TECHNISCHE UNIVERSITÄT BERLIN

Faculty IV - Electrical Engineering and Computer Science
Department of Telecommunication Systems
Chair for Next Generation Networks (AV)

# Self-Adaptable IP Control in Carrier Grade Mobile Operator Networks

## - Engineering Doctorate Dissertation -

Author: Dipl.-Eng. Marius-Iulian Corici

Scientifically Defended in Berlin, on the 27th of November, 2013
Berlin, 2014
D 83

R&D in cooperation with Fraunhofer Institute for Open Communication Systems

TECHNISCHE UNIVERSITÄT BERLIN

# Self-Adaptable IP Control in Carrier Grade Mobile Operator Networks

vorgelegt von
Dipl.-Ing.
Marius Corici
aus Bukarest/Rumänien

von der Fakultät IV - Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften
- Dr.-Ing. -

genehmigte Dissertation

Promotionsausschuss:

| | |
|---|---|
| *Vorsitzender :* | Prof. Dr. Sebastian Möller |
| *Gutachter :* | Prof. Dr. Thomas Magedanz |
| *Gutachter :* | Prof. Dr. Axel Küpper |
| *Gutachter :* | Prof. Dr. Hans Schotten |

Tag der wissenschaftlichen Aussprache: 27.11.2013

Berlin, 2014
D 83

This work represents my engineering doctorate dissertation, defended on the 27th of November 2013 in Berlin.
The academic requirements have been fulfilled at the Technische Universität Berlin, within Faculty IV - Elektrotechnik und Informatik - Chair for Next Generation Networks.
The practical R&D has been performed in cooperation with the Fraunhofer Institute for Open Communication Systems (FOKUS). The resulting source code copyrights belong in large parts to Fraunhofer FOKUS.

www.tu-berlin.de
www.av.tu-berlin.de
coricim@gmail.com

www.fokus.fraunhofer.de/go/ngni
www.openepc.net
marius-iulian.corici@fokus.fraunhofer.de

---

**Impressum**

Druck und Verlag: epubli GmbH, Berlin, www.epubli.de
ISBN 978-3-8442-9462-0
Zugl.: Berlin, Technische Universität, Diss., 2014

## Acknowledgments

Research is foremost about community and writing a dissertation is a truly community development act. Many people helped in the creation of this dissertation.

First, I would like to thank to Prof. Dr. Thomas Magedanz who has guided and encouraged me as well as acting as a role model for my entire research career at Fraunhofer FOKUS. I am extremely grateful for his guidance and support throughout these years. Likewise, I am indebted to Prof. Dr. Hans Schotten and Prof. Dr. Axel Kuepper for their valuable suggestions and for reviewing my research work.

Furthermore, I have to thank to the people that through brief encounters or through long time exposure made me who I am, my models, my mentors: Prof. Ion Savu, Prof. Dr. Nicolae Tapus, Prof. Dr. Valentin Cristea, Dr. Dorgham Sisalem, Prof Dr. Radu Popescu-Zeletin, Dr. John Williams Floroiu, Dr. Stefan Arbanowski, Dr. Mikhail Smirnov, Dr. Stephan Steglich, and Hans Einsiedler.

I thank to all my colleagues at Fraunhofer FOKUS Institute for their technical support through the research projects, especially to Alin Murarasu, Mihai Capota, Elena Apostol, Catalin Iacob, Diana Potolog, Alberto Diez, Julius Mueller, Ancuta Onofrei, Jordi Jaen Pallares, Frank Schulze, Jakub Kokur, Marc Emmelman, Lukas Wollner, Mihai Constantin, Dana Satryia, Benjamin Reichel, Oliver Keil, Rudolf Roth, Giuseppe Carella and to the external project partners Kees den Hollander, Suwon Lee, Cornel Pampu, Zhou Qing, Mirko Schramm, Nico Bayer, and Ivano Guardini. I could not have done it without you!

I am very grateful to my friends who supported me throughout the long duration of developing myself and of my research career Cristian Zamfir, Andrei Aron, Andrei Damian, Valeriu Ohan, Vama Veche, Tudor Golubenco, Razvan Petre, Monica Sarbu, Cristian Constantin, Cosmin Mogos, Bogdan Harjoc, Bogdan Pintea, Oliver Friedrich, Nikolai Tscholciev, and Jacob Schramm.

My long lasting thanks go to Dragos Vingarzan and to Jens Fiedler who always pushed me forward to places where my courage did not dare to let me go. I would have not accomplished half without your advice and your day-to-day guidance.

The community starts with the family. My parents and my hole family were very supportive in pressuring me to keep the path towards my goal, especially in the last years of development and change. For this I will always be in dept to you.

My special gratitude goes to my wife, Ancuta, for her patience and loving in the many hours I have dedicated to this dissertation. Thank you for being so determined and resolute. Last and also most important, I have to thank to my son Andrei Ilarion who decided by himself from the beginning to be a very happy little man, my life, my light.

I thank God this work is completed.

**Abstract:** The current trend in wireless operator networks is towards deploying in parallel a large number of heterogeneous access network technologies and core network functions in order to reach a higher capacity and thus accommodating the communication needs of a high number of connected devices. However, the wireless operator network is designed for handling uniformly all subscribers from the perspective of mobility management, subscription information, resource reservation and charging. The current architecture corroborated with a high increase of connected devices requires a new level of scalability to be attained which is not realizable.

This dissertation provides a framework, named SelfFit, for subscriber oriented automatic adaptation of connectivity from core network perspective for access network selection, efficient usage of the core network resources and the data path adaptation, addressing the previous scalability challenge. In these three directions, a set of five innovative self-adaptation concepts for dynamic carrier grade network infrastructures was developed. The concepts were immediately exemplified as additions to the current 3GPP Evolved Packet Core (EPC) architecture, for proving their value in a standard, real-life deployed architecture. Additionally, the functionality was prototyped on top of the Fraunhofer FOKUS OpenEPC toolkit and evaluated on complete realistic testbed setups, providing a practical view on how it can be realized as part of future core network products.

For assessing the limitations of existing technologies as well as of the proposed solutions, a rigorous generic evaluation framework was developed including the requirements definition, a simplified network model, qualitative metrics for assessing heterogeneous technologies, and comparative qualitative testbed measurements. The evaluation framework enables the reader to grasp a clear evidence of the innovation value included in this dissertation from innovative idea to pre-product prototyping.

**Zusammenfassung:** Der aktuelle Trend bei Mobilfunknetzen besteht im parallelen Einsatz einer großen Zahl von heterogenen Zugangsnetztechniken und Kernnetzfunktionen, um eine höhere Kapazität zu erreichen und so den Kommunikationsbedürfnissen einer großen Zahl von angeschlossenen Endgeräten gerecht zu werden. Allerdings ist das Mobilfunknetz darauf ausgerichtet, alle Teilnehmer in Bezug auf Mobilitätsbetreuung, Teilnehmerinformationen, Ressourcenzuteilung, sowie Rechnungslegung gleich zu behandeln. Die aktuelle, durch den großen Anstieg der angeschlossenen Geräte bestärkte Architektur, erfordert das Erreichen eines neuen Maßes an Skalierbarkeit, das nicht realisierbar ist.

Die vorliegende Dissertation stellt einen Rahmen Namens SelfFit bereit, für eine teilnehmerorientierte, automatische Anpassung der Konnektivität aus der Perspektive des Kernnetzwerks zur Zugangsnetzauswahl, zur effizienten Nutzung der Kernnetzressourcen und der Datenpfadanpassung zur Verfügung, und geht damit auf das bestehende Skalierbarkeitsproblem ein. In diesen drei Richtungen wurde ein Satz von fünf innovativen Selbstanpassungskonzepten für dynamische Netzwerk-Infrastrukturen auf Betreiberniveau entwickelt. Die Konzepte wurden sofort beispielhaft als Ergänzung zur aktuellen 3GPP Evolved Packet Core (EPC) Architektur eingesetzt, um sich in einer normalen, praxisnahen Einsatzarchitektur zu bewähren. Zusätzlich wurde die Funktionalität prototypisch auf dem Fraunhofer FOKUS OpenEPC Toolkit aufgesetzt und auf vollständigen realistischen Prüfstandaufstellungen bewertet, wodurch sichtbar wurde, wie sie in der Praxis als Teil zukünftiger Kernnetzwerk-Produkte realisiert werden könnte.

Für die Beurteilung der Grenzen der bestehenden Technologien sowie der vorgeschlagenen Lösungen wurde ein strenger generischer Bewertungsrahmen entwickelt, einschließlich der Anforderungsdefinition, eines vereinfachten Netzwerkmodells, qualitativer Kennzahlen zur Beurteilung heterogener Technologien, sowie vergleichender qualitativer Prüfstandmessungen. Der Bewertungsrahmen liefert dem Leser einen klaren Beweis des in der vorliegenden Dissertation enthaltenen Innovationswerts, von der innovativen Idee bis zur Konstruktion des Vorprodukt-Prototyps.

# Contents

# List of Figures

# List of Tables

## Chapter 1

# Introduction

This chapter describes the motivation and the objectives of this thesis giving a general introduction into the context of the work.

## 1.1 Motivation

During the last two decades, the telecommunication technologies have reached a high level of acceptance, mainly due to the development of the Internet which allowed previously unimaginable levels of information exchange, [1] and due to the mobile communication which empowered the users with continuous reachability and communication [2]. Following this trend, the mobile communication industry is currently passing into a novel massive broadband communication age through the harmonization of previously voice centered services with the Internet technologies [3]. Five factors are of significance for this evolution.

First, novel radio technologies became available and are currently rolled out by carrier grade network operators like LTE and wide space WiFi. They come to supplement the already deployed technologies like HSPA, UMTS, CDMA and WiFi [4]. Although heterogeneous as capacity, delay, packet loss, and operational costs, the access networks altogether are offering wireless remote communication with an increased level of throughput, thus being able to accommodate larger levels of communication.

Secondly, the users accustomed with both the Internet and the mobile communication technologies are gradually adopting, as commodity, devices such as smart phones, tablets and laptops. The adoption of the heterogeneous devices corroborated with the data traffic increase enlarge the current subscriber based communication, while at the same time requiring an extended support from the core networks [5].

Following, due to the opportunity of delivering services anywhere, anytime, and to a large set of users, novel services and applications are foreseen especially targeting mobile subscribers. Added to this, the current applications used over fixed lines access networks are adapted to the mobile networks, offering a large variety of services to the mobile subscribers [6].

Additionally, the fast adoption of cloud based technologies [3], [7] enables the sharing of the data centers' computation and data storage capabilities, fast and on-demand service deployment, and through a wide range of cost efficient applications addressing mass markets and especially opening the opportunity for niche services.

Also, due to the massive deployment of access networks offering remote communication at reduced costs, other industries such as energy, automotive, and security are considering the usage of operator infrastructures for a novel type of mobile communication for which the human interaction is limited, generically named *Machine-2-Machine (M2M)*. For this, it is estimated an increase with one order of magnitude of the connected devices and a high diversification of their capabilities [8].

Finally, motivated by an increased awareness on environment issues and by the high urbanization and globalization [9], also sustained by the foreseen diversification in devices and services and the reduction in communication and service deployment costs, a novel Smart Cities initiative aims at fostering the energy efficiency, smart energy distribution, mobility and transportation through the support of novel remote communication mechanisms [10].

However, in order to be fully accepted, the massive broadband mobile communication core network infrastructure is expected to reach similar quality and reduced operational costs as the evolution to *Next Generation Networks (NGN)* brought to the fixed lines communication [11]. This presumes that the wireless network operators are able to use efficient carrier-grade architectures. As defined by the Webster Telecom Dictionary, the carrier grade property refers to *"hardware or software that is durable and reliable to satisfy the demands of a carrier, i.e. incumbent telephone company, competing telecommunication service providers, or Internet service provider (ISP)"* [12]. In the context of this thesis, a carrier grade operator network refers to the systems deployed by network providers addressing a large number of subscribers such as mobile network operators.

As the wireless networks are natively cellular, the devices located at a specific moment of time at a specific location compete for the same resources, thus requiring resource mediation. Also, in order to be able to maintain the reachability of mobile devices, the network has to be able to offer mobility support considering mobile device location changes. These two features of the mobile technology as well as the deployment limitations such as renting of antenna sites, make the mobile technology more expensive to use and requires separate architecture design [13].

As depicted in Figure 1.1, the massive broadband wireless environment is characterized by an increased acceptance of mobile devices and mobile application brought by the deployment of novel access networks and service platforms, as well as by a high variation in parameters and characteristics for these new deployed technologies. For sharing the mobile network infrastructure and for enabling service support when required and at reduced operational costs, a high level of scalability has to be reached, which can be attained only by developing a novel system architecture, based and evolved from the Next Generation Networks [11].

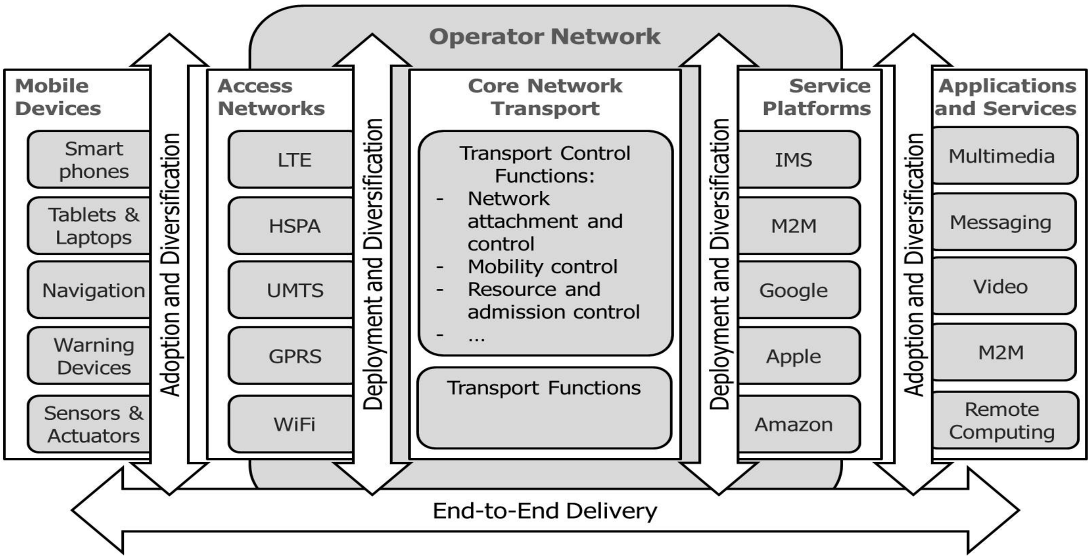


Figure 1.1: Mobile Network High Level Functional Split

As defined in [14], *scalability is "the ability of a system to accommodate an increasing number of elements or objects to process growing volumes of work gracefully, and/or to be susceptible to enlargement"*. Translated to the operator infrastructure, scalability represents the ability of the core network to accommodate an increasing number of subscribers and to handle their signalling and data traffic. Specifically, the scalability increase required for the operator core network architecture relates to the load scalability - handling a larger number of subscribers - through the usage of structural scalability - implementation of efficient infrastructure and algorithms enabling the accommodation of these subscribers.

In order to reach the required scalability, the first evolution of the architecture, denominated as Next Generation Networks (NGNs) separated the transport functionality from the actual services. *The transport stratum was defined by ITU-T as "the part of an NGN which provides the user functions that transfer data and the functions that control and manage transport resources to carry such data between terminating entities"* [15]. The functionality addresses solely the transmission of the data information between the different communicating entities, enabling the connectivity between mobile devices and service platforms, while leaving the specific service data, control and management for the service stratum.

