the TEEM to reduce the load through the affected iTN by a given percentage.
3. The TEEM then computes alternative paths and after finding a suitable solution it triggers the reconfiguration of the affected network elements. Upon completion, the TEEM sends an acknowledgement back to the RAC module.

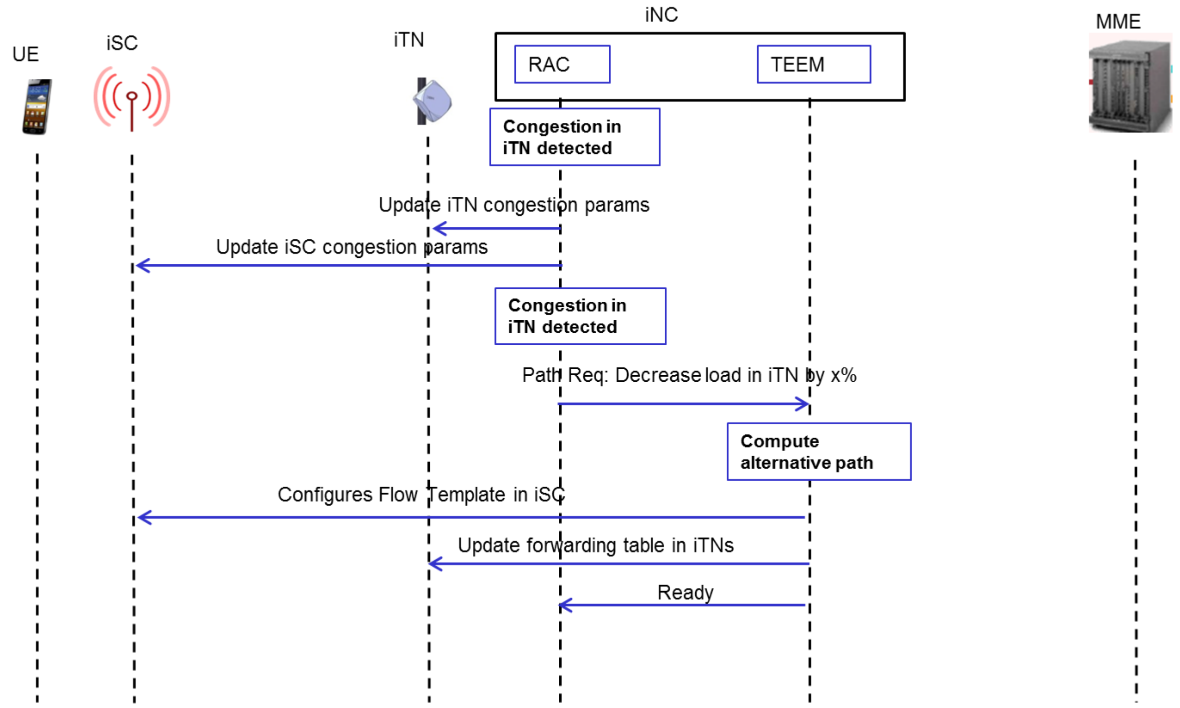

Figure 5-4 Interaction during Congestion Management (no UE relocation needed)

The second congestion management procedure is depicted in Figure 5-5 and corresponds to the case where a UE needs to be relocated. The envisioned sequence of operations is the following:

1. After congestion detection in an iSC, the RAC module may decide to relocate a UE, which triggers a UE relocation request to the AMM module.
2. The AMM module then determines the new iSC where the UE should be attached to and sends a path request to the TEEM in order to reserve backhaul resources for the new UE traffic.
3. The TEEM computes the new backhaul state, configures the affected iSC and iTNs, it and notifies the AMM.
4. Once the resources in the backhaul have been assigned, the AMM sends a request to the MME to trigger a handover. The AMM informs the MME about the new iSC where the UE should be attached to.
5. The MME triggers the corresponding handover and upon completion notifies to the AMM that the UE has been moved to the desired iSC.
6. Finally, the AMM notifies to the RAC module that the requested UE has been relocated.

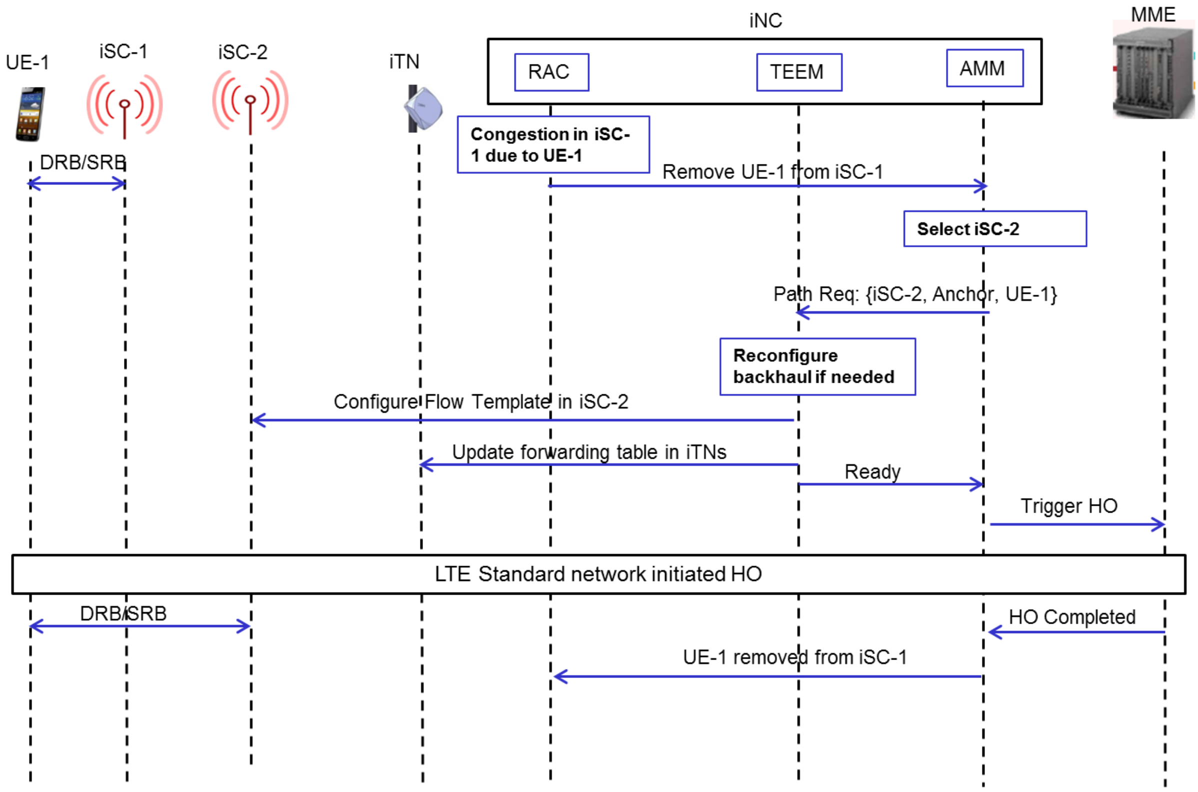

Figure 5-5 Interaction during Congestion Management (UE relocation needed)