Based on the NGN architecture, the Next Generation Mobile Networks (NGMN) Alliance provided a coherent view on the direction the carrier grade operator community is directed in relation to the users communication [16]. NGMN defined the major requirements as recommendation for the next generation of mobile network platforms. The target architecture defined by these recommendations is an optimized NGN network architecture. It enables a smooth migration from the existing wireless access networks towards an IP network with improved cost competitiveness and broadband performance. Thus, the same IP technologies which made the Internet and the fixed access NGN successful are considered to be adapted as transport stratum to the mobile environment.

At a functional level, the future mobile networks transport stratum is separated into three distinct planes, as defined by ITU-T [15], including distinct types of functions:

- Data Plane - functions which transfer the data between the different entities.
- Control Plane - functions which control the operation of the specific layer entities and the support functions.
- Management Plane - functions used to manage the specific layer entities and the support functions

Also, for reaching the required scalability, NGMN considers that the novel core network architecture has to be able to converge the heterogeneous access networks and to offer a unified mechanism for handling the mobile device communication [17]. For reducing the network complexity, a single common mechanism is considered for the communication with the service platforms through which they can transmit application requirements [15].

Following the NGMN list of requirements, the 3rd Generation Partnership Project (3GPP) [18], standardizes convergent network architectures, mediating the connectivity characteristics between services and devices according to the momentary conditions in the access and the core network as well as based on the application and device capabilities. These architectures, including the Evolved Packet Core (EPC) [19] and IP Multimedia Subsystem (IMS) [20] are currently in deployment, bringing momentum to the mobile telecommunication industry.

Especially, EPC, which encompasses the connectivity for the LTE and the other heterogeneous accesses, is a requirement for operator network. However, EPC was designed considering a limited amount of devices, access networks and services, thus supporting the required scalability only to a certain extent and only for specific services. For fitting a larger number of devices and their data traffic the architecture design has to pass a new evolutionary step.

## 1.2 Problem Statement

The vision of the future massive mobile broadband network is to enable efficient communication to a high number of subscribed devices over a multitude of overlapping heterogeneous access networks according to the diversified requirements of the different service platforms according to the application diversification foreseen [3], [16], [6], [10]. It is expected that the number of connected devices will increase with at least one order of magnitude in the next ten years [8]. Similarly, in the next five years, the data traffic is expected to almost double every year at the same time highly varying in the resources required from the network [3]. To enable this vision, the above described operator core network evolution in technology imposes a novel set of functions into the architecture enabling higher levels of adaptation. Figure 1.2 provides an overview of the operators' issues for the core network architecture, considering the required evolution in scalability.

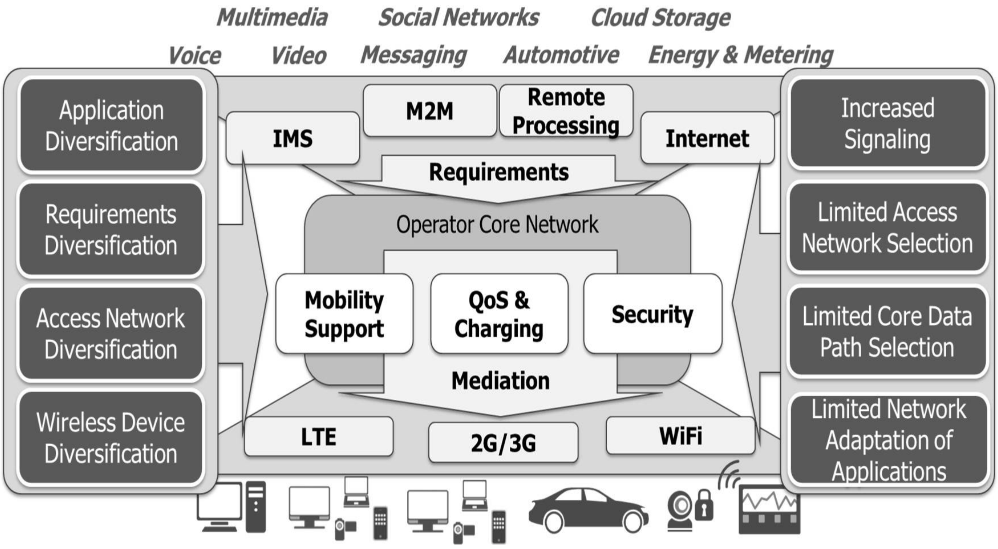


Figure 1.2: Massive Mobile Broadband Communication Concerns

In fact, there is a stringent requirement for more scalable features in the core network as to be able to reduce the operational costs through the adaptation of the resources available for each subscribed device as close as possible to the resources required without deterring the communication from the perspective of the end users. The limitations of the core network which reduces the scalability and the flexibility of the overall system currently foreseen are:

- **Increased Signaling** - Here signalling is defined as *"internal communication functions that must take place within a network in order to ensure that it* *operates properly [...] in contrast to the communications of user data, or payload"* from [12]. The signalling feature enables the control of a specific system.

The current mechanisms require that a mobile device will signal each service each time a specific application is initiated, modified or terminated. This allows the devices to negotiate with the applications a service agreement, which is then transmitted in the form of resource requirements to the core network and which in its turn guarantees the resources for the specific session. This signalling increases proportionate to the number of connected devices. However, the information transmitted through the signalling is to a large extent already available in the network, thus making the signalling redundant. Reducing redundant signaling, the radio and the core network functionality can be freed from part of the current attributions and thus, being able with the same capacity to serve multiple subscribers.

- **Limited Access Network Selection** - the current access network selection mechanisms are limited to select statically based on some subscription information a set of policies allowing the mobile device to prioritize the selection of a specific access network. [16]. However, this functionality is static to the actual network conditions and to the requirements of the mobile devices, thus making the selection inaccurate in case of multiple highly overlapping access networks. An extended functionality, able to gather the required information from the core network and to make more accurate decisions, enables the core network to better steer the devices to the most appropriate momentary available access networks, and thus to better balance the usage of the radio resources [21].

- **Limited Core Data Path Selection** - due to mobility support designed in the form of overlay circuits, the current core network can not change seamlessly the data path of subscribers. Specific entities such as the ones that act as anchors for the mobility of the device can not be dynamically balanced, while maintaining the communication seamless for the mobile devices. Also, no energy efficiency or maintenance operations can be considered through optimal geographic distribution of the data path of the subscribers [22].

- **Application Adaptation to Network Context** - current mobile core network was designed to offer a transparent data connection between devices and core network, without considering the specific time constraints and resources variation of the communication [23]. Each time some event happens in the network with the mobile device e.g. it requires to communicate, it changes the access network etc. the mobile device has to announce the applications

> in order to adapt to the new network context. However, the subscriber information is available in the core network and can be received by the applications without communication over the wireless network. Using the available information, the wireless link communication is drastically reduced during resource adaptation to the network conditions.

The current network architecture is not able to cope with these challenges mainly due to the limited communication with service platforms and to mobile devices on one side, as well as due to the lack of flexibility in the communication between the different core network functional features. In the next section, the scope of this dissertation is presented from a functional perspective.

## 1.3 Taxonomy of User Communication

Connectivity is the only service offered by the core network to the mobile devices of the users enabling them to exchange data with remote entities. For offering this service to a large number of users, the network provider deploys a network infrastructure. In case of mobile communication, the infrastructure consists from one or multiple fixed or wireless access networks suc as the Digital Subscriber Line (DSL), the Long Term Evolution (LTE) or Wireless Fidelity (WiFi) and one or more core network infrastructures e.g. 3GPP EPC, Worldwide Interoperability for Microwave Access (WiMAX) core network, etc. The core network infrastructure is then interconnected to other operator infrastructures through the Internet Backbone. In this dissertation, *An operator core network infrastructure is defined as a set of functional components that enable the connectivity between the mobile devices and the service platforms by using one or more radio access networks* [12].

Depending on the characteristics of the core network infrastructure deployed, user communication can be classified into two high level classes: unmanaged and managed communication, as depicted in Figure 1.3

### 1.3.1 Unmanaged User Communication

The unmanaged user communication is based on the best effort delivery of data packets between users and service providers. In this case, the infrastructure nodes are not aware of the communication characteristics and therefore do not guarantee the data packets' delivery. This type of communication is widely adopted by the development of ad-hoc and Internet communication. In this infrastructure, the user devices have to compete for the communication environment, which leads to lower Quality of Experience (QoE) for the end users, due to the limited resources available.

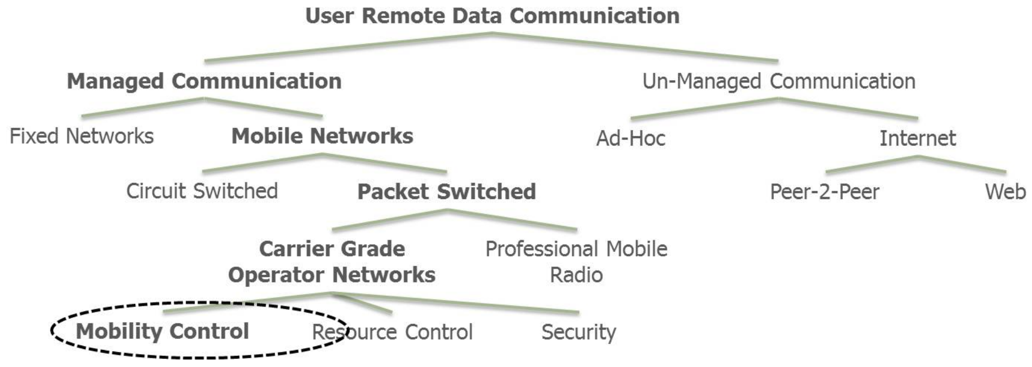


Figure 1.3: Taxonomy of User Communication

### 1.3.2 Managed User Communication

The user communication through a managed core network infrastructure is controlled on an end-to-end path from the connected devices through the access network, through the operator core network and then through the IP backbone according to the various set up Service Layer Agreements (SLAs) between the user and the service providers as well as between the service providers and the network providers. In this case, the connectivity is governed by the SLAs among all the involved parties which guarantee QoS parameters e.g. throughput per second, packet loss, delay, etc., security and privacy, as well as transparent or seamless mobility support between different locations and different access networks.

Compared to the unmanaged user communication, the managed fixed and mobile networks are offering the connectivity over an infrastructure owned by a single network provider which maintains a set of SLAs for the delivery of data towards other network providers (over the IP backbone), through this being able to ensure a complete end-to-end delivery. Being able to control the infrastructure allows the operator to further engineer the functions and components as well as the communication path of the user devices and the allocated resources for the communication data flows in order to obtain a more efficient network infrastructure and by this to accommodate more subscribers using the same physical resources and to reduce the operational costs per subscriber.

The research of this dissertation approaches only the managed mobile networks from the perspective of carrier grade operator infrastructure, as they present multiple challenges due to the structure of wireless communication and due to the generic concepts involved compared to the specialized Professional Mobile Radio (PMR) infrastructures. For carrier grade operator networks, there are multiple architectures which can be deployed: based on the 3GPP standardization or based on other architectures, such as 3GPP2, ITU-T or WiMAX Forum.

3GPP has as scope the standardization of multiple access technologies such as GPRS, UMTS, HSPA and LTE families as well as the core network enabling their connectivity. EPC is the latest standard for the core networks from 3GPP. EPC is the only core network standardized for the LTE access, therefore any deployment of this access technology relies on the deployment of an EPC infrastructure. Additionally, the EPC is able to integrate all the heterogeneous IP-based access technologies in a convergent manner thus enabling policy based control and charging, as well as security through a single generic set of mechanisms [24], [21].

Therefore, those concerned with the research in the area of carrier grade mobile communication should pay particular attention to EPC architecture for the following reasons:

- Currently EPC standardization is in its infancy. Although a first high level version of the architecture was defined and several initial deployments were made by the major operators, EPC was not tested into full capacity scenarios. Therefore, another major advancement of the architecture will be required before the actual deployments consisting into the introduction of new functions able to cope with the high increase in number of connected devices and their afferent data traffic.

- Most of the major operators committed to LTE deployment and as EPC represents the single standard means to deploy a core network infrastructure for the LTE access, it is foreseen that in the next years EPC will be widely used in the mobile communication, establishing it as the basis for the further research.

- Additionally EPC includes all the evolved functionality of the 3GPP core network architecture standardized for the previous access networks, providing currently the most scalable perspective on how massive mobile broadband can be achieved for a high number of subscribers. From this perspective, it is expected that 3GPP will further adopt architecture refinements and extensions such as the ones presented here, while supporting of future communication requirements.

- EPC was designed to encompass the functionality necessary for establishing a carrier network using other 3GPP or non-3GPP accesses. Especially for the WiMAX and for the WiFi access networks, it offers the further functionality towards the core network which was not in the goal of the specific standardization bodies. Through the EPC, the various heterogeneous access networks remain transparent to the users and to the service platforms, enabling seamless intra- and inter-technology mobility.

- Finally, EPC offers a complete integration of state of the art features related to operator IP connectivity. This includes items related to uniform authentication and authorization mechanism, encryption and compression for the radio connectivity, policy and charging control enabling subscriber profile based resource reservation and accounting and mobility including reachability and service continuity.

For these reasons, the analysis and the evaluation of the concept and models proposed in the scope of this dissertation are exemplified and critically assessed against the 3GPP EPC. Furthermore, this work will limit the scope to advancements in the area of mobility support with additional development of its implications to the resource control as depicted in Figure 1.3.

In this dissertation, mobility is defined in conformance with the ITU-T as *"the ability for the user or other mobile entities to communicate and access services irrespective of changes of the location or technical environment"* [11]. Mobility support as a network feature includes along with the ability to communicate while the subscribers are moving through the physical environment, also the same ability when the connectivity features or parameters change without the physical changes such as communication through another access network. Due to this dual scope, the term mobility support as specified by IETF [25] is preferred in detriment of mobility management from ITU-T [26]. Mobility support includes the entities and their functionality which sustains this ability from the network side including authentication, authorization, location update, reachability, and transparent or seamless service continuity.

The resource control functionality is one of the main counterpart functions in the network of the Quality of Service (QoS). QoS is defined by ITU-T as *"the totality of characteristics of a telecommunication service that bear on its ability to satisfy stated and implied needs of the user of the service"* [26]. As considered by ITU-T, among others, QoS includes parameters related to the trafficability performance - *"the ability of a system to meet traffic demands of a given size and other characteristics, under given internal conditions"* [26] - and transmission performance of a network. These two features are supported from the network side through the QoS functionality which includes the network functions for classification of data traffic, its prioritization and retention as well as its forwarding and derivation, connection admission, and bandwidth allocation, reservation and protection. The term resource control is preferred in this dissertation for the respective functionality.

## 1.4 PhD Scope

Following the vision of future massive broadband mobile networks, the goal of this research work was to develop a self-adaptation management framework integrated with the core network infrastructure, enabling scalable network resources usage including access network, core network, and applications delivery.

For understanding the scalability issues of the core network, the state-of-the-art technologies are analyzed. By examining the relevant requirements for a target platform, managing the efficient data delivery for a large number of diverse devices, over heterogeneous access networks, convergent core network and various applications and services, a set of basic components is identified for providing self-adaptation functionality. To allow the self-adaptation at the core network level, the following major aspects should be considered:

1. Communication between mobile devices and services
2. Data packets delivery optimizations
3. Security support in the core network
4. Best path selection and mobility support in the core network
5. Resource reservation and charging in the core network

The main focus of this dissertation is on the mobility support and QoS in regard to mobility including the data path selection over the radio and core network and the resource reservation adaptation to network conditions. As the first two points represent application level communication, the second data forwarding optimizations, and the third addresses security issues, they are not included. Figure 1.4 illustrates the overall research framework, following the NGN layering approach, defined by ITU-T [15].