5.3.4 Energy Optimisation Procedures

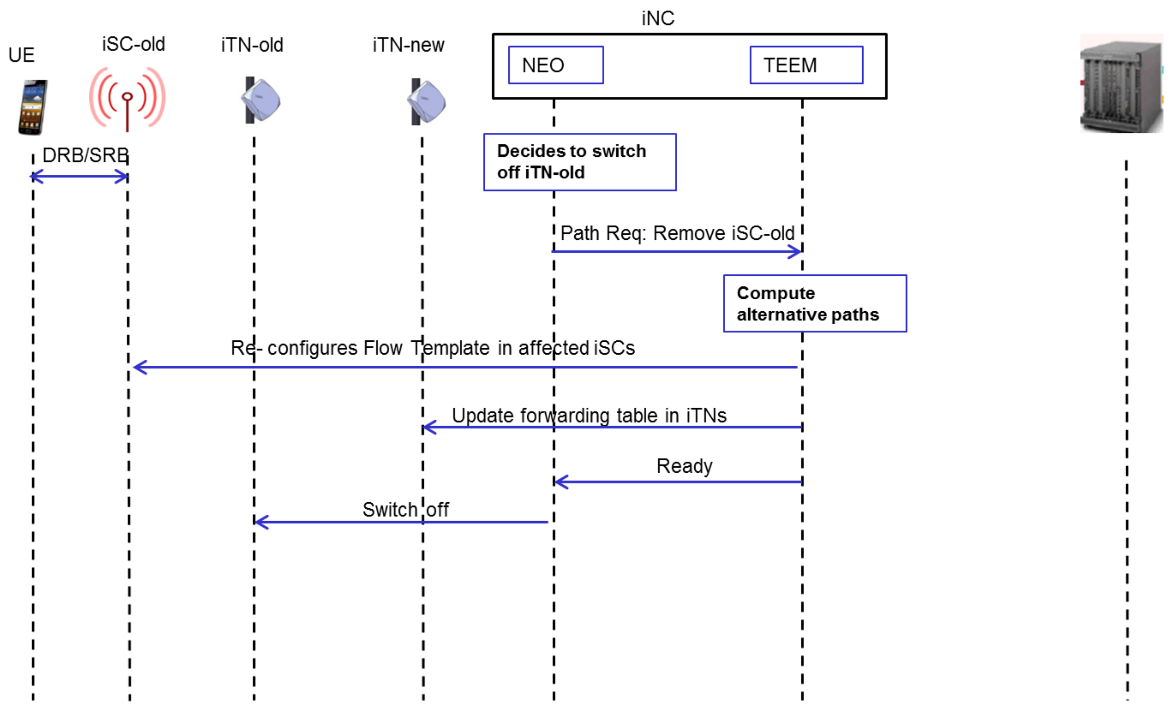

Figure 5-6 Interaction during Energy Optimisation implying switch-off of iTN (no UE relocation)

We now describe procedures related to energy optimisation in the radio access and backhaul network. Similar to the Congestion Management procedures, we split the description of these procedures in two cases, depending on whether a UE relocation is required or not.

Figure 5-6 depicts the case where an iTN can be switched off without requiring a UE relocation. The envisioned sequence of operations is the following:

1. The NEO module monitors the network wide energy consumption and decides to switch off a certain iTN. For that purpose, the NEO module sends a request to the TEMM to remove backhaul flows being routed through the selected iTN.

2. The TEEM performs the required computations and after identifying a suitable solution, it reconfigures the affected network elements and sends an acknowledgement back to the NEO module.

3. Finally, the NEO module sends a command to the iSC to power off.

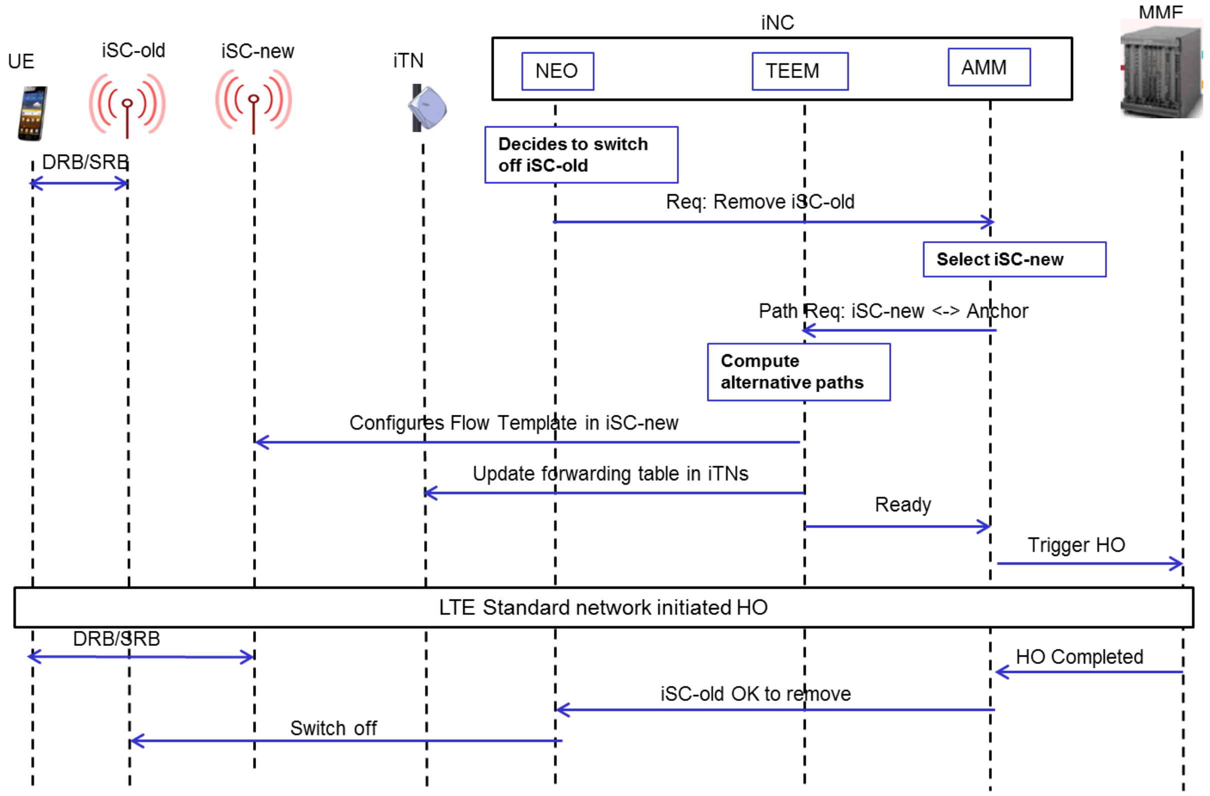

Figure 5-7 Interaction during Energy Optimisation implying switch-off of iSC (and UE relocation)

The second energy optimisation procedure is depicted in Figure 5-7 and corresponds to the case where a UE relocation is required. The envisioned detailed steps are the following:

1. The NEO module decides that a given iSC needs to be switched off. Therefore, the NEO module sends a request to the AMM in order to relocate the UEs attached to the selected iSC.

2. The AMM selects new iSCs for the affected UEs and sends a request to the TEEM to check if the required resources can be accommodated in the backhaul.

3. The TEEM determines whether the backhaul can accommodate the required resources and reconfigures the affected iSCs and iTNs.

4. Once the backhaul resources have been prepared, the AMM requests the MME to trigger HOs in order to relocate the affected UEs.

5. Finally, after the UEs have been relocated, the AMM sends an acknowledgement back to the NEO module which then proceeds with switching off the corresponding iSC.

6 Additional Considerations on the Use of Software Defined Networking

In this section we perform a first assessment of the impact of using SDN technologies in the iJOIN architecture, as well as explore the associated benefits. We also identify the SDN protocols that are currently foreseen to be adopted by WP4 CTs and the potential needs to extend them.

6.1 Impact of the Use of SDN

We first assess the impact of an SDN architecture on the RANaaS concept, by focusing on two main characteristics of RANaaS that are enabled by SDN:

- The support of functional mobility between network elements with different degrees of centralisation and distribution depending on the scenario to be supported.
- The ability to support over the same infrastructure backhauling and fronthauling requirements associated with the Cloud RAN concept.

The support of functional mobility by an SDN architecture can be considered from two different viewpoints: the mobility of backhaul functionality and the mobility of the functions that use the backhaul infrastructure. With respect to the first viewpoint, SDN may be used to support the distribution of the following backhaul functionalities:

- Transport services: MPLS, MPLS-TP, VLAN, IP, and related protocols.
- Routing functionality for network sharing and reliability.
- Basic functionalities such as security (IPsec, MACSec 802.1AE, IDP/IPS), physical layer synchronization (frequency, phase), and reliability.
- Added value functionalities such as caching, transcoding, traffic engineering, and data collection.

With respect to the mobility of functions, SDN may be used to locate different processing elements in different nodes. It is possible to distinguish

- Data processing, i.e., baseband processing, scheduling.
- Control processing, i.e., mobility support, load balancing.
- Non-Access Stratum (NAS) signalling processing.

SDN may also be an enabler for supporting different functional splits which imply different infrastructure requirements. This includes conventional X2/S1 information flows as well as quantised IQ signals associated to central baseband processing. The differences in terms of required bandwidth, latency, and reliability make it difficult to define a single technical solution which is able to cope with all requirements. SDN allows for an advanced network design and operation allow for this functional split, e.g. adaptive congestion control and traffic engineering.

CTs, which are presented in WP4, make use of an SDN approach, which allows for a logically centralised monitoring, configuration and control of the different involved entities. In an ideal network, logical centralisation provides clear advantages as it allows for an optimal solution, i.e., the controlling node has all information about the network and it can predict the impact that a configuration action would have on the network before actually taking this action. However, networks are far from being ideal and therefore the requirements imposed by using an SDN approach as well as the associated drawbacks need to be carefully analysed. This feasibility analysis will be an important aspect during the design of the different CTs. To provide an example of the differences and trade-offs between using a non-SDN approach and an SDN approach, we detail downlink common scheduling functionality in the following.

Figure 6-1 illustrates the downlink scheduling process in LTE. The downlink scheduler controls the assignment of user terminals to timeslots and resource blocks of the Downlink Shared Channel (DL-SCH). In addition, it controls the transport-format selection, i.e. selection of transport-block size, modulation scheme, and antenna mapping, and it controls the logical-channel multiplexing for downlink transmissions.

As a consequence, the RLC segmentation and MAC multiplexing will be affected by the scheduling decisions.

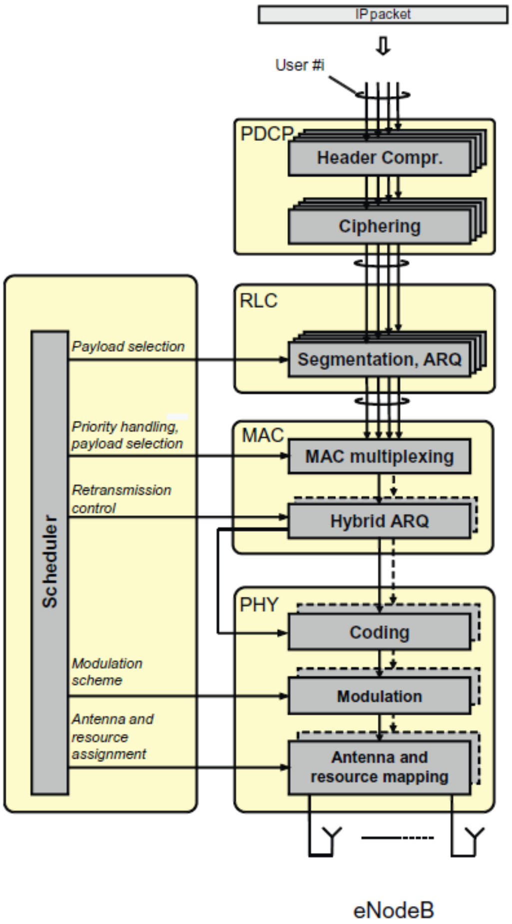

Figure 6-1 Downlink scheduling process in LTE

Figure 6-2 illustrates common scheduling, which allows for the selection of resources in different base stations. Its objective is to minimize inter-cell interference and maintain the scheduling gain in each cell. One implementation of a common scheduler assigns a dedicated base station which controls the scheduling processes of its own and other cells. Another possibility is a decentralised implementation across all involved base stations which exchange the corresponding scheduling information in order to achieve a consistent decision.