The functionality developed addresses the transport stratum, defined by ITU-T as *"the part of an NGN which provides the user functions that transfer data and the functions that control and manage transport resources to carry such data between terminating entities"* [15]. The functionality addresses solely the data transmission between different communicating entities enabling the connectivity between mobile devices and service platforms.

The scope of this dissertation relates to the definition of a self-adaptable subscriber oriented management framework, named SelfFit which dynamically manages the parameters of control plane, modifying its behavior in specific situations, while relying transparently on the data plane. The functionality addresses the transport stratum of the mobile core networks, leaving out of scope the service level communication between mobile device applications and service platforms.

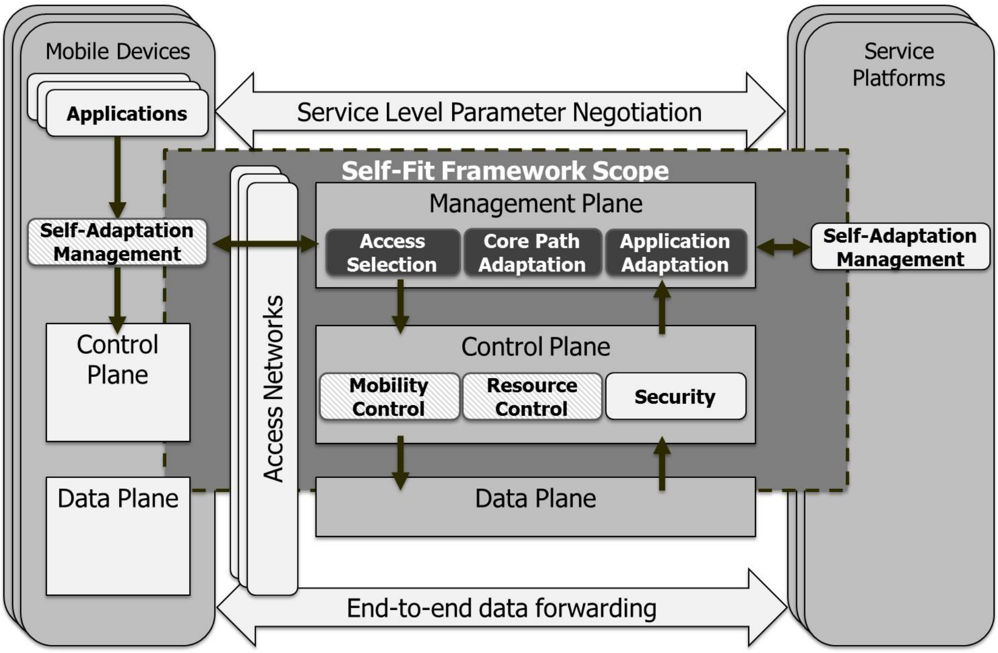


Figure 1.4: Overall Research Framework

The new features of SelfFit address the adaptation of the functionality of the *Mobility Support* in relationship with the *Resource Control*. The rest of the control functionality is considered unmodified, although integrated with the self-adaptation solution into an overall functional system. Additionally, the SelfFit interfaces to the mobile devices and to the service platforms are developed, enabling the core network to retrieve the required information and to transmit the dynamic system policies and recommendations.

The SelfFit framework has as main goal the dynamic adaptation of the decision parameters of the control plane in order to have customized data transport according to the current network status. The parameters are updated for each subscriber individually. From this perspective, SelfFit is an extension of the subscriber oriented control plane towards the management plane.

This framework completely differs in goal and approach from the Self-Organizing Networks (SON) concept [27]. SON concentrates on automatic update of the management plane parameters based on network triggers received from the management layer of other network components [28]. Through these means, the various functions can co-exist and optimize their behavior without requiring human interaction, thus solving the management complexity problem. SON concentrated only on data plane optimizations through extending the flexibility of the management plane, while SelfFit aims at the automatic adaptation of the control plane through introduction of a new specifically tailored management framework. Additionally, SON interactions and decisions are seen from an entity level and not from a subscriber perspective. Thus, the parameters updated will refer to all the subscribers which use the specific function for their connectivity. This dissertation presumes that the subscriber granularity level used in control plane and further in SelfFit framework may also be beneficial for SON. However, this hypothesis is not further detailed.

In order to reach the goals of the SelfFit platform, based on the state of the art technologies and to the requirements of the future massive broadband mobile communication, a model is defined focusing on the scalability of the overall system. The scalability is approached from the architecture optimization perspective. In this thesis, an architecture is considered more scalable when:

- its features can be customized according to the communication requirements - the control features used for a specific part of the communication can have a reduced redundancy. This implies that the communication for each subscriber can be adapted to the momentary best levels of control, thus reducing the network functionality necessary per subscriber.
- the features can be dynamically customized through self-adaptation mechanisms - during runtime of the system the parameters of the communication are continuously adapted to the most efficient communication level available. This adaptation enables that the network functionality per subscriber is appropriate to the current needs of the devices.
- the overall operations of the system are more efficient than the currently available operations - an operation more efficient for a subscriber i.e. with less signalling and processing at the communicating entities as well as with an optimized data path between the mobile device and the correspondent nodes both on the access and in the core network implies that more subscribers can be served with the same infrastructure.

This dissertation focuses on a Self-adaptation management functionality with extensions towards the control functionality, including reference communication points towards devices and service platforms, containing the following research topics:

1. **Automatic Access Network Selection** - The main objective of this topic is to examine and to optimize the selection of the access network used at a specific moment of time by a specific device through means of automation and efficiency of the handover procedures. It presumes the examination and the analysis of the access network requirements. Further, it will define

the self-adaptation management mechanism and the functional split between the core network and the mobile device, as well as communication points with other network components. The reference architecture will be based on specific standard mechanisms and extended to support a more scalable functionality. The degree of customization and the overall system efficiency are balanced against the complexity of the self-adaptation mechanisms depending on different use cases.

2. **Dynamic Data Path Self-Adaptation** - The objective is to develop a new mechanism enabling the core network to adapt the data path for one or multiple subscribers based on the information available in the core network. The focus for this topic is on the dynamic mechanisms for selecting the optimal network entities for parts the mobile subscribers communication and on the seamless data path adaptation procedures. The data path self-adaptation management is integrated with the automatic access network selection mechanism, in order to be able to offer a harmonized management for the complete data path optimization for the device through access and core network.

3. **Application Self-Adaptation Core Network Support** - The objective is to develop a core network management functionality enabling applications to self-adapt to the momentary network conditions of the wireless connected devices. It presumes the development of a novel functionality able to receive the resource requirements of the applications, differentiated on possible network conditions. The self-adaptation function is able to automatically adapt the communication parameters to the momentary network conditions of the devices and to inform the application on specific data path events.

4. **Mobility Control Support for the Self-adaptation features** - The objective is to extend the current mobility control of the core networks to support the features developed for the self-adaptation management functionality. This includes the extension of the reachability and seamless service continuity features for specific use cases and the optimization of the afferent procedures.

5. **Resource Control Support for the Self-adaptation features** The objective of this topic is the extension of the Resource Control functionality of the core network in order to support the application self-adaptation management functionality. This includes the extension of the resource reservation and event notification features, leaving the charging control out of scope.

These key functions are missing and at the point of writing of this dissertation. Figure 1.4 illustrates the roles of the functionality and the interaction within the core network architecture. This dissertation is specifically targeted to packet switched core networks. For exemplification over a standard core network architecture the 3GPP EPC was chosen, although the concepts here presented may be directly applicable on other network architectures. The proposed solution is independent of the specific applications from service level. The design of the architecture as well as the related components follows the 3GPP design for the core network functionality.

The development of the SelfFit framework follows the study, technology and dissemination steps illustrated in Figure 1.5. First, based on the future massive broadband carrier grade operator requirements, a set of technologies are selected and combined from the area of access network selection, mobility support, resource control and security resulting in a new architecture which best answers the scalability requirements.

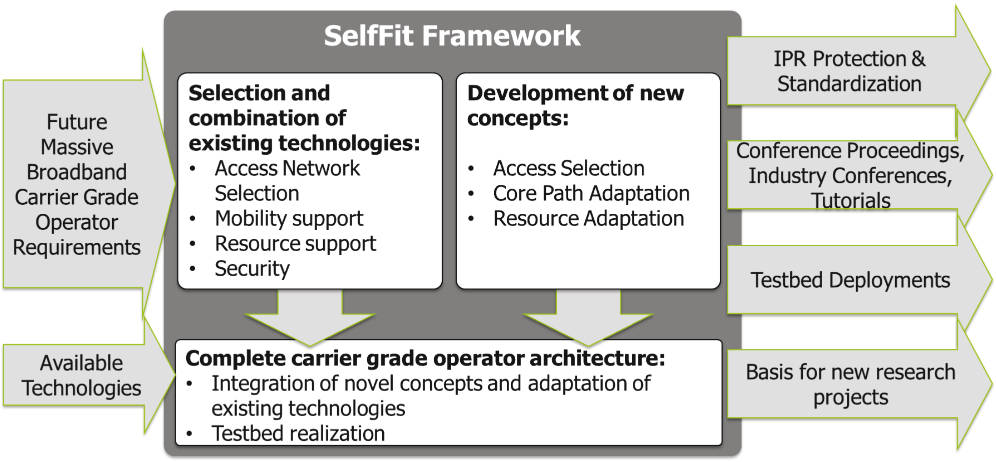


Figure 1.5: SelfFit Development Roadmap

As the current technologies are not able to answer to all the massive communication requirements foreseen, SelfFit framework will contain a set of new concepts for access network selection, core path adaptation and resource adaptation for active services. These concepts are separately developed enabling their independent deployment, as well as the possibility for evaluation their specific advantages and benefits.

Based on the combined state of the art technologies and on the novel developed concepts and their realization, a complete carrier grade architecture is developed including the SelfFit framework and other operator network functions. The architecture is then realized as part of a testbed toolkit enabling the experimentation and the evaluation of the proposed technology.

The results of the novel concepts and their integration as part of a migration path to the future mobile network architecture are disseminated in the form of

Intellectual Property Right (IPR) evaluation and protection and part of the standardization of the system architecture through the help of the industry partners from the projects through which the SelfFit framework was developed ([29], [30], [31], [32], [33], [34], [35], [36], [37]).

Additionally, the concept development, the experimentation and the evaluation results will be published through scientific proceedings and participation with specific chapters to technology related books ([38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59]).

For raising the awareness on the novel technologies in general and on the specific concepts of the dissertation, the results are presented as part of presentations in industry R&D conferences ([60], [61], [62], [63], [64], [65]) and as part of half-a-day technology introduction tutorials ([66], [67], [68], [69], [70], [71]).

The practical realization of the novel concepts and their integration with the current technologies will be included as part of the Fraunhofer FOKUS OpenEPC testbeds ([72]) which are deployed in the operator, vendor and academia environments enabling the hands-on building of know-how in the R&D departments of the institute's partners such as Deutsche Telekom Innovation Laboratories, British Telecom, Portugal Telecom, Qualcomm, etc. ([73]).

Additionally, the SelfFit framework will be brought as background technology in novel research projects, consisting the basis of the further research, projects including the European Commission Future Internet Public Partnership Project (FI PPP) Future Internet Core Platform (FI-WARE) ([74]).

The text of this dissertation is based and cites the public presentations, publications and tutorials through which the work was publicly disseminated prior to its defense.

## 1.5 Methodology

This thesis will provide a five-fold approach addressing the problems described above, as depicted in Figure 1.6.

1. Analysis and classification of principles and technologies from existing existing core network architectures to identify and discuss scalability issues in a future massive broadband mobile network.

2. Definition of the requirements from the perspective of the network operator, the service providers and the subscriber.

3. Discussion and modeling of a novel self-adaptation management functionality addressing the requirements previously identified especially targeting the future machine type communication and network self-configuration use cases.

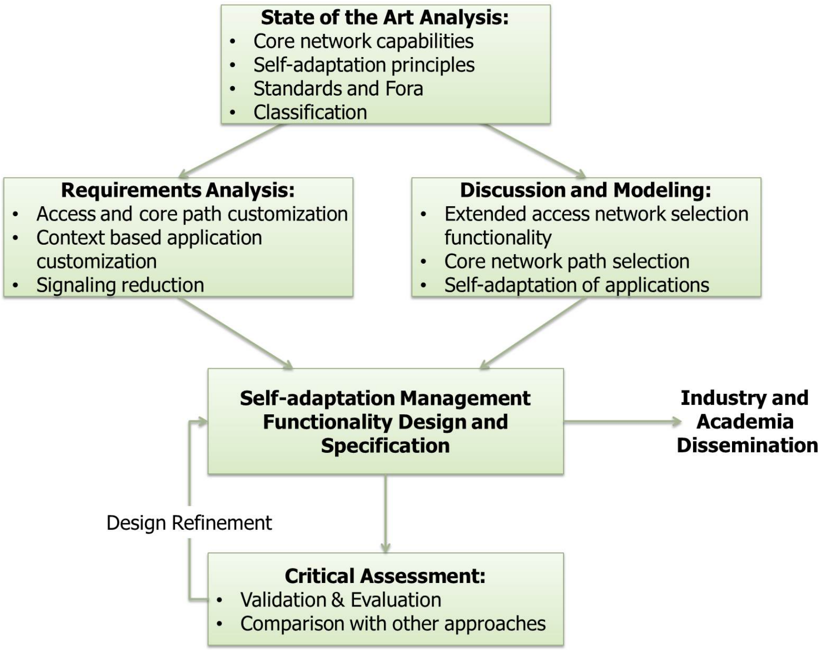


Figure 1.6: Research Methodology

4. Definition and specification of a self-adaptation management function allowing the efficient support for the previously defined use cases. The environment enable complete data path flexibility through access and core network as well as application self-adaptation to network conditions. The functionality is exemplified and integrated as part of the current core network architecture.

5. Finally, the self-management functionality and components are critically assessed through theoretical evaluation, proof of concept implementation and through comparison with other approaches.

## 1.6 Major Contributions

This work follows the foreseen evolution of the core network infrastructures, designed using mobile operator network principles in which specific features are guaranteed to the users and integrating Internet technologies for a more efficient control of the communication.

In summary, this dissertation brings the following contributions:

1. The analysis of the current network architecture from the perspective of the scalability in the future massive broadband mobile communication.

2. The classification of the requirements for a more efficient network architecture considering the foreseen evolution of the wireless environment.
3. Design and development of the SelfFit platform including the following key features:
    (a) The design and the development of an automatic access network discovery and selection function considering the momentary network status in the vicinity of the device and the most probable path of the mobile device, enabling a more accurate adaptation of the access network connectivity to the subscriber needs.
    (b) The integration of the access network selection mechanism with the core network entities enabling shorter handover duration through efficient core network information sharing
    (c) The design and the development of a dynamic data path self-adaptation functionality transparent to the mobile device and to the correspondent nodes enabling shorter data path for the data traffic of the subscriber and core network balancing for maintenance and energy efficiency purposes.
    (d) The design and the development of a core network entity reselection mechanism transparent to the subscribers enabling the network operator to balance the communication of the device through multiple entities of the same type located in specific geographic areas close to the subscribers.
    (e) The design and the development of a self-adaptation functionality for applications with reduced signalling over the wireless link and between the core network and the service platforms enabling efficient network context aware services.
4. The integration of the previous developed functionality into a common self-adaptation management functionality. The exemplification of the previously developed management functionality as enhancement to the current mobile core architecture.
5. Providing a reference implementation of the developed core functionality and disseminating the results through IPR protection, scientific publications and contributing through project partners to the related standardization organizations.

These topics are discussed throughout this dissertation and the related results are summarized in the final chapter.

## 1.7 Thesis Structure

This dissertation is structured in the following seven main parts:

- Chapter 2 provides an overview of the fundamental features of the mobile communication relevant to this dissertation
- In Chapter 3 the requirements for subscriber oriented self-adaptation in mobile systems are presented followed by their analysis, providing recommendations for the target architecture.
- In Chapter 4 an overview of the relevant technologies and standards in the context of this dissertation and with regard to customized connectivity from the perspective of access networks, core networks and service delivery.
- In Chapter 5 the design and specification of the customized core network architecture is presented. The architecture has as main goal to provide efficient service delivery for large number of mobile subscribers connected over heterogeneous access networks.
- Chapter 6 describes in detail the implementation an of the architecture as evolutionary step to the current network architectures, and the applied tools and technologies.
- In Chapter 7, the proposed solution is assessed against other approaches. Then the validation of the implementation carried out at Fraunhofer FOKUS in industry and research projects is presented in comparison with other solutions and implementations.
- Finally, Chapter 8 gives a summary of this dissertation and introduces open research items and future works.

## Chapter 2

# Fundamentals of Mobile User Communication

This chapter gives an overview of the fundamental technologies used for user communication in mobile systems.

## 2.1 Introduction

The subscriber communication is the main driving force behind the rapid increase of mobile communication market, followed at a large distance by the communication between different services and machine devices. Following the success of fixed communication, mobile communication is requiring the development of an own network infrastructure in order to circumvent specific physical limitations, to provide service with the required quality of service to a large number of users, and to maintain a reduced operational cost for the network operator.

In the first section an overview of the main features of user communication in carrier grade operator networks is given while in the second section, the current core network architecture is presented including the access networks, the core network and the inter-connection with the application platforms, exemplified using the 3GPP EPC.

## 2.2 Fundamental Features of Mobile Networks

The communication of the mobile users is realized through the connectivity service offered by the network. This service is composed of a multitude of features which may not all be required in all communication scenarios. These features co-exist in the same network infrastructure, each supporting independently a specific characteristic of the communication.

Some of the features relate to a common perspective of the communication, and therefore are grouped under the same category. The functional decomposition of the connectivity service is depicted in Figure 2.1.

The connectivity features compose the transport stratum of a Next Generation Network (NGN) [11]. They are categorized into three classes, depending on the

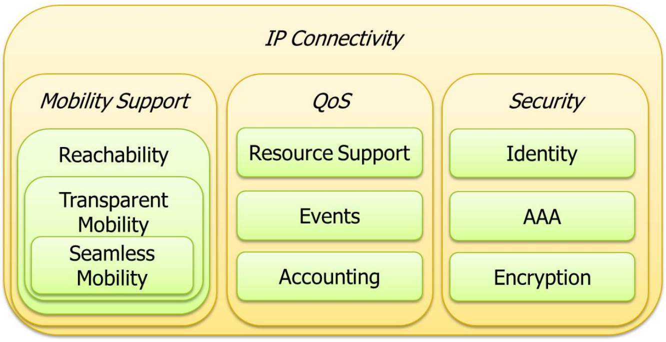


Figure 2.1: IP Connectivity Features

perspective on the communication: mobility support, QoS control, and security. The three sets of features can be considered separately. However, in general they rely on each other for a more efficient implementation of the functionality.

As considered by the NGN architecture design introduced by ITU-T [75], each of these features spans over the data, control, and management planes. For each of the features there is specific functionality which is part of the planes and of the afferent procedures.

### 2.2.1 Security

The security features ensure that the communication is executed between the appropriate parties at the privacy level requested by applications and by mobile devices. Security includes the following functions, depending on the role in active communication:

- **Identity** - in carrier grade operator networks, each mobile connected device has an own identity which allows the core network to determine to whom the communication addresses to and to execute the appropriate mobility, resource and security functions for establishing the connectivity of the device. The identity is stored in the mobile devices in Subscriber Identity Modules (SIMs) [76].
- **AAA** - the authentication, authorization and accounting function enables the network operator to determine whether the device uses its own identity and to authorize the usage of the network resources accordingly. In case of current network architectures, the feature of authorization may be part of other functions such as the reachability and the resource reservation. This functionality is application dependent [77], [24], [21], [78].
- **Encryption** - the encryption feature enables the core network to protect the

privacy of the mobile device communication over the access and core network [12].

### 2.2.2 Quality of Service

The Quality of Service (QoS) feature of the mobile network system enables the resource reservation between the mobile device and the application for which the subscriber may be charged [26]. Seen from the perspective of the transport stratum, the QoS features include the following functions:

- **Guaranteed Resources** - the communication requirements are received by the core network from either the mobile device or the service platform. Based on these requirements and on the subscription information of the mobile device, the core network guarantees the resources of communication on the data path which include a specific level of throughput with a guaranteed end to end delay and a specific guaranteed limitation on packet loss or jitter. The guarantee of resources enables the application to reduce their internal functionality used over best-effort networks and to achieve the service level agreement between subscriber and network provider [26].

- **Charging** - the charging functions are able to track, to associate, and to differentiate data packets based on flows; to meter, to rate the communication of each subscribers, and to transmit it in an aggregated form to the billing systems [79].

- **Event Notification** - as the core network is aware of events happening at the data path, it can transmit event notifications to the service platforms or applications which subscribe for such information. This feature empowers the applications with knowledge on the current context of the mobile devices [79].

### 2.2.3 Mobility

The mobility feature enables correspondent nodes to reach a mobile device regardless of the momentary location in the network. Depending on the support level, mobility can be separated into the following features:

- **Reachability** - from the moment a device attaches to the network, it receives apart from its unique identity, a temporary reachability address like an IP address making it available for remote communication. This address enables third parties to reach and to forward appropriately data packets to the device. The device is reachable throughout the duration of its connectivity to the network, regardless of its momentary network location, as long as the correspondent parties are aware and can forward the data traffic to the current reachability context [80].

- **Transparent Mobility** - in case the reachability context of the device does not change due to location changes, mobility is transparent to the correspondent nodes. Thus, end-to-end sessions do not have to be re-established [80].
- **Seamless Mobility** - mobility support with no data packet lost is named seamless mobility [25] or Seamless Service Continuity [81].

The main procedure related to mobility feature in cellular carrier grade operator networks is named handover. A handover [25] is the process by which a mobile node changes its point of attachment to the network or when such a change is attempted. A synonymous term for handover is handoff. The term handover is preferred in this dissertation.

## 2.3 Network Architecture

This section provides an overview of the IP-based network architecture including the access and core networks and their interconnection with the application platforms.

### 2.3.1 Access Technologies

Due to the advancement in mobile radio, providing massive mobile broadband to a large number of subscribers is becoming a reality [3]. These access technologies complement each other in terms of cell size, available throughput, packet loss, transmission delay, and operational costs, as depicted in Figure 2.2.

Due to the deployment of new accesses as an extension of the already existing technologies, in the near future it is expected to see a significant growth of heterogeneous wireless environments in which mobile devices are able to connect and to exchange data over multiple access technologies at the same location [82].

#### 2.3.1.1 3GPP Wireless Accesses Technologies

3GPP [18] has the mission to standardize the large wireless area networks, currently taking the lead of the initial standardization phases for the 4th Generation (4G) of mobile communication systems [83], [84], [85] while supporting 2G and 3G accesses.

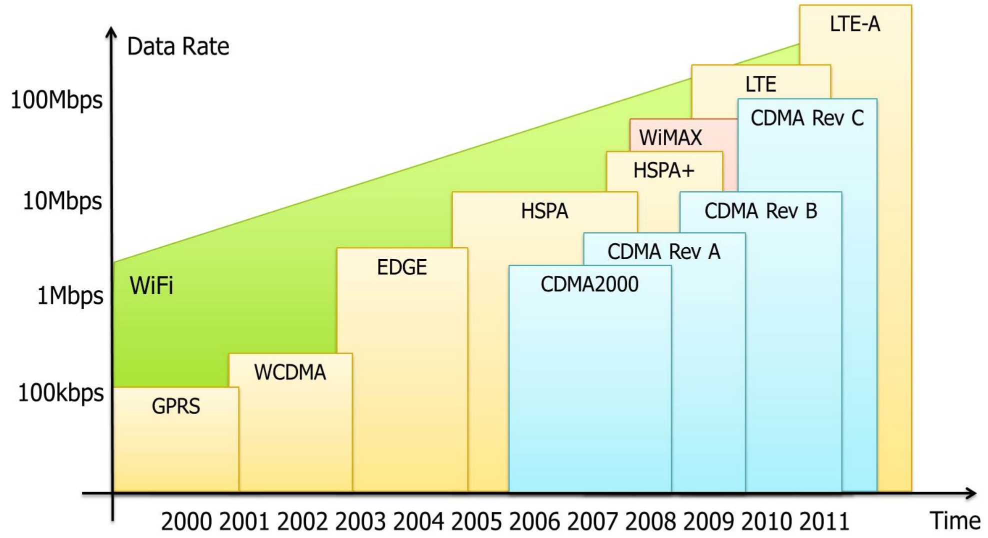


Figure 2.2: Wireless Technology Evolution

Although an unprecedented success, the GSM (Global System for Mobile Communication, originally Groupe Special Mobile) was standardized as a mobile voice centric access network, capable only of limited circuit data communication. The General Packet Radio Services (GPRS) was a packet switched data communication extension of the GSM network more suitable for data applications as it relied on sharing of communication channels between several mobile stations. Following, the EDGE (Enhanced Data rates for GSM Evolution) access technology provides an increased data rate and lower delays [86].

The UMTS (Universal Mobile Telecommunication System) comes as the first 3rd Generation (3G) wireless system. Having a redesigned network architecture, it can offer a highly increased throughput for an even more increased number of subscribers. Following, the HSPA (High Speed Packet Access) technology further developed the 3G communication towards increased throughput with reduced transmission delay.

A break-through was obtained in 2009, when the LTE (Long Term Evolution) access network was first standardized, reaching new levels of capability and performance. The uplink and downlink data rates are significantly higher, requiring 3GPP to further standardize also a new network core architecture and other technology enhancements. Additionally, LTE provides a reduced cost/bit rate compared to the previous access technologies, enabling carrier grade operators to address mass market needs for data communication.

However, LTE is the first step in the standardization of a new type of access network which is meant to fulfill and surpass the requirements of ITU for the 4th

Generation of radio systems [87]. Consolidating the results obtained by the first LTE deployments, an LTE-Advanced evolution of the technology is currently in standardization [18].

Through the evolution of 3GPP access technologies, the same basic concepts are used [19]. Each technology is build as a next step of the previous one, enabling the research designed for a specific access network to be further disseminated in previous and in future accesses.

In order to further increase the available resources for the subscribers in specific locations, the carrier grade operator networks are considering the deployment of small and medium size networks through the usage of small coverage cells of the same access as the wide area network technologies named home nodes or femtocells [88], [89].

Currently, the HSPA technology is deployed as an addition to the GPRS and EDGE networks by most of the carrier grade operator networks [4]. LTE access technology is deployed by a limited number of operators around the world, followed by a large number of commitments in the following years [90]. It is foreseen that the 3GPP accesses will be deployed in addition to the previous ones and not as a replacement mainly due to the support for legacy mobile devices.

#### 2.3.1.2 Non-3GPP Access Technologies

Apart from the 3GPP accesses, operator networks are currently diversifying the access networks which are provided to the mobile devices [82]. In this dissertation, these wireless technologies are named non-3GPP accesses. Following, the main set of non-3GPP wireless technologies is presented.

One of the large scale deployed access technologies is the IEEE 802.11 Wireless Fidelity (WiFi). WiFi is mainly used as a wireless local area network with a reduced coverage range and with a large throughput [86]. Using unlicensed spectrum, medium deployments are also possible such as in enterprise environments [91].

The IEEE 802.16e Mobile WiMAX (Worldwide Interoperability for Microwave Access) standard is specifying a wireless broadband access technology which is capable of a wide throughput for large areas. Containing new radio technology features, WiMAX supports a wider coverage as WiFi and a better control of the radio environment through the introduction of resource allocation mechanisms [86].

In parallel with the 3GPP 2G and 3G standards, the 3GPP2 standardization organization developed the CDMA standards which follow the same radio concepts, however differing in used spectrum, technology realization and core network architecture. Currently, 3GPP2 continues the evolution of the standard technologies without attempting to develop a 4G technology equivalent to LTE [92].

As a high number of the current fixed network deployments are terminated with a wireless access of the previously presented 3GPP or non-3GPP technologies, fixed networks are considered in this dissertation as part of the non-3GPP accesses through which mobile devices can connect and exchange data. For fixed connectivity, various technologies are used. Three major technologies for fixed connectivity are currently widely deployed. First, the Digital Subscriber Line (DSL) is a modem based access which uses the telephone lines for transmitting high speed data towards subscriber premises. In order to further increase the data rates, Passive Optic Fiber (PON) is currently deployed as a replacement for the DSL type of access, offering higher capacity at a reduced delay [93]. The IEEE 802.16d Fixed WiMAX provides a wireless alternative for the last mile connectivity, having distinct requirements and architecture compared to the Mobile WiMAX - no mobility support, no energy constrains and loose resource limitations in the endpoint device [94].

### 2.3.2 3GPP Evolved Packet Core

With the progress towards high data rate wireless access technologies, a complimentary evolution was required from the side of the core network architecture. As observed from the fixed communication evolution, the services that the user require over broadband access are not any more limited to voice and reduced data exchange such as email. Instead, they are using all of the resources available, which transforms the mass broadband wireless environment into an IP data dominant one. In order to face this challenge, the novel architecture is considered to rely on IP communication and to use a similar forwarding mechanism as the Internet, adapted to the wireless environment while being more efficient than the fixed infrastructures.

Following the indications of the NGMN Alliance [16], 3GPP initiated the standardization procedures for EPC [95], an all-IP core network architecture specifically designed for the communication requirements of the LTE access and offering IP convergence for the heterogeneous access networks in a transparent manner to the applications as depicted in Figure 2.3 [24], [21].

As EPC is the only core network architecture supporting the LTE access technology and as LTE is gradually rolled out with a time horizon of more than ten years until complete deployment [96], the EPC architecture represents a technology landmark for the core network development. Furthermore, the further scientific development of the EPC core network will most probably be integrated in the next generations of the core network architecture, as it happened until now throughout the evolution of the mobile core networks.

Other architectures are available from other standardization organizations such

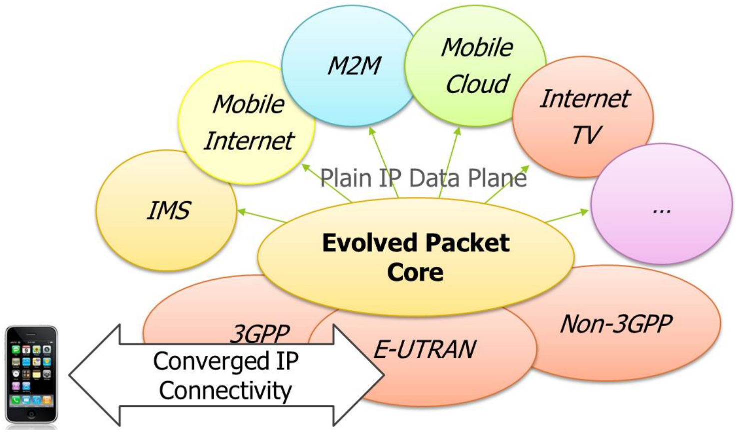


Figure 2.3: 3GPP Evolved Packet Core: All-IP Access and Core Network Convergence

as the NGN architecture from ITU [97]. These architectures include the same functional features as EPC because they offer the same type of connectivity support. However, either their standardization status is still in very early stages or the specific access technologies for which they are developed are not having a global momentum similar to LTE. Therefore, although their features are considered as part of the background literature, they are not further considered for the development of the framework introduced here.