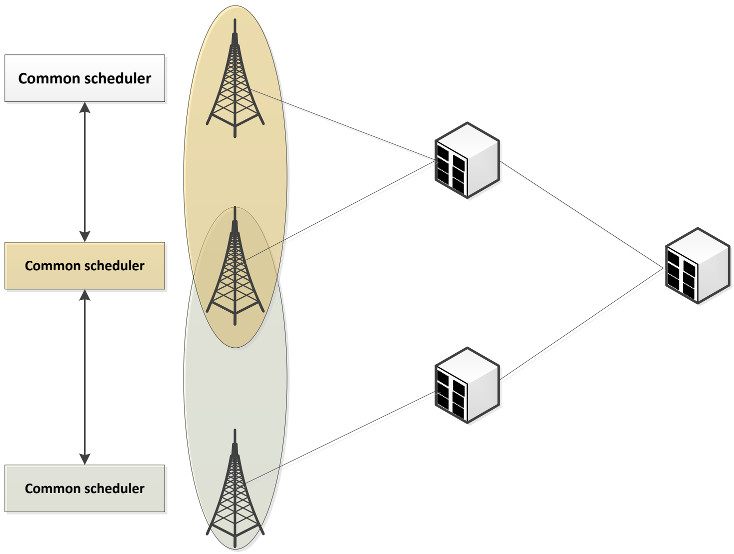

Figure 6-2 Common scheduling across base stations

Figure 6-3 illustrates a third option. It involves an external node that is connected to the cooperating base stations. This node can perform the common scheduling process. It receives IP packets from the core network and sends transport blocks to the cells with an indication of assigned timeslot and resource block. This division should be performed such that the scheduler is able to communicate directly with the cooperating cells.

This option would require an implementation of the common scheduling functionality in several nodes of the backhaul network. Hence, it implies some potential drawbacks, e.g. an increased complexity of the nodes, the need for coordination in order to select the most adequate node which supports the scheduling function[3], impact of the introduction of new base stations and changes in the network topology.

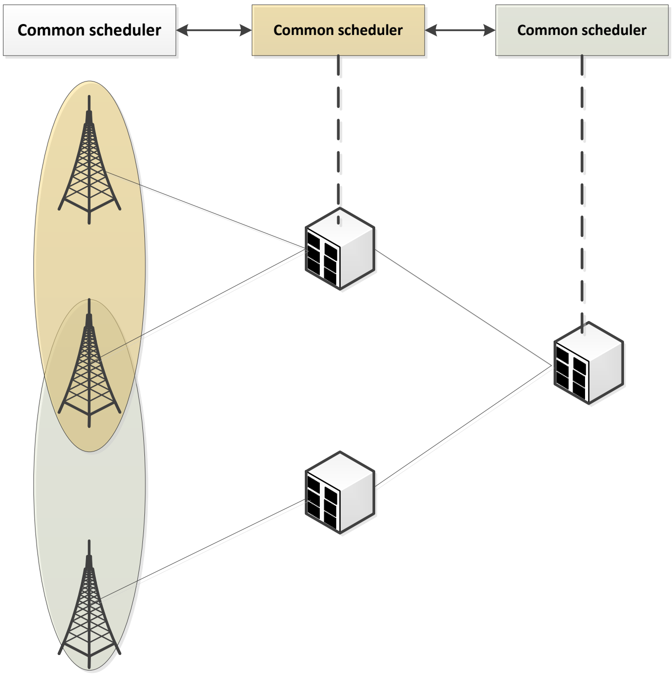

Figure 6-3 Scheduling through an external node

Figure 6-4 illustrates one potential way to address these challenges. The illustrated solution is the implementation of a SDN-like architecture for the backhaul network with a centralised control that supports the common scheduling functionality, alongside with other functionality which is more related to backhaul such as explained earlier.

The common scheduler should reside in a node that is accessible to all nodes in the backhaul network. It would have two basic functions:

- Selection of the appropriate modulation and coding scheme, resource elements, and antennas based on CSI information reported by the cooperating cells and the buffer status.
- Repackaging the downlink IP packets into transport blocks consistent with the scheduling decision.

One of the most critical challenges is the potential imperfectness of CSI due to delays. The impact of this imperfectness will depend on the UE mobility and interference characteristics. However, under normal operating conditions this delay should not exceed a few milliseconds.

[3] This would require implementing a signalling mechanism that right now is not contemplated in the standards.