In this dissertation, EPC was chosen for exemplifying the concepts. This is due to the level of maturity of the technology and to the prospective carrier grade operator deployments realization. Compared to EPC, the WiMAX Forum architecture is designed only for WiMAX access and does not consider the requirements of heterogeneous access networks. On the other hand the ITU architecture is still in a high level design state without being applied to the characteristics of the various access technologies, thus, lacking the possibility of being deployed in a foreseeable future.

As LTE is still in its first releases in standardization, it is foreseen that EPC architecture is not yet completed. Major requirements are expected to be further received from the first deployments as well as with the increase in scale of the number of connected subscribers and of the number of applications especially tailored for the mobile environment [98], [99]. Because of this a large space for novel concepts and innovation is envisaged.

Additionally, EPC sustains also in an integrated manner the other heterogeneous accesses such as 3GPP e.g. UMTS, EDGE, GPRS and non-3GPP e.g. WiMAX, CDMA etc., being able to interconnect with any available or future wireless technologies including fixed communication [24], [21]. For mobile devices connected to these accesses, EPC provides full convergence at IP connectivity level including the support for identity, authentication and authorization, policy and charging control and mobility support.

Through the convergence at network level, EPC is enabling efficient connection of mobile devices to the network and their communication with different service platforms such as IP Multimedia Subsystem (IMS) [100] or Internet, being able to control the resources allocated to various devices and offering transparent service continuity across the same or different access technologies and by this offering to the service platforms an independent connection over the wireless accesses. From service platform perspective, EPC manages event oriented policy based access control, resource management and mobility in single carrier grade operator core network together with accounting and charging functionality based on the subscription profile and the active services of the mobile device. These characteristics make EPC the next generation transport level for signaled services.

#### 2.3.2.1 EPC Architecture

EPC is designed based on the concepts of previous 3GPP architectures ([79], [24], [21], [101], [102]). It represents a long term evolution of the UMTS architecture, which was designed for voice and data transport [78]. Additionally, it maintains the subscriber-based resource reservation and mobility concepts from the IMS architecture.

EPC contains a clear delineation between the control and the data path. Its main components are depicted in Figure 2.4. A correspondent mobile node is also considered, named User Endpoint (UE). In this dissertation UE is denominating the specific EPC device while Mobile Node (MN) is considered a broader, more generic naming. EPC components are classified based on their functionality as follows:

- **Subscription Data Entities (e.g. HSS, AAA Server, etc.)** - These entities store, update, transmit notifications on the users' subscription profile, and perform authentication and authorization procedures. Additionally, they contain dynamic user information such as current allocated IP addresses.

- **Control Entities (e.g. PCRF, MME, ANDSF)** - These entities make policy based decisions regarding mobile device connectivity, access control and allocated resources. Based on various subscriber related triggers, they make decisions which are afterwards installed on mobile devices and data path components.

- **Gateways (e.g. Serving GW, PDN GW etc.)** - These data path entities are forwarding the data traffic and ensure the access control, QoS and mobility support according to the decisions made by control entities. They

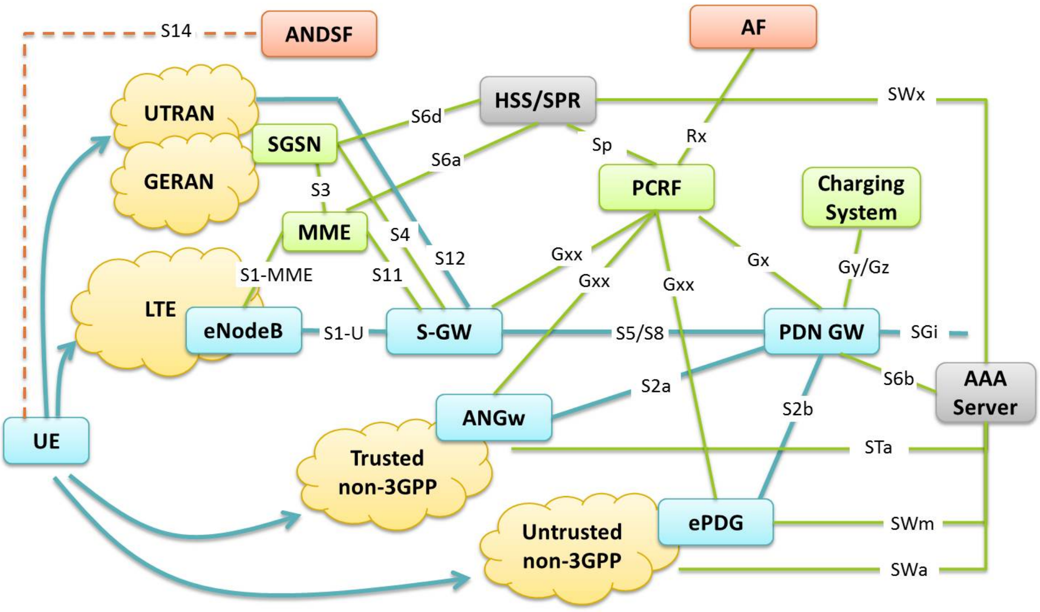


Figure 2.4: 3GPP Evolved Packet Core Architecture

are also able to transmit subscriber related events to the control entities in order to adapt the policy decisions accordingly.

Together these entities enable that the IP connectivity is provisioned and the resources are allocated according to the profile of the user and based on the requirements of each application. The applications are considered external to the EPC architecture and they are generically named Application Functions (AFs). The role of the AF can be taken by an IMS core, by a service broker, by an intermediary node of the operator on the application path or directly by infrastructures of third party service providers. Additionally, the operator may deploy traffic detection tools which can also transmit events which generate policy decisions modifying the IP connectivity.

Furthermore, EPC contains accounting functions enabling the charging of the IP connectivity. The charging is based on the data path sessions and can be executed synchronous or asynchronous to the data exchange.

The control functionality of the EPC is based on subscriber information and does not consider that during the policy decisions information on the network status is available. Because of this, even though the decisions are taken considering the highest level of resources which may be allocated to each subscriber, they might not be available in the radio and on the data path. Because of this, the policy decisions are taken faster as they lack the complexity of input parameters.

However, in exception cases, which are more frequent due to the missing information during decision taking, the procedures are highly extended.

### Subscription Data Entities

The Subscription Data Entities store and are able to send notifications for the subscription profile of the UE. They perform the authentication and the authorization upon the attachment of the mobile device to an access network.

The Home Subscriber Server (HSS) is imported from the previous 3GPP IMS architecture. In EPC, it maintains the subscription information for each user, including restrictions for the attachment and resources that can be allocated over the different access networks. It also maintains dynamic information on the network location of the UE and entities which serve it in the control and in the data path [103], [104], [105].

Another repository with similar functionality is the Subscription Profile Repository (SPR) [79]. It is a logical function which maintains the subscription related information necessary for the policy based decisions for access control and resource reservation. As the current standards do not specify the internal data structure of this repository and following the standardization direction in 3GPP to unify different subscription data entities, the SPR is associated here with the HSS, because of their similarities.

The AAA Server was initially defined for inter-working between the 3GPP and WiFi and it was extended in EPC for supporting authentication and authorization in non-3GPP accesses [106], [104]. It retrieves and updates the subscriber profile information from HSS allowing also report generation for charging based on the authentication requests received from the non-3GPP access networks.

### Control Entities

The control functions of EPC manage the access control, the mobility and the resource reservations. The control mechanisms of EPC are triggered by events including the user profile modification, the attachment of the UE, resource required changes or access network context changes.

The Mobility Management Entity (MME) is the central management entity for LTE access [24], [101]. It is responsible for connecting UEs, by selecting the gateway through which messages are to be exchanged and a level of resources. It also provides authentication, authorization, location tracking, and intra-3GPP mobility support e.g. between 2G/3G and LTE.

The Policy and Charging Rules Function (PCRF) is the control entity making policy based decisions for service data flow detection, admission control, QoS and flow based charging [79], [107], [108], [109], [110]. It maintains a complete subscription state for each UE with decisions made on a per-user base. PCRF cannot be associated with a management entity as it does not maintain information on the resources available in the different access networks. Instead for each decision an enforcement procedure is to be executed in order to determine whether the requirements can be fulfilled by the gateways.

The Access Network Discovery and Selection Function (ANDSF) makes subscription based decisions and transmits to the UE information on the operator's preferences for accesses discovery and handover decisions [111], [102]. ANDSF uses UE's location, a Coverage Map and subscription information, to make these decisions. As inside 3GPP accesses the preference of the operator is controlled by the MME, ANDSF addresses only handovers with non-3GPP accesses.

**Gateways**

In EPC, the gateways ensure the forwarding of data packets between UE and network core within the parameters enforced by the control entities. They also support the execution of different mobility protocols and transmit notifications to control entities for data path related events.

The Packet Data Network Gateway (PDN GW) is the break out point for data traffic in EPC. It provides connectivity from UEs to external packet data networks. Additionally, it acts as central mobility anchor. The PDN GW performs policy enforcement and packet filtering for each data flow of each subscriber. For this, it maintains the context for each connection of the UE, traffic flow templates for the active services, QoS profile and charging characteristics.

EPC is able to offer a unified enforcing of policies for different access networks. For each class of access networks, a different gateway is defined. For 3GPP accesses a Serving GW (S-GW) is used. Additionally, a Serving GPRS Support Node (SGSN) is used for the 2G/3G access technologies. The non-3GPP technologies are separated based on the security level the operator has to provide over the access network in trusted and un-trusted non-3GPP accesses. In the untrusted non-3GPP accesses additional security levels are to be established. As gateways, for the untrusted non-3GPP accesses an evolved Packet Data Gateway (ePDG) is used while for the trusted non-3GPP a generic Access Network Gateway (ANGw) is deployed, differing for each access network type.

The gateways may include a Bearer Binding and Event Rules Function (BBERF) which provides policy based enforcement and event notifications from the wireless link to PCRF and gating functionality for data traffic. They also include the attachment and mobility related functionality for both control and data path.

#### 2.3.2.2 EPC Key Protocols

EPC bases on IP protocols, as standardized by IETF and OMA, enhanced by the 3GPP specifications. It is able to communicate over IPv4 or over IPv6 or over a mixture of the both. Three main categories of protocols are considered, depending on their goal in the EPC: active control protocols, mobility protocols and remote mobile device management.

Diameter is used as a singular control protocol [112]. The basic IETF standard was enhanced by 3GPP for supporting not only authentication and authorization, but also policy based decisions and event notifications related to the connectivity and the different data flows of the subscribers. Diameter is deployed on the interfaces between the subscription repositories, the control entities and the data path entities supporting the operations required for access control, QoS and event notifications[79], [103], [104], [107], [108], [109], [110].

Multiple mobility protocols can be used like GPRS Tunneling Protocol (GTP) [105], Mobile IPv4 (MIPv4) [80], Dual Stack Mobile IP (DSMIP) [113] and/or Proxy Mobile IP (PMIP) [114], [115]. Each of these protocols has different capabilities and is in a different development stage. Although they may function independent of the architecture itself directly on the data path, as specified by IETF, in EPC, the entities defined in these protocols are interconnected with Diameter control interfaces which supply parameters and through which different events and operations are to be triggered. All these mobility protocols have some similar characteristics. First, they are transparent to correspondent nodes, enabling the communication to any IP based service which may or may not pertain to the carrier grade operator which deploys the EPC. Secondly, the mobility protocols rely on a centralized anchor node which enable the mobility of all the devices in all the accesses and in all the locations covered by EPC.

In EPC, the access network discovery and selection functionality was included as part of the operator controlled management of UE. Thus, OMA Device Management (DM) protocol was chosen as transport protocol, because it is already integrated in existing mobile network architectures [116]. A novel Management Object was defined specifically for the access network discovery and selection for conveying the in-formation from the network to the mobile devices [111].

#### 2.3.2.3 EPC Functional Features

EPC is designed to provide a number of key capabilities supporting seamless and efficient service delivery. The main functionality includes:

- Network access control functions
- Resource Control

- Mobility Support

### Network Access Control Functions

The network access control functions enable UEs to connect to EPC and then over EPC to different service providers. They include the authentication and authorization, admission control, selection of the entities which will serve the UE (e.g. the PCRF, the PDN GW etc.) and the establishment of a minimal context which enables the UE reachability in the IP domain and its basic communication with service platforms.

The authentication and authorization procedures are dependent on the access network selected for attachment. For LTE, MME retrieves the subscriber profile from HSS during the attachment, for other 3GPP accesses, SGSN controls the authentication and authorization while for non-3GPP accesses, ANGw or ePDG request the information through the AAA Server from the HSS [117], [118].

Then, the IP reachability context is created for the UE. For GTP or PMIPv6, upon an address request from the UE, a data tunnel is established between the gateway of the specific access and the PDN GW. In case of MIP or DSMIP, an address is allocated locally and the tunnel is established directly between the UE and the PDN GW.

During tunnel establishment, the PDN GW initiates the procedures for the basic resource reservations which allows the UE to communicate with service platforms. In this procedure, the PCRF makes a policy based decision whether the UE is allowed to communicate over the access network and which default level of resources is to be allocated for the basic communication.

After this decision is enforced and the mobile device receives the IP address over which it is able to communicate, the attachment procedure is completed and it may exchange data over access network and EPC - it has a Packed Data Network (PDN) connection.

The network access control functions allow the EPC to coherently apply policies during the attachment and detachment of UEs according to the subscription profile of the user in the SPR/HSS.

### Resource Control

EPC has a central notion for resource reservations, the IP Bearer, representing an aggregation of IP data flows that receive a common QoS treatment (forwarding, scheduling, queuing, shaping etc.) established between UE and PDN GW.

In EPC there is a clear distinction between access network and core network resource reservation. Access resource reservation is technology specific and does not influence directly EPC operations. On the other hand, EPC was designed as support architecture for LTE, thus the MME is managing resources inside this access technology.

For all access technologies, the core network resource reservation can be triggered by the UE or by service platforms. The UE trigger is received by the gateway of the specific access which forwards the request to PCRF. The service platforms triggers are transmitted directly to PCRF. By this centralization into PCRF, the consistency of the QoS rules enforced for an UE is maintained. Then the PCRF makes the subscription based policy decision whether the UE is allowed to reserve the required resources. The decision is enforced to the PDN GW and to the gateways of the specific access networks and from here using specific mechanisms to the wireless access network.

As PCRF does not manage the resources available on the different wireless accesses and gateways, this enforcement may fail. From this perspective, PCRF is not a resource control entity as it is not able to consider the context in different accesses, but as a user control entity because it sustains its decisions on input regarding the UE and its applications. This allows the concentration of the functionality on the UE itself and not as much on the momentary context of the accesses, enabling a scalable flat decision mechanism.

In order to reduce the complexity, EPC considers that the default bearer established at the attachment to the EPC and the subsequent bearers, requested by UEs or by service platforms, are reserved using the same procedures. This allows to include network access control functions as resource control functions.

#### Mobility Support

EPC integrates multiple access technologies. For supporting mobility between the different access networks, a set of mechanisms was defined which include access network discovery and selection and correlated attachment to target access and detachment from source access relying on existing mobility protocols like GTP, PMIP and MIP.

LTE was designed to be a direct extension of the other 3GPP accesses, as UMTS was an extension for GPRS. The Intra-3GPP mobility support bases on the MME to select the next attachment cell and to initiate a preparation procedure. When it is completed, the MME transmits a handover command to the UE.