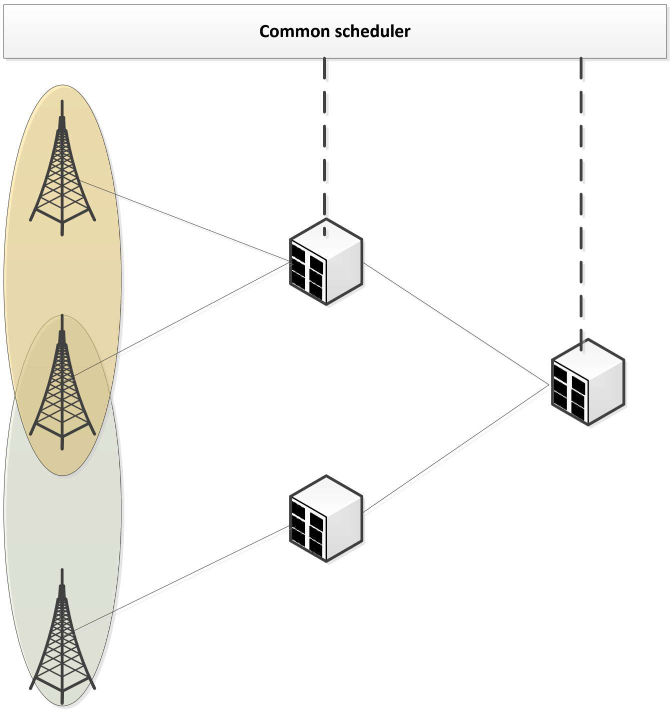

Figure 6-4 SDN architecture for common scheduling

6.2 SDN Protocols

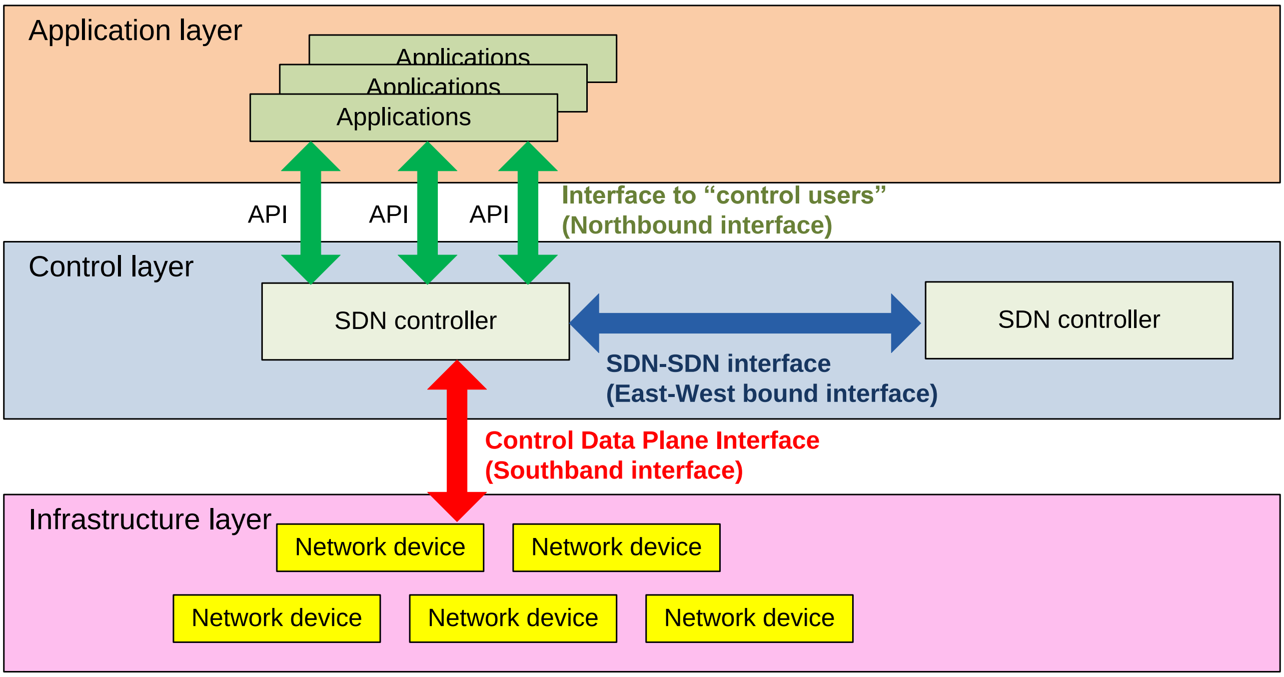

Figure 6-5 SDN protocols architecture

Figure 6-5 illustrates the SDN protocol architecture with particular focus on the interfaces. The most important interface is the so-called "southbound" interface which allows the centralised controller to program the network devices. Currently, the most widely adopted protocol for this interface is OpenFlow, other protocols include for instance ForCES. In iJOIN, we have adopted OpenFlow as base protocol, which will be extended based on the needs of the CTs and in order to fulfil our requirements.

The configuration of network devices may be done by using OpenFlow-CONFIG which is a protocol related to OpenFlow. This protocol still misses some important functionality, such that iJOIN needs to develop new extensions or to propose reusing other existing protocols.

For the interface between the SDN controller and its different users (the so-called "northbound" interface), specific iJOIN interfaces will be designed between the different modules (as shown in Figure 5-1).

The interface between SDN controllers is a matter of very active research (the so-called "East-West bound" interface). It is not yet clear whether iJOIN would require such an interface which would imply the existence of multiple iNCs that need to be coordinated. This will be subject of further study and explained in further detail in the next deliverable. If such an interface is deemed necessary, proposals from literature will be analysed regarding their applicability.

In summary, iJOIN aims to re-use existing SDN protocols and use ONF as baseline for future extensions. iJOIN aims at contributing those iJOIN extensions to the relevant SDOs, e.g. ONF and IETF.

7 Summary

In this deliverable, the state-of-the-art of the main topics in the joint RAN and backhaul network design of iJOIN in the network layer perspective is presented. The SDN architecture, which serves as the common platform for WP4 candidate technologies, is firstly introduced in the state-of-the-art section. These topics to be explored in the candidate technologies are also presented. This includes literature on various network layer topics, including mobility management, energy optimisation, congestion control, load balancing and also a backhaul network analysis technique based on stochastic geometry. The design of WP4 candidate technologies aims at improving the performance based on those literatures.

Based on this state-of-the-art, several candidate technologies for the joint radio access and the backhaul network design are also presented. The candidate technologies generally follow the SDN-based architecture, in which iNC plays a key role in managing the network based on the intelligence gained from the global network knowledge. Those candidate technologies aim to improve the network performance in terms of mobility, energy consumption, routing/scheduling and congestion control. There is also one candidate technology exploring the stochastic analysis of backhaul networks. The assumptions and requirements of each candidate technology are also explicitly mentioned and the detailed design is presented after that in each section.

Finally, the WP4 functional architecture is presented. This includes details on the modules running on iNC and the interactions between modules are also specified. Additional considerations on using SDN protocols in WP4 scope are also analysed.

Acknowledgements and Disclaimer

This work was partially funded by the European Commission within the 7th Framework Program in the context of the ICT project iJOIN (Grant Agreement No. 317941).

References

[1] Jain Sushant et al., "B4: Experience with a globally-deployed software defined WAN", Proceedings of the ACM SIGCOMM conference on SIGCOMM, IEEE, 2013.

[2] iJOIN Project, "D5.1 - Revised definition of requirements and preliminary definition of the iJOIN architecture, " October 2013.

[3] C. J. Bernardos, A. De Domenico, J. Ortin, P. Rost und D. Wübben, "Challenges of designing jointly the backhaul and radio access network in a cloud-based mobile network", ICT Future Networks and Mobile Summit (FuNeMS 2013), Lisboa, July 2013.

[4] F. Giust, A. de la Oliva and C. J. Bernardos, "Analytic Evaluation and Experimental Validation of a Network-based IPv6 Distributed Mobility Management Solution", submitted to IEEE Transactions on Mobile Computing, September 2013.

[5] H. Chan et al., "Requirements for Distributed Mobility Management", IETF Internet Draft: draft-ietf-dmm-requirements-03 (work-in-progress), December 2012.

[6] D. Liu et al., "Distributed Mobility Management: Current practices and gap analysis", IETF Internet Draft: draft-ietf-dmm-best-practices-gap-analysis-01 (work-in-progress), June 2013.

[7] D. Kutscher, R. Winter, S. Krishnan, Y. Zhang, C. J. Bernardos, "Mobile Communication Congestion Exposure Scenario", IETF Internet Draft : draft-ietf-conex-mobile-02, (WG document), July 2013.

[8] S. Sezer, S. Scott-Hayward, P.K. Chouhan, B. Fraser, D. Lake, J. Finnegan, N. Viljoen, M. Miller and N. Rao, "Are we ready for SDN? Implementation challenges for software-defined networks", Communications Magazine, IEEE, vol.51, no.7, pp. 36-43, July 2013.