In case of non-3GPP technologies, 3GPP cannot influence directly the handover procedures, due to the various standardization groups involved for access networks specifications. Thus a complete network controlled handover cannot be realized. In order to limit the effect of the UE independent selection of target access which leads to congestion of some accesses while others are capable to sustain seamlessly the communication, ANDSF was introduced in the architecture. Based on the location information, it transmits to the UE the operator preferences for target accesses in case a handover becomes necessary. The functionality of ANDSF restrains to this general indication. The UE has to consider the policies received and decide independently when a handover is to be executed.

The effective handover procedures are separated logically in attachment to target access network and detachment from source access network. In the handover with optimization case, like in the intra-3GPP handovers, a preparation phase is executed which creates a user context on the target access network. However this is not possible in handovers from and to non-3GPP accesses, as the handover decision is taken by the UE. In this case, the mobility relies on the mobility protocol deployed, e.g. GTP, PMIP, classic MIP. It presumes that at the moment when the UE attaches to the target access network, data traffic is forwarded to the new access even though the source access network connection is still available. From this perspective EPC is able to execute soft intra-3GPP handovers and hard handovers with the non-3GPP accesses.

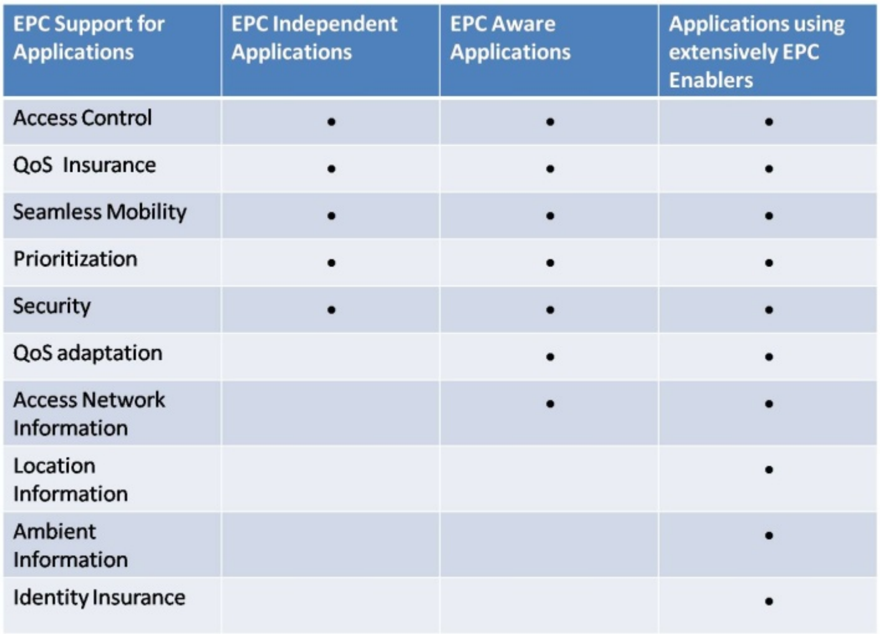
As all the mobility protocols deployed in the EPC rely on a centralized anchor node and as handover procedures are triggered by changes of UE location, currently the EPC does not provide any mechanism in which data path can be adapted for each subscriber depending on current network status and according to application delay requirements.

### 2.3.3 Interconnection with Application Platforms

EPC provides a transparent convergent network layer for the IP applications. From the perspective of a service provider without a modification of the application, it enables a degree of satisfaction similar to the fixed IP applications, by transparently supporting features like access control, QoS insurance, seamless mobility between different access networks, prioritization and security. Based on this it is foreseen that the mobile application environment will adapt and integrate the ones previously deployed on fixed Internet.

Also due to the resource reservation mechanisms, the services have a guarantee of the quality of the communication which is an addition to the typical IP communication and a high added value for broadband communication on mobile devices with reduced processing power.

EPC provides also a control interface between the service platform and the network core [109]. Through this interface, the EPC aware applications can transmit indications on resources that have to be reserved for specific users. They can also receive upon request information on events happening at link and network layers e.g. UE lost connectivity or a handover to another access network occurred. By these mechanisms, the applications can be adapted to the momentary UE context and to offer services customized not only based on the service level user profile, but also to the UEs in use and to the surrounding network context. In this class of applications may enter the services offered by the operator or by third parties



| EPC Support for Applications | EPC Independent Applications | EPC Aware Applications | Applications using extensively EPC Enablers |
|---|---|---|---|
| Access Control | • | • | • |
| QoS Insurance | • | • | • |
| Seamless Mobility | • | • | • |
| Prioritization | • | • | • |
| Security | • | • | • |
| QoS adaptation |  | • | • |
| Access Network Information |  | • | • |
| Location Information |  |  | • |
| Ambient Information |  |  | • |
| Identity Insurance |  |  | • |

Figure 2.5: Application Support in EPC

having an agreement with the operator, like IMS services, mobile cloud computing etc.

Although not yet standardized, EPC is able to export a set of enablers to the applications which offer even more flexibility to the user in controlling the service delivery. For example, services may use the UE location or even ambient information and UE subscriber identity, for further adapting to the environment conditions and to ensure a more secure communication. In this category fall the future mobile applications adapted to subscriber and to his surrounding environment.

With the development of the mass broadband environment offering data dominant communication, a novel evolution of the applications that mobile users will be using is foreseen. This is due to the complete personalization of the mobile device as communication instrument which allows the specific user to exchange data any time and anywhere. In order to face this challenge, the novel architecture is limiting its goal in offering added value at the level of network connectivity and of different enablers for service platforms which enhance the communication (e.g. resource reservation enabler). This increases the acceptance of the overall system and of the novel mobile applications and thus the acceptance of the mass broadband wireless environment.

CHAPTER 3

# Requirements for Mobile Broadband Connectivity

This chapter defines the requirements for an efficient core network architecture for mass mobile broadband communication through self-adaptable subscriber oriented control.

## 3.1 Introduction

The perspective of multiple actors is adopted in this dissertation considering mobile device users, service providers and especially network providers. The requirements are presented as arising from the opportunity brought by the radio technology advancement for a new telecommunication environment enabling a massive number of subscribed devices to be attached and to communicate over packet data networks.

Based on an extended set of requirements, standardization organizations such as 3GPP and ITU continued their architecture specification efforts. However, at the moment of writing this dissertation, for the requirements selected here, the standardization organizations did not complete the description of the specific features, therefore being still open for further academic study and investigation, prototyping and proof of concept and then by trials and architecture consolidation which are usually followed by the technology deployment.

First, the key properties expected from the massive carrier grade network communication are presented from the view of different use case actors. Then requirements related to the access and core network from the perspective of the network provider and of the generic service platforms which the operators are foreseen to deploy. Based on these requirements, a set of evaluation criteria are drawn, against which the related technologies and the provided solution are later evaluated.

## 3.2 Carrier Grade Foreseen Key Properties

With the evolution of mobile technologies, the distinct actors require and have to adapt to a novel wireless system. As technology evolves towards providing a higher level of resources in specific network areas, it offers from the perspective of users and of the application providers extended opportunities for communication.

However, as the mobile communication relies on cellular access networks, resulting in competition for the wireless environment and as the available network capacity increases, the core network has to be as efficient as possible in delivering the communication at the levels required by mobile users and by application providers.

In this section these key properties of the carrier grade communication are further detailed and motivated.

### 3.2.1 Mobile User Communication

From mobile user perspective, with the capacity increase, new opportunities of communication are foreseen. This key characteristic of user communication is augmented by the diversification of the available communication devices such as smartphones, tablets and laptops and by the adoption of novel applications [4].

Currently, a large pallet of smartphones is available, in a market which is consolidating around a limited number of basic architectures. This enables a reduced production cost which further increases the acceptance rate of the users [119].

The gradual adoption of smartphones is creating new application opportunities especially tailored for mobile communication, enabling a further service diversification. Remote data communicating machines are further diversifying the requirements on the core network [6].

Additionally, the users of carrier grade wireless communication are using laptops which were extended with additional modem components to be able to connect to operator provided networks [82]. The users of these devices are considering as key property of the mobile system, the maintaining of the similar experience level as in the fixed lines while using their applications. Devices which were traditionally connected to fixed or to home access networks, are now requiring for their communication the functionality provided by mobile core networks. However, services used by this devices are mainly designed for fixed networks, translated for the mobile environment.

Coming to bridge the gap between smartphones and laptops, mobile tablets are providing a mixture of the features of the previous two devices features, thus mixing the adapted fixed and the especially designed mobile applications [82].

In the search of an increased user experience, current mobile devices are featuring more than one access network modem enabling multi-homing scenarios [82]. the devices are able to connect and to exchange data over multiple heterogeneous accesses, providing different wireless capabilities in terms of coverage, data rates, delay etc [120].

For fully adopting the data communication, a key feature required by mobile users is the reduced cost. To reduce the costs of mobile subscribers, carrier grade network providers have to further reduce the operational costs.

In conclusion, from mobile user perspective, the following key properties should be provided by carrier grade networks:

- Support for data exchange with applications. It includes applications translated from fixed communication and applications especially designed for mobile.
- Support for the experience levels to which the users are accustomed from fixed line communication. For providing the same quality of experience, specific additional features should be considered in the mobile networks such as mobility support.
- Support for multi-homed devices. The core network has to be able to sustain multiple wireless communication paths for a device.
- Reduced costs. The cost plays an important role for users, especially in environments with massive number of subscribers.

### 3.2.2 Mobile Service Platforms

The increased capacity of wireless networks and the reduced cost of the network communication represent high incentives for developing a large variety of applications targeting independent classes of subscribers [6].

The reduced latency, high data rates and reduced packet loss over wireless links, augmented by processing power increase in mobile devices, opens a new opportunity for applications which consume high level of resources such as video delivery, intensive multimedia communication, gaming, as well as remote computation and storage using cloud computing paradigm, peer-to-peer based communication and services and content offered by mobile devices [82].

Additionally, the new mobile services provided can be adapted momentarily available resources and thus, reducing the cost of the momentary communication while maintaining an acceptable quality of experience for subscribers.

The application diversification is also sustained by the deployment of novel end communication devices, autonomous to human intervention generically named machine devices. Sensing and actuating the environment, these devices are foreseen in the next five years to consume more network data capacity that half of the current data traffic [82].

The novel applications designed for the mobile environment are using specific information such as location and presence information in order to adapt their intelligence and their communication accordingly [121], [122]. This information may be provided by mobile devices at an extended delay and signalling cost or directly by core network.

Spanning from small sensors and actuators, to live high-definition streaming devices, the number of machine connected devices will increase with at least one order of magnitude the overall connected devices [8], requiring connectivity for a new generation of M2M services, addressing other markets than current communication, such as Smart Cities, eHealth, Smart Grids, Automotive, etc. Additionally, they will require network support for their communication over the wireless environment.

This property required from the core network is two folded. On one direction, in order to be able to deliver the services to all the appropriate devices, those devices have to be able to communicate over the network. This side of the key property is addressing the network capability to enable the communication for a multitude of devices with very heterogeneous resource requirements.

On another direction, the service platforms have also a similar device number scalability problem [123]. In order to alleviate this issue, the network can provide various features transparently to application platforms such as access network selection, resource reservation adaptation and dynamic data path support. When these features are supported from the core network, service platforms require to maintain less functionality per subscriber, thus, being able to sustain more users with the same infrastructure and thus reducing the costs per user.

Additionally, mobile service providers are requiring a reduced cost from the network providers, which can be achieved through an efficient design of the mobile architecture.

In conclusion, from the perspective of mobile service platforms, the following key properties should be provided by carrier grade networks:

- Support for the application delivery, including seamless mobility support and resources at the required levels including throughput, delay and packet loss and the seamless mobility support.

- Support for the maintaining the communication active for the service delivery, including the core network signaling for supporting the often communication state transitions of the mobile devices, required for application paradigms such as peer-to-peer, cloud computing, and short message communication.

- Network level information sharing, including data path events related to active data sessions and context information such as location and presence.

- Support for service adaptation through providing information on the current network status of the devices.

- Core network support for a high number of devices i.e. connectivity support.
- Transparent delegation to the core network of application level functionality, reducing the processing and the information stored in the applications for each device.
- Reduced costs. A reduction in cost of the overall service delivery can be realized by optimizing functionality in the service platform or in the core network.

### 3.2.3 Network Platform

The technology evolution enables the network providers to offer IP connectivity to a larger number of subscribers. Additionally, more resources can be allocated for each mobile device, through this increasing the capabilities offered.

In order to be able to offer more capacity in term of connected device and available resources, network operators are currently deploying multiple access networks in the same locations, highly overlapping and enabling complimentary connectivity service with different characteristics specific to heterogeneous wireless technologies [82]. These heterogeneous access networks have to be supported in an uniform manner, including transparent mobility inside one or between multiple radio technologies. Additionally, the uniform network should consider the resources that have to be allocated on a target access in case of handover and a convergent mechanism for mobile device identification, data traffic accounting and charging [16].

With the increasing number of devices and data traffic, a novel level of efficient handling of IP connectivity is reached: the network has to scale as signalling and as data transport in order to enable the communication for all the mobile users. This may be reached only through the functionality distribution. Distribution requires new algorithms for flexible selection of access and core network data path entities able to ensure the reliability and the easiness of maintenance of the network.

Without any optimization, signalling is increasing at least linearly with the increase in number of connected devices. An increase with one order of magnitude of connected devices implies the same increase in number of functional components handling the devices inside the network. Therefore, it is utmost important that the communication between devices and network as well as the communication inside the network are reduced.

Additionally, the foreseen exponential data traffic increase, requires a reshaping of the data paths in the core network, closer to current fixed networks. However, a complete transfer to a fixed network infrastructure is not possible, mainly due to transparent mobility support requirement [124].

Furthermore, the data traffic exchanged by the mobile devices is based on IP technology. Because of this, in order to be more efficient in connectivity handling, the network has to natively support the IP communication paradigms, however adapted to the mobile communication [11].

The devices which are foreseen to be deployed in future carrier grade network infrastructures are highly differing as capabilities, including device interfaces which connect to the heterogeneous access networks and the processing and storage capabilities, overall requiring from the network a customized support for the communication [82].

Considering the increase in data traffic and in number of connected devices, the key property of the future core network is the capability of customizing signalling, data path and application delivery to the different requirements of different classes of devices. The customization presumes that for each device, network resources will be consumed to a minimal level in which the connectivity is not under the quality of experience level expected by the users. In the next section a set of requirements are derived from this key property required from the core network.

Additionally, with the more efficient handling of signaling and data traffic, it is expected that the communication infrastructure is able to sustain an increased number of devices without requiring the deployment of any extra functionality, through this reducing the cost per user both for the wireless system deployment and for the operations.

In conclusion from the perspective of network platform providers, the following key properties should be sustained:

- Support for heterogeneous access networks, enabling the communication of a larger number of connected devices, realized through supporting a large radio spectrum.

- Support for multi-homing. For an increased data rate of the same connected device multiple simultaneous accesses may be used. The network provider has to sustain transparently these multiple accesses.

- Distributed architecture. With the increase in number of devices, a distributed architecture is required enabling the flexible adaptation of communication, through this reducing the communication delay and localizing the data path. Additionally, it offers new possibility for easiness of maintenance and reliability.

- Reduced signalling per device. For accommodating a large number of connected devices, the network provider has to reduce the signalling with the mobile devices without implying a reduction in state information maintained for each device.

- Optimized data traffic handling. The mobile data traffic has to be transported through the access and the core network to the application platforms. With an exponential data traffic increase novel mechanisms similar to the IP technology should be considered to optimize the data forwarding.
- Reduced costs. A reduction in deployment and operational costs of the network can be obtained through all the other key properties.
- Customization. A key property of the future network infrastructure is the adaptation of the resources which are used for establishing the IP connectivity of the devices to a minimal level without deterring the quality of the services.