[9] X. Costa-Pérez, A. Festag, H.J. Kolbe, J. Quittek, S. Schmid, M. Stiemerling and H. van der Veen, "Latest trends in telecommunication standards", ACM SIGCOMM Computer Communication Review, vol.43, pp. 64-71, April 2013.

[10] NEC, "Mobile backhaul evolution and convergence, White Paper", January 2010.

[11] K. Pentikousis, Yan Wang and Weihua Hu, "Mobileflow: Toward software-defined mobile networks", Communications Magazine, IEEE, vol.51, no.7, pp. 44-53, July 2013.

[12] G. Hampel, M. Steiner and Bu Tian, "Applying Software-Defined Networking to the telecom domain", IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp.133-138, April 2013.

[13] J. Kempf, B. Johansson, S. Pettersson, H. Luning and T. Nilsson, "Moving the mobile Evolved Packet Core to the cloud", IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp.784-791, October 2012.

[14] Kunz, S. Palanimuthu, J. Song and H.Kolbe, "Introducing OpenFlow in the 3GPP Evolved Packet Core", IEEE COMSOC MMTC E-Letter, vol.8, no.40-42, May 2013.

[15] X. Costa-Perez, J. Swetina, Tao Guo, R. Mahindra and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks", Communications Magazine, IEEE, vol.51, no.7, pp. 27-35, July 2013.

[16] P. Dely, A. Kassler and N. Bayer, "OpenFlow for Wireless Mesh Networks", Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN), pp. 1-6, 31 July 2011 - 4 August 2011.

[17] NGMN, "A requirement specification by the NGMN alliance, NGMN optimised backhaul requirements", August 2008.

[18] Chiosi et al., "Network Functions Virtualisation", SDN and OpenFlow World Congress, Darmstadt, Germany, October 2012.

[19] C. Perkins, "IP Mobility Support for IPv4", RFC 5944, IETF, November 2010.

[20] D. Johnson, C. Perkins and J. Arkko, "Mobility Support in IPv6", RFC 3775, IETF, June 2004.

[21] 3GPP TS 29.274, "Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C)", September 2011.

[22] S. Gundavelli, K. Leung, V. Devarpalli, K. Chowdhury and B. Patil, "Proxy Mobile IPv6", RFC 5213 (Proposed Standard), IETF, August 2008. [Online] Available: http://www.ietf.org/rfc/rfc5213.txt.

[23] H. Soliman, "Mobile IPv6 Support for Dual Stack Hosts and Routers", RFC 5555 (Proposed Standard), IETF, June 2009. [Online] Available: http://www.ietf.org/rfc/rfc5555.txt.

[24] H. Chan, "Problem statement for distributed and dynamic mobility management", IETF Internet Draft, March 2011. [Online] Available: http://tools.ietf.org/html/draft-chan-distributed-mobility-ps-01.

[25] C. J. Bernardos, A. de la Oliva, F. Giust, "A IPv6 Distributed Client Mobility Management approach using existing mechanisms", IETF Internet Draft: draft-bernardos-dmm-cmip-00 (work-in-progress), July 2013.

[26] C. J. Bernardos, A. de la Oliva, F. Giust, T. Melia, R. Costa, "A PMIPv6-based solution for Distributed Mobility Management", IETF Internet Draft: draft-bernardos-dmm-pmip-02 (work-in-progress), July 2013.

[27] P. Seite, P. Bertin, J. H. Lee, "Dynamic Mobility Anchoring", IETF Internet Draft: draft-seite-dmm-dma-06.txt (work-in-progress), January 2013.

[28] H. Chan, F. Xia, J. Xiang and H. Ahmed, "Distributed Local Mobility Anchors", IETF Internet Draft: draft-chan-netext-distributed-lma-03 (work-in-progress), March 2010.

[29] J. C. Zúñiga, C. J. Bernardos, T. Melia, A. de la Oliva, R. Costa, A. Reznik, "Distributed Mobility Management: a Standards Landscape", accepted for publication in IEEE Communications Magazine, issue on Telecommunications Standards, March 2013.

[30] F. Giust, C. Bernardos, S. Figueiredo, P. Neves and T. Melia, "A Hybrid MIPv6 and PMIPv6 Distributed Mobility Management: the MEDIEVAL approach", IEEE Workshop on multiMedia applications over Wireless Networks (MediaWiN), organised in association with the IEEE Symposium on Computers and Communications (ISCC 2011), June 2011.

[31] F. Giust, A. de la Oliva and C. Bernardos, "Flat Access and Mobility Architecture: an IPv6 Distributed Client Mobility Management Solution", 3rd IEEE International Workshop on Mobility Management in the Networks of the Future World (Mobiworld 2011), in conjunction with the 30th IEEE International Conference on Computer Communications (IEEE INFOCOM 2011), April 2011.

[32] F. Giust, A. de la Oliva, C. J. Bernardos and R. P. Ferreira Da Costa, "A Network-based Localized Mobility Solution for Distributed Mobility Management", International Workshop on Mobility Management for Flat Networks (MMFN), collocated with WPMC 2011, October 2011.

[33] P. Bertin, S. Bonjour and J.-M. Bonnin, "A Distributed Dynamic Mobility Management Scheme Designed for Flat IP Architectures", New Technologies, Mobility and Security (NTMS), November 2008.

[34] P. Bertin, S. Bonjour and J.-M. Bonnin, "An Evaluation of Dynamic Mobility Anchoring", IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), September 2009.

[35] M. Fischer, F.-U. Andersen, A. Kopsel, G. Schafer and M. Schlager, "A Distributed IP Mobility Approach for 3G SAE", 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), IEEE, September 2008.

[36] L. Yu, Z. Zhijun, L. Tao and T. Hui, "Distributed mobility management based on flat network architecture", 5th Annual ICST Wireless Internet Conference (WICON), March 2010.

[37] P. McCann, "Design of a flat wireless Internet Service Provider network", 14th International Symposium on Wireless Personal Multimedia Communications (WPMC), pp.1-5, October 2011.

[38] R. Wakikawa et al., "Separation of Control and User Plane for Proxy Mobile IPv6", IETF Internet Draft: draft-wakikawa-netext-pmip-cp-up-separation-00.txt (work-in-progress), July 2013. [Online] Available: http://tools.ietf.org/html/draft-wakikawa-netext-pmip-cp-up-separation-00.

[39] S. Matsushima et al., "Separation Stateless user-plane architecture for virtualized EPC (vEPC)", IETF Internet Draft: draft-matsushima-stateless-uplane-vepc-01.txt (work-in-progress), July 2013.

[40] IEEE standard, "IEEE P802.21/D14 Draft Standard for Local and Metropolitan Area Networks: Media Independent Handover Services", IEEE, 2008.

[41] 3GPP, TS 36.300, v10.9.0, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 10) ", December 2012.

[42] 3GPP, TS 25.304, v10.7.0, "Technical Specification Group Radio Access Network; User Equipment (UE) procedures in idle mode and procedures for cell reselection in connected mode (Release 10)", December 2012.

[43] 3GPP, TS 23.401, v12.2.0, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access", Release 12, September 2013.

[44] 3GPP, TR 23.859, "LIPA Mobility and SIPTO at the Local Network", July 2011.

[45] W. Hahn, "Flat 3GPP evolved packet core: Improvement for multiple network connections", International Symposium on Wireless Personal Multimedia Communications (WPMC), October 2011.

[46] W. Hahn, "3GPP Evolved Packet Core support for distributed mobility anchors: Control enhancements for GW relocation", International Conference on ITS Telecommunications (ITST), August 2011.

[47] C. J. Bernardos, J. C. Zúñiga, A. Reznik, "Towards flat and distributed mobility management: A 3GPP evolved network design", IEEE International Conference on Communications (ICC), pp.6855-6861, June 2012.

[48] L. Wang, Y. Zhang, and Z. Wei, "Mobility Management Schemes at Radio Network Layer for LTE Femtocells", IEEE 69th Vehicular Technology Conference (VTC Spring), pp. 1-5, April 2009.

[49] A. Ulvan, R. Bestak, and M. Ulvan, "The Study of Handover Procedure in LTE-based Femtocell Network", Wireless and Mobile Networking Conference (WMNC), Third Joint IFIP, pp. 1-6, October 2010.

[50] A. Ulvan, R. Bestak, and M. Ulvan, "Handover Scenario and Procedure in LTE-based Femtocell Networks", the 4th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM), October 2010.

[51] Z. Fan and Y. Sun, "Access and Handover Management for Femtocell Systems", the 71st Vehicular Technology Conference (VTC Spring), IEEE, pp. 1-5, May 2010.

[52] H. Li, X. Xu, D. Hu, X. Chen, X. Tao, and P. Zhang, "Pseudo-Handover based Power and Subchannel Adaptation for Two-tier Femtocell Networks", Wireless Communications and Networking Conference (WCNC), IEEE, pp. 980-985, March 2011.

[53] H. Zhang, X. Wen, B. Wang, W. Zheng, and Y. Sun, "A Novel Handover Mechanism Between Femtocell and Macrocell for LTE Based Networks", the 2nd International Conference on Communication Software and Networks (ICCSN), pp. 228-231, February 2010.

[54] Z. Chowdhury, W. Ryu, E. Rhee, and Y. Jang, "Handover between Macrocell and Femtocell for UMTS based Networks", the 11th International Conference on Advanced Communication Technology (ICACT), February 2009.

[55] M. Chowdhury and Y. Jang, "Handover Control for WCDMA Femtocell Networks", the Journal of Korea Information and Communication Society, vol. abs/1009.3779, September 2010.

[56] J. Kim and T. Lee, "Handover in UMTS Networks with Hybrid Access Femtocells", the 12th International Conference on Advanced Communication Technology (ICACT), vol. 1, pp. 904 –908, February 2010.

[57] S. Wu, "A New Handover Strategy between Femtocell and Macrocell for LTE-Based Network", the 4th International Conference on Ubi-Media Computing (U-Media), pp. 203 –208, July 2011.

[58] K. K. Yap, M. Kobayashi, D. Underhill, S. Seetharaman, P. Kazemian, and N. McKeown, "The stanford openroads deployment", In Proceedings of the 4th ACM international workshop on Experimental evaluation and characterization, pp. 59-66, September 2009.

[59] D. Liu et al., "Mobility Support in Software Defined Networking", IETF Internet Draft: draft-liu-sdn-mobility-00.txt (work-in-progress), July 2013.

[60] 3GPP, TR 32.826, v10.0.0, "Technical Specification Group Services and System Aspects; Telecommunication Management; Study on Energy Savings Management (ESM)", Release 10, March 2010.

[61] 3GPP, TS 32.551, v10.1.0, "Technical Specification Group Services and System Aspects; Telecommunication Management; Energy Savings Management (ESM); Concepts and Requirements", Release 10, March 2011.

[62] 3GPP, TR 36.927, v11.0.0, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Potential Solutions for Energy Saving for E-UTRAN", Release 11, September 2012.

[63] Deliverable 3.2, EARTH Project, EU FP 7, "Green Network Technologies", January 2012.

[64] ETSI, TS 102.706, Environmental Engineering (EE), "Energy Efficiency of Wireless Access Network Equipment", August 2009.

[65] ETSI, TS 102.530, Environmental Engineering (EE), "The Reduction of Energy Consumption in Telecommunication Equipment and Related Infrastructure", June 2008.

[66] J. Quittek, R. Winter, T. Dietz, B. Claise, M. Chandramouli, "Requirements for Energy Management", IETF Internet Draft, November 2011.

[67] ITU-T, Series G Sup. 45, "GPON Power Conservation", May 2009.

[68] IEEE 802.3az, "Energy Efficient Ethernet Task Force", 2010.

[69] Paolo Monti, Sibel Tombaz, Lena Wosinska and Jens Zander, "Mobile Backhaul in Heterogeneous Network Deployments: Technology Options and Power Consumption", the 14th International Conference on Transparent Optical Networks (ICTON), IEEE, July 2012.

[70] S. Tombaz, et al., "Impact of Backhauling Power Consumption on the Deployment of Heterogeneous Mobile Networks", Proceeding of IEEE GLOBECOM, December 2011.

[71] Z. Hasan, H. Boostanimehr, and V.K. Bhargava, "Green cellular networks: A survey, some research issues and challenges", Communications Surveys & Tutorials, IEEE, 2011.

[72] Tao Han, N. Ansari, "On greening cellular networks via multicell cooperation", Wireless Communications, vol. 20, issue 1, IEEE, February 2013.

[73] Tao Han, N.Ansari, "Optimizing Cell Size for Energy Saving in Cellular Networks with Hybrid Energy Supplies", IEEE GLOBECOM, December 2012.

[74] M. Sawahashi, "Coordinated Multipoint Transmission/Reception Techniques for LTE-Advanced (Coordinated and Distributed MIMO) ", IEEE Wireless Communications, June 2010.

[75] Yan Chen, Shuqing Zhang, Shugong Xu and G.Y.Li, "Fundamental Trade-offs on Green Wireless Networks", IEEE Communications Magazine, vol. 49, issue 6, June 2011.

[76] S. Floyd and V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance", IEEE ACM Transactions on Networking, August 1993.