## 3.3 Carrier Grade Operator Requirements

In this section, a set of requirements for the carrier grade operator communication networks is introduced as derived from the previously presented key properties. The requirements are classified into four categories, one class addressing the complete architecture, the others depending on the logical part of the core network they are addressing: access network, core network and connection with the service platform.

### 3.3.1 General Requirements

The following requirements are addressing the overall wireless system from the perspective of the network operator which is providing heterogeneous access networks convergent through a single core network for a large number of subscribers.

In mobile networks, apart from the dynamic initiation and termination of the applications, three other factors are affecting and increasing the dynamics of the overall system. Firstly, due to the own mobility and to the mobility of the other devices, the network context in the vicinity of each connected node is continuously changing. Secondly, when multiple access networks of the same or different wireless technologies are available at a single location, then mobile devices have the possibility to execute a handover to another access network and through this to adapt to another network context. Thirdly, the mobile device communication can be done using multiple data paths through the core network with different momentary operational costs and with different end-to-end latency.

For maintaining the communication at the required level, the network has to be able to adapt the communication based on these factors for each subscribed device independently. The main requirement is the self-adaptation of the access, core and application delivery. A core network which is able to select the most appropriate access network and core network data path and the appropriate level of resources with which the applications are delivered is required in face of the high diversity of access and core network as well as applications foreseen in the future mobile networks.

Additionally, the core network has to be able to reduce the time required for the adaptation procedures as they involve the reservation of resources on both the initial and on the adapted states.

With the increase of the number of mobile devices, the carrier grade core network has to reduce the signalling allocated to each mobile device. This may include reduction in the signalling required for access network handovers, core network data path optimizations and in communication with the service platforms for service delivery.

Additionally, a reduction in number of policy based decisions executed by the core network for each device is required in order to serve more subscribers using the same network infrastructure. The number of decisions that are to be taken is highly influenced by the number of adaptations required and by the number of possible states in which the device connectivity may enter.

Furthermore, as the network has to maintain a complex state for each device for which an adaptation is executed including the initial and the final states, a reduction in the number of adaptations is desired for a more efficient usage of the network resources, especially by eliminating ping-pong effects.

Additionally to these requirements, some other are to be considered due to the high distribution of the access and core network components including flexibility in entity selection, easiness of maintenance of control and data path entities, and reliability of the communication. However, this dissertation concentrates on subscriber oriented self-adaptation and considers these features only from this perspective, therefore differentiating from the large research in management brought by the self-organizing networks paradigm.

As mobility is not native to data network infrastructures, it requires various modifications through additional functions and interfaces in the network apart from the basic data forwarding functionality. For a solution to be applicable in the context of IP networks, these modifications should not propagate towards other entities which are part of the wireless system. Therefore, any self-adaptation solution should have minimal implications on mobile devices while maintaining all the features transparent to correspondent nodes. However, this requirement should not limit the possible advantages that could be obtained by the communication of specific correspondent nodes with the service platform such as in the case when the mobile devices are exchanging data with service platforms.

As the proposed solution is designed to be applied to different network architectures, it must be easily applicable as an addition to the current network architectures such as the 3GPP Evolved Packet Core. The reduced integration complexity enables a solution to be easily accepted by the standardization community and adopted as part of the technology.

### 3.3.2 Access Network Related Requirements

The general requirements have a specific form when related to the access networks. Mainly, they relate to access network selection and handover procedures as the adaptation procedure. The access network selection enable the mobile devices to appropriately the select the target access when required. The handover procedure is active adaptation to the new access network executed when the selection happens.

For access network selection, the main requirements are related to the consideration as part of the access decision parameters the available resources in the vicinity of the mobile device, the required resources for the delivery of various applications and the subscriber profile information as part of the selection procedure. Additionally, any change in required resources and in subscription information should trigger an adaptation procedure in order to be able to efficiently support the new mobile device characteristics.

For the handover procedures a reduction in duration compared to the current procedures is required as well as a reduction in the number of operations that have to be executed.

All these operations have to be executed transparently to the correspondent nodes and with a minimal impact on the mobile devices.

### 3.3.3 Core Network Related Requirements

From core network perspective, the general requirements are related to the data path adaptation through the core network including the making of the decision and its execution through signaled orchestration between the different network components.

The decisions should be made considering the current status on the different network paths, the resource requirements of the applications and of the current service platforms used.

For data path adaptation procedure, the same requirements of reduced signalling and low delay are considered.

Additionally, the adaptation procedures should be contained in the core network, therefore to be transparent to mobile devices and to correspondent nodes. For procedures such as fail over mitigation, maintenance or energy consumption reduction procedures, the mobile device should be able to remain attached to the same access network.

### 3.3.4 Service Delivery Related Requirements

The service adaptation is related to the flexible change of the communication parameters between a mobile device and a service platform while the access network and the core network path are adapted.

The adaptation requires a reduced signalling between core network and service platform. Additionally, the core network has to take over functionality from the service platform which is already partially sustained such as mobility support and application resources requirements handling through policy based decisions.

### 3.3.5 Mobile Device Related Requirements

The realization of the requirements imposed on the operator networks including the radio access network, the core network and the service delivery platforms imply also modifications on the mobile devices. As the number of mobile devices is very large and distributed to their users, the modifications brought to the mobile device should be minimal.

The Subscriber Identity Module (SIM) and the mobile device firmware or operating system should not be modified. Additionally, modifications to the mobile device should easy be integrated in existing devices. Thus all the distinct modifications should be realized as applications which can be downloaded through software updates.

### 3.3.6 Requirements Summary

The requirements are summarized in Table 3.1.

## 3.4 Evaluation Criteria

For the evaluation of the related technologies and for the final evaluation of proposed system, a set of criteria are defined in this section. The criteria are selected based on the requirements previously described. This section introduces these criteria and discusses the evaluation metrics, enabling the framework for comparison of different technologies. The following criteria are not ordered by their importance towards the carrier grade operator network self adaptation.

1. **Resource based access network selection** - Solutions handling the selection of access networks have to make a policy decision triggered also based on the required and available resources.

| Requirement | Description |
|---|---|
| Adaptation Enabling | The network should provide the means to appropriately select the access network, the core data path and the level of resources required by the applications according to the network status, by this providing the best available connectivity for each subscribed device |
| Adaptation duration reduction | The network should reduce the duration of the adaptation procedures |
| Signaling reduction | The network should be able to reduce the signalling per subscriber |
| Policy Decisions Number reduction | The number of policy decisions for each subscriber should be limited to the a minimum which ensures the self-adaptation of the subscriber characteristics through the network |
| Adaptation procedure number reduction | The number of adaptation procedures should be reduced to the minimal level while maintaining the characteristics of the communication |
| Correspondent Node Transparency | The adaptation procedures should be completely transparent to correspondent nodes as to make them feasible to be deployed in the current network architecture. Containment to the operator network infrastructure also reduces the duration of the procedures |
| Minimal Mobile Device Modification | The modifications brought to mobile devices should be minimal and easily integrated into state of the art devices. |
| Service Platform Integration | Application delivery requirements should be part of the adaptation decision |
| Standard Architecture Applicability | The solution should be easy applicable as an extension to current network architectures such as and not restricted to the 3GPP EPC |
| Modularity | The solution should be composed of multiple enablers which may be used independently |

Table 3.1: Requirements Summary

This criterion provides a means evaluate how close to the ideal access network selection a decision may be, considering the required and the available resources.

2. **Data path stretch** - Solutions handling mobility have to maintain a mobile device reachable by its correspondent nodes despite its mobility. For this reasons, solutions may not always be able to maintain the best route between mobile devices and correspondent nodes.

   This criterion provides a means to evaluate path optimality. That is, how close to optimal routing is a solution, based on the theoretical measuring of the number of links between the mobile device and the correspondent nodes, pondered by the metrics allocated to these links

3. **Service adaptation support** - As the mobile communication environment

is highly changing, even when the devices do not change their network location, the service platforms require that their data delivery is adapted to the dynamic network status.

This criterion provides the means to evaluate how close to the optimal the service adaptation is, considering the available network resources and the possible session profiles negotiated between mobile devices and service platforms.

4. **Subscription profile and user preferences** - With mobile device diversification, access networks and services, the user has multiple choices for application delivery. The choice is done through the subscription profile stored in the network and through dynamic user preferences.

   This criterion evaluates if and how much a specific solution is considering the user related information into their decisions making for adapting of the subscriber communication. A solution is considered better if it uses all the information available in the mobile system.

5. **Required signaling** - Current solutions have to maintain a binding between the location of the mobile device and its identity to enable reachability and service delivery. For example, in case of access network and core network adaptation, this information has to be maintained in the core network while for the application adaptation, it should be contained in the service platforms. For this reason, adaptation triggers have to be signaled, information has to be collected and then the decisions have to be enforced. Different levels of signalling are considered for the communication inside the network domain, then for the case when correspondent nodes are involved.

   This criterion evaluates how many messages are required for executing an adaptation procedure. Additionally, if a procedure has to be executed multiple times until the system reaches a stable state, mainly due to lack of available information at the initial decision, the signalling is counted as part of the same overall adaptation procedure. A reduced signalling it is considered better as it requires less functionality in the network and a lower delay.

6. **Required policy decisions** - Current solutions require a policy decision for each adaptation. For example, in case of access network reselection, a decision is made for handover and another one for adaptation of the service. This criterion evaluates the number of policy decisions taken during an adaptation. Considering the same final result, a reduced number of decisions is considered better as it requires less functionality in the network per device.

7. **Adaptation duration** - During any adaptation procedure there is a certain critical time interval in which a complex state is maintained containing both the initial and the target information. This state has a high resource consumption and possibly impacts the communication of the subscribers e.g. through packet loss, jitter etc. The adaptation duration depends on signalling delay and on the network processing. A shorter adaptation time should be considered.

   This criterion measures the time difference between the moment an adaptation is requested and the moment when it is completed as noticed from the data path. Some control procedures may happen afterwards, however their delay will not affect directly the data communication.

8. **Solution flexibility** - this criterion evaluates the capability of a solution to administratively power down one or more network entities. It enable to evaluate the easiness of network topology change due to maintenance or energy efficiency reasons. A solution is evaluated on its capability to adapt for each of the affected users without service interruption. A better solution is the one which can flexibly select and replace more functional entity serving a mobile nodes.

9. **Reliability** - This criterion evaluates the ability of a solution to perform its functionality even in exceptional cases in which a part of the system is not functional. A system is considered better when more components can be replaced seamlessly to subscribers at any failure case.

10. **Impact on mobile devices** - This criterion evaluates how much functionality, if any, is required to be introduced in the mobile nodes, supporting the features of the network solution. A minimal modification is considered the introduction of a software component, while an extended modification is considered a modification of the TCP/IP stack.

    Additionally, as part of the impact on mobile devices is considered the support of the multi-homing and for seamless communication.

11. **Impact on the correspondent nodes** - This criterion evaluates whether the solution requires correspondent node modifications. The correspondent node is generically representing all the other possible remote parties including mobile devices, fixed devices and service platforms which in a great majority require complete transparency to a newly deployed solution. From this perspective, a solution which is transparent to the correspondent nodes is considered better.

12. **Impact on the service platforms** - This criterion evaluates whether a solution is able to reduce the service platform functionality by taking over functionality, reduce the signalling during adaptation procedures, and by maintaining the changes transparently.

    Additionally, the system should allow the service platform to transmit their requirements to the core network and to receive the minimum number of messages in form of events to allow service adaptation. From the application perspective, a better solution is the one takes over part of the functionality while reducing the signaling.

13. **Applicability** - This criterion evaluates the easiness of integrating a specific solution in the current network architectures. A solution which is already integrated or is easy to be added in the current running carrier grade systems is considered better as it can be easily deployed in the network. However, this criterion should not be considered as decisive for the scientific value of a specific solution.

As a complete orthogonality between the evaluation criteria was not possible to be achieved, due to high interconnection between different features, they should not be considered equal as importance. However, even having the equality perception pit-fall, a large set of criteria was preferred for achieving a complex image of the network environment.

Some of these criteria are logically opposing each other. For example, with a high increase in signalling and with virtually an infinite adaptation delay, any adaptation procedure will produce close to perfect results. For solution feasibility reasons, these opposing criteria were desired. Through good innovation and design, it is expected that solutions will have better performance even on opposing criteria, resulting in progress beyond the current technology state.

In the following chapter the various state of the art and related solutions are evaluated against these criteria. After presenting the proposed solution, in the critical assessment chapter, these criteria are used to further demonstrate its advantages.

In order to be able to offer a homogeneous comparison between heterogeneous technologies, to each evaluation criterion, a metric was associated. The scale for the metrics is detailed in Annex A.2.1.

The evaluation criteria are summarized in Table 3.2.

| **Evaluation Criteria** | **Description** |
|---|---|
| Resource based access network selection | Evaluates the usage of resource related information in access network selection decisions |
| Data path stretch | Evaluates how close to the optimal routing is a solution |
| Service adaptation support | Evaluates how close to the optimal the service adaptation is |
| Subscription profile and user preferences | Evaluates how much information related to the subscriber is used |
| Required signaling | How many messages are required for executing and adaptation procedure |
| Required policy decisions | Number of policy decisions that have to be taken until the adaptation procedure succeeds |
| Adaptation duration | The time difference between the adaptation trigger and the completed procedure observed on the data path |
| Service Platform Integration | Application delivery requirements should be part of the adaptation decision |
| Solution flexibility | Capability of a solution to adapt to network topology changes |
| Reliability | Capability of a solution to adapt to fail-over cases |
| Impact on mobile devices | How much functionality, if any, has to be added to the mobile devices |
| Impact on the correspondent nodes | How much functionality, if any, has to be added to the correspondent nodes |
| Impact on the service platforms | How much functionality, if any, has to be added to the service platforms |
| Applicability | How easy is to introduce the solution in the current core network architecture |

Table 3.2: Evaluation Criteria Summary

Chapter 4

# Related Technologies

Various research efforts tackle the main features of the subscriber based adaptation according to the requirements from the previous section chapter. The following sections present several research activities on which this thesis is built or which relate to the work presented.

## 4.1 Access Network Selection Technologies

With the deployment of multiple access networks of various technologies, the telecommunication environment is evolving towards a dense wireless environment where multiple access networks overlap and complement each other in terms of bandwidth, transmission delay and operational costs. It became obvious that the access network selection is transforming from a trivial problem on selecting the only possible access network to an extended issue in which a high number of access networks of various access technologies are available at the same location. This section presents initial approaches and solutions for access network adaptation aiming at seamless and optimized service continuity.

### 4.1.1 Media Independent Handover

The IEEE Media Independent Handover 802.21 standard (MIH) [125] represents a solution for fast network detection and selection, integrated with the IP connectivity level, enabling fast handover procedures. MIH concentrates on vertical handovers, by offering convergent triggering, control, and information exchange pertaining to each access network. For this, it offers mechanisms for network discovery, network selection and handover negotiation, named handover initiation. Also, it offers optimizations for the handover preparation phase including the physical, link and IP layers.

As depicted in Figure 4.1, MIH is concerned with the optimization of the handover provisioning in the area between access network link layer connection technologies and network layer. The key component is the MIH Function MIHF offering three services: event triggering, control messages and information service.