[77] S. Athuraliya, S. Low, V. Li, and Q. Yin., "REM: Active Queue Management", IEEE Network, May 2001.

[78] 3GPP, TR 22.805, "Feasibility study on user plane congestion management (Release 12)", December 2012.

[79] NGMN, "Next Generation Mobile Networks Use Cases related to Self Organising Network, Overall Description", Deliverable, December 2008.

[80] F. Baccelli and B. Blaszczyszyn, "On a coverage process ranging from the boolean model to the Poisson-Voronoi tessellation with applications to wireless communications", Advances in Applied Probability, vol. 33, no. 2, pp. 293-323, 2001.

[81] F. Baccelli and B. Blaszczyszyn, "Stochastic Geometry and Wireless Networks Volume 1: THEORY", Foundations and Trends in Networking, vol. 3, pp. 249-449, 2009.

[82] V. Suryaprakash, A. F. dos Santos, A. Fehske, and G. P. Fettweis, "Energy consumption analysis of wireless networks using stochastic deployment models", GLOBECOM, IEEE, 2012.

[83] V. Suryaprakash, A. Fehske, A. F. dos Santos, and G. P. Fettweis, "On the impact of sleep modes and bw variation on the energy consumption of radio access networks", VTC-Spring, IEEE, 2012.

[84] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular Networks", CoRR, vol. abs/1009.0516, 2010.

[85] H. Dhillon, R. Ganti, F. Baccelli, and J. Andrews, "Modeling and analysis of k-tier downlink heterogeneous cellular networks", IEEE Journal on Selected Areas in Communications, vol. 30, no. 3, pp. 550–560, 2012.

[86] V. Suryaprakash and G. Fettweis, "A stochastic examination of the interference in heterogeneous radio access networks", Proceedings of Wireless Optimization Conference (WiOpt), IEEE, May 2013.

[87] A. Sharma, A. Roy, S. Ghosal and R.Chaki, "Load Balancing in Cellular Network: A Review", the 3rd International Conference on Computing Communication & Networking Technologies (ICCCNT), July 2012.

[88] E. Yanmaz, S. Mishra, O. K. Tonguz, H. Wu and C. Qiao, "Efficient dynamic load balancing algorithms using ICAR systems: A generalized framework ", IEEE Vehicular Technology Conf. (VTC Fall), vol. 1, pp.586-590, 2002.

[89] L. Du, J. Bigham and Laurie Cuthbert, "Towards Intelligent Geographic Load Balancing for Mobile Cellular Networks", IEEE Trans. on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol. 33, issue 4, pp. 480-491, November 2003.

[90] S. Kim and P.K. Varshney, "Adaptive Load Balancing with Preemption for Multimedia Cellular Networks", IEEE WCNC, pp. 1680-1684, 2003.

[91] Lexi Xu, Yue Chen and Yue Gao, "Self-organizing Load Balancing for Relay Based Cellular Networks", the 10th IEEE International Conference on Computer and Information Technology (CIT), pp. 791-796, 2010.

[92] A. Awada, B. Wegmann, I. Viering and A. Klein, "A game-theoretic approach to load balancing in cellular radio networks", IEEE International Symposium on Personal, Indoor and Mobile Radio Communications Conference, September 2010.

[93] Kolios P, Friderikos V, Papadaki KP, "Load balancing via store-carry and forward relaying in cellular networks", IEEE global telecommunications conference, 2010.

[94] H. Lee and C. Lee, "An Integrated Multi-hop Cellular Data Network", the 58th Vehicular Technology Conference (VTC-Fall), IEEE, 2003.

[95] B. Fortz and M. Thorup, "Optimizing OSPF/IS-IS Weights in a Changing World", IEEE Journal on Selected Areas of Communication, vol. 20, no 4, May 2002.

[96] A. Oliveira and G.R. Mateus, "Using Genetic Algorithms to LSP Setup in MPLS Networks", Proc. XXIV Simpósio Brasileiro de Redes de Computadores, pp. 705-721, 2006.

[97] G. Conte, P. Iovanna, R. Sabella, M. Settembre and L. Valentini, "A Traffic Engineering Solution for GMPLS Network: A Hybrid Approach Based on off-line and on-line routing methods", ONDM Conference, February 2003.

[98] M. Zhang, B. Liu and B. Zhang, "Multi-Commodity Flow Traffic Engineering with Hybrid MPLS/OSPF Routing", IEEE Globecom, December 2009.

[99] Tao Guo, Atta UI Quddus, Ning Wang and Rahim Rafazolli, "Local Mobility Management for Networked Femtocells Based on X2 Traffic Forwarding", IEEE Transaction on Vehicular Technology, vol.2, Issue 1, pp. 326-340, August 2012.

[100] Zubin Bharucha, et al., "Small Cell Deployments: Recent Advances and Research Challenges", the 5th International Workshop on Femtocells, February 2012.

[101] Dimitri Staessens, et al., "Software Defined Networking: Meeting Carrier Grade Requirements", the 18th IEEE workshop on Local & Metropolitan Area Networks (LANMAN), October 2011.

[102] 3GPP, TR 36.927 v11.0.0, "Potential Solution for Energy Saving for E-UTRAN", Release 11, September 2012.

[103] Marco Ajmone Marsan, Luca Chiaraviglio, Delia Ciullo and Michela Meo, "On the effectiveness of single and multi base station sleep modes in cellular networks", INRIA, Research report, RR-8186, November 2012.

[104] J. Wells, "Multi-gigabit wireless technology at 70GHz, 80GHz and 90GHz", RF Design, pp. 50-58, May 2006.

[105] 3GPP, TS 36.300, v11.7.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2", Release 11, September 2013.

[106] 3GPP, TS 23.203, v12.2.0, "Technical Specification Group Services and System Aspects; Policy and charging control architecture", Release 12, September 2013.

[107] IEEE, "IEEE Standard for Local and Metropolitan Areas", May 2006.

[108] Xin Jin, Li Erran Li, Laurent Vanbever and Jennifer Rexford, "SoftCell: Scalable and Flexible Cellular Core Network Architecture", Princeton University and Bell Labs, 2012.

[109] Mor Harchol-Balter, Alan Scheller-Wolf and Andrew Young, "Surprising results on task assignment for high variability workloads", Stanford University, 2009.

[110] L.C. Schmelz, J.L. Van Den Berg, K.Zetterberg, M.Amirijoo, K. Spaey, I.Balan, N. Scully and S.Stefanski, "Self-organization in Wireless Networks - Use Cases and their Interrelation", Wireless Word Research Forum, May 2009.

[111] ONF, "OpenFlow Configuration and Management Protocol OF-CONFIG 1.0", June 2012.