The Media Independent Event Service (MIES) offers information on the physical link to the upper layers, aiming at a reduced disruption during the handover

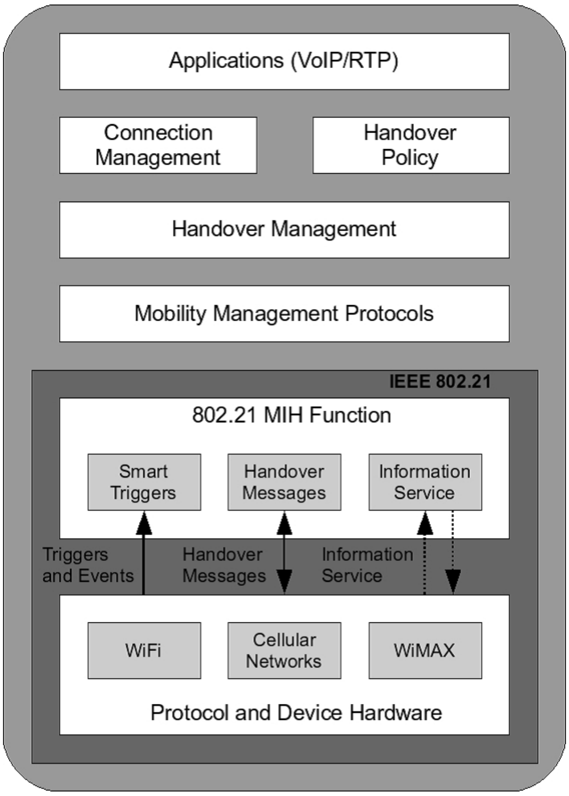


Figure 4.1: Media Independent Handover Stack

by announcing changes in access network availability, including loss of connectivity predictive triggers. MIES enables the mobile device to convey link layer triggers into media independent format which are transmitted to higher levels of the network stack or to remote network entities.

The Media Independent Control Service (MICS) enables the transmission of handover related commands to the link layer of the mobile devices.

The Media Independent Information Service (MIIS) offers the necessary information for access network selection. It is usually located on the network side. Based on the mobile device characteristics, it can determine an ordered list of available networks. The parameters transmitted to the device are mainly static parameters, enabling a fast processing of the requests. However, this implies that the actual presence of the network in the specific location can not be determined i.e. the network might not be available due to energy saving, fail-over or maintenance mechanisms, etc.

MIH standard was developed for ensuring fast connectivity continuity with no consideration for seamless services handover. Having a static MIIS, MIH is not aware of the momentary status of the access networks, not guaranteeing the connectivity and the preservation of resource reservations.

MIH aims only at handovers due to loss of connectivity over source access network, therefore, enabling a fast communication over the network in the time interval when a handover has to be executed, without supporting out of band information exchange. Further optimizations in this direction and in the application of the technology for different access network are available in the literature [126], [127], [128], [129], [130], [131] however maintaining the same time constraints limitations.

Because of this, MIH was initially standardized as a link layer communication protocol. The network stack of the mobile devices and of the base stations of different radio technologies had to be modified.

MIH conceptual approach was extended through multiple mechanisms. In a notable approach, from EFIPSANS project, the information transmitted considered also the type of connectivity the subscriber requires and the current available resources in the specific access network [132]. Although being able to balance the subscribers between the different access networks depending on network load, the operations are executed only at the attachment of new devices.

In order to be able to communicate over multiple device interfaces at the same time, the EU Ambient Gateways [133], [134], [135] project extended MIH layer with data routing capabilities over multiple radio links. This intermediary Generic Link Layer controls the availability of specific links, while for the data transmission it makes routing decisions in the mobile device selecting the most appropriate interface.

### 4.1.2 IETF

IETF tackled the issue of access network selection from a network oriented perspective. Although recognizing the role of the availability of the mobile device required resources in the target access network for a seamless service continuity and thus, the major role of the communication between selection functionality and network provider, IETF solution concentrated on the network level missing features in the IEEE MIH standard, mainly in the direction of the transport level protocols [136].

A novel abstraction level was defined on top of the network layer, which enables the exchange of information between any entities pertaining to access network selection services. In this framework two main functions were considered: the service discovery and the service transport [137].

For the service discovery, two complimentary mechanisms were considered. One uses DHCP for transmitting the address of the network mobility service to mobile devices [138]. The other uses DNS for service discovery in the specific access network [139]. In both cases, a standard protocol is extended through known means to support mobility service discovery.

A basic transport mechanism was developed in which the data is encapsulated into UDP messages [140]. However, a more general protocol was preferred [136] being able to support both TCP and UDP transport. Both protocols are data carriers without having support for mobile device authentication and authorization or for encryption. These mechanisms enable differentiation between the messages that have high time constraints and those which contain large amount of information, without standardizing a specific mechanism for either.

Another solution was proposed for the transport of the MIH messages in the operator core network using the IP Multimedia Subsystem (IMS) signalling mechanism [141]. Although this solution enables the exchange of data packets with application servers especially designed for service continuity, it remains tributary, as any transport solution to the selection state machines at the end nodes, thus, to MIH limitations.

### 4.1.3 ITU-T

The ITU Telecommunication Standard Group on Future Networks, proposed a high level architecture for the mobility support and control in the Next Generation Networks [97], which is representing a high level view of carrier grade operator networks, similar to the 3GPP Evolved Packet Core architecture, having a Resource Admission and Control Function (RACF) mapped to the 3GPP PCRF and a Handover Decision Function (HDF) similar with the 3GPP ANDSF, due to its reselection functionality and due to its ability to receive information for discovery procedures from a Mobility Location Management Function (MLMF). A simplified architecture from the access network selection perspective is depicted in Figure 4.2.

In ITU-T architecture, the access network reselection decisions are taken based on user profile information and on the serving link events such as reduction or loss of signal the availability of the required resources on the target access network. Although, during the handover decision the level of available resources is requested from the RACF function, the RACF is not able to initiate handover procedures. Therefore, when the mobile device required resources are changing, this is not considered as reselection trigger.

However, as the ITU standard is regarded as recommendations for further standardization in other organizations, such as 3GPP, it do not detail how this procedure should be executed. Therefore, the concepts presented in this dissertation in relationship to access network adaptation are easy to be integrated as additions also to ITU architecture.

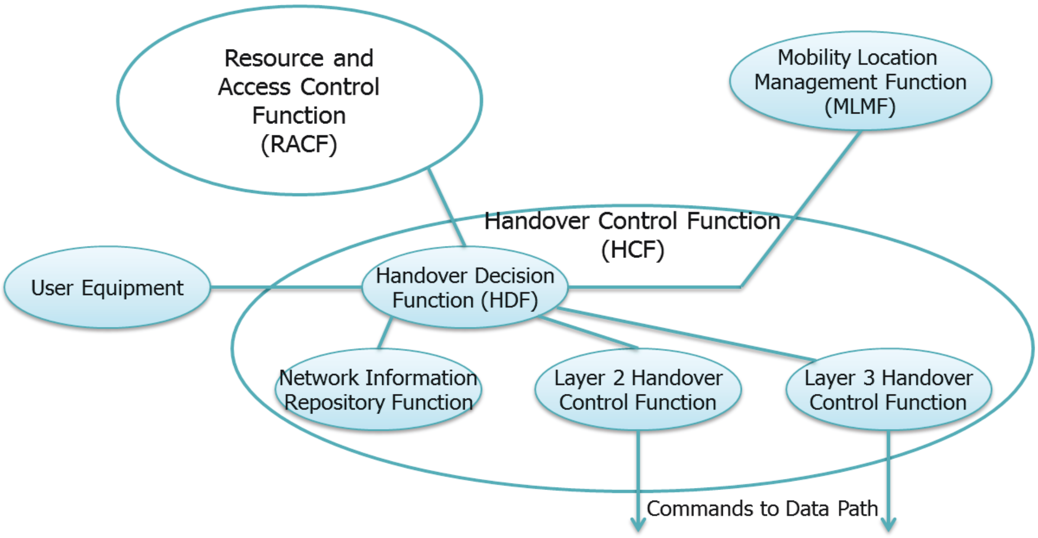


Figure 4.2: ITU-T Access Network Discovery and Selection High Level Architecture

### 4.1.4 IEEE P1900.4

The IEEE P1900.4 work group is developing the architectural building blocks enabling network-device distributed decision making for optimized radio resource usage in heterogeneous wireless access networks. The standard is part of the cognitive radio activities of the IEEE.

P1900.4 protocol [142] defines a set of building blocks in the mobile devices and in the network, handling the resource management and the information that has to be exchanged between these functional entities, aiming at a coordinated decision for an optimized radio resource usage.

The goal is to improve through cognitive radio mechanisms the overall composite capacity of heterogeneous wireless systems by defining appropriate protocols facilitating the usage of radio resources, in particular the information exchanged between network and mobile devices addressing single- and multi-homed devices.

From the perspective of access network selection in carrier grade operator networks, the P1900.4 functional architecture is offering an extensive functionality in the area of reselection of physical resources, on assigned spectrum, on momentary resources required by the mobile device, and on the requirements of the subscribed services [143].

However, the standard still requires further advancement and research, especially in the area of resource allocation orchestration at the overall system level and at the integration while legacy mobile devices. Missing extensions include the common managing of decisions taken in specific network locations and the integration with subscriber oriented policy decisions, allowing the differentiation between the subscribers and their required resources.

P1400.4 solution was extended through projects such as EC FP6 E3, which concentrated the most on autonomous spectrum allocation. In the area of access network selection, the protocol was adapted having similar features as the MIH [144], [145]: only link level protocol communication and no relationship with subscriber communication.

### 4.1.5 3GPP Access Network Discovery and Selection

The access network selection functionality in the 3GPP Evolved Packet Core is differentiated based on the access networks between which the handovers are assumed to be executed. One mechanism is presumed for handovers between 3GPP access networks and another one when the handovers involve at least one non-3GPP access network. In this section, the two mechanisms are described.

#### 4.1.5.1 Intra-3GPP Access Network Selection

The handover between 3GPP access networks is controlled by network management entities which are aware of the mobile device location and of its reserved resources, of the access network coverage, and of resource availability. For handovers between GPRS or UMTS accesses, SGSN is supporting this functionality [78], while for the handovers to LTE accesses, MME is the control entity [24]. As the concept which is followed is similar for both these entities, in the following only the MME will be further analyzed.

The MME is controlling the intra-3GPP mobility and maintains the current location of the UE, including the tracking area information, the LTE cell identity, the age since this identity was acquired, the authentication information, and the UE access capabilities. Its main functions include authentication and authorization of UEs, location tracking, and network assisted cell change [103], [101].

For location functionality, MME maintains information on the cells located in UE's vicinity named Tracking Area. This functionality along with the cell identity information with handover requirements of the base stations due to loss of signal and with the network topology in the area controlled by the MME, enable the MME to make decisions and to prepare and orchestrate a handover.

In order to reduce the interruption when executing a cell change between the different 3GPP access networks, MME provides proactive handover information to the target cell, enabling a preparation phase in which the data packets are forwarded between the source and the target access networks. For this, MME has to first discover the appropriate Serving GW, eNodeB or SGSN to which the handover will be performed.

In the MME solution, control entities are managing the complete handover procedures on behalf of the UE, providing a means specific to 3GPP accesses, in which the network controls the access network selection. Even though it maintains high level of information, the system is rather static in regard to UE's required resources, reacting only to radio signal or subscription profile modifications.

Additionally, as the MME is able to command handovers, the system is ensured that a handover to the target access network will be executed and thus that no redundant resources are consumed. This deters the applicability of the same mechanism for non-3GPP accesses,where UEs connect with a different device interface based on the preference of the user.

#### 4.1.5.2 Non-3GPP Access Network Selection

In 3GPP EPC, the access network selection for non-3GPP accesses is based on the ANDSF which transmits indications on the access network located in the vicinity of the UE [21], [102]. The UE uses this information to discover access networks and then to select the most appropriate one.

At the moment of writing of this dissertation, the ANDSF functionality offers to the UE only guidance in selecting a next point of attachment. It does not imply that the UE prior to the attachment procedures has to transmit to the ANDSF neither its location nor the networks located in its vicinity. For this, the ANDSF is aware of UE's identity and of a coverage map which contains all accesses deployed by the specific carrier grade operator.

By knowing the networks available for communication in the area where the UE is located and based on the selection policies received from the ANDSF, the UE selects one access. For non-3GPP accesses, EPC leaves much of the control to the UEs not being able to dynamically balance the UEs between different accesses based on their required resources and the momentary network availability. Therefore, the UE is selecting to which access network to connect to according to the policies, which selection can be observed in EPC only when attachment procedures are initiated [105].

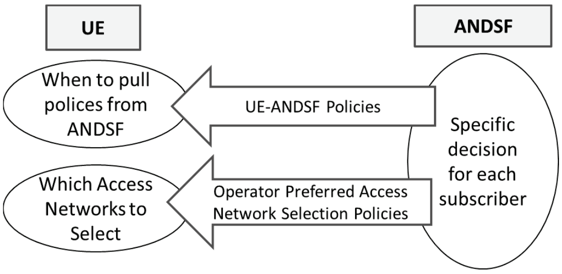


Figure 4.3: ANDSF Policies

Two types of policies are to be exchanged between UE and ANDSF, differentiated by their conditions and their actions, as depicted in Figure 4.3:

- *UE-ANDSF policies* - these policies refer to the communication between UE and ANDSF. Their conditions mention when and how the communication between ANDSF and UE has to take place. This type of policies is implicit, being separated into ANDSF discovery and UE registration for service (for both PUSH and PULL communication models).
- *Operator Preferred Accesses Policies* - these policies include the main functionality of ANDSF. The policies state in which conditions a UE has to initiate the discovery procedures for a new attachment and which access network to select from the list of discovered networks.

ANDSF has only one interface, defined for the communication with the UE. Using this interface, the UE initiates the communication with the ANDSF for operator preferred access network discovery when some specific conditions set up within the policies available in the UE are met or when the user requests a manual selection.

The communication between UE and ANDSF can be executed in PUSH, PULL or both modes. For PULL mode, UE has to initiate the communication, thus the UE-ANDSF policies are triggering the exchange. For PUSH mode, the network has more control of UEs, being able to update the policies based on network events. Similar to PULL mode, PUSH mode bases also the Operator Preferred Accesses Policies exchange on the UE-ANDSF policies, because the UE has to initially discover the ANDSF and register to it.

For the case of switch-on or recovery from lack of signal, the UE has to attach to a first access network based only on stored policies [102]. In this case, the UE has to be able to select first the technology over which to communicate, the necessary credentials and to execute the attachment procedures. In order to be able to receive these policies, UE has to attach and to be authenticated and authorized at ANDSF. The initial attachment to ANDSF may provide a solution to this problem. However, from the perspective of the 3GPP standards this remains as an open issue.

By this mechanism, the operator has a very loose control of the next point of attachment to the access networks, as the UE does the actual selection of the next access to attach to.

3GPP underlined the importance of location for selecting the appropriate Operator Preferred Access Policies. As stated before, the policies are transmitted to the UE prior to the technology specific discovery procedures, thus the ANDSF is not aware of the exact access networks located in UE vicinity, having to rely on subscription profile and on the Coverage Map information. A more granular location information would provide a more suitable set of policies, accustomed to UE's context.

At his moment, in the 3GPP standard, it is uncertain which information is kept in the ANDSF for each UE. Thus, it is not clear whether ANDSF maintains the full location and bearer information e.g. accesses used, momentary GBR, momentary MBR etc. or whether, it maintains only static information that may apply to multiple UEs at the same time. The full state information enables the ANDSF to track the UE attachments and to avoid the ping-pong effect. Also it can command handover procedures with optimization, as being able to determine the next point of attachment.

The full state can be maintained at a specific processing cost. When ANDSF maintains only static information, more users can be served using the same capacity. A trade-off between these solutions may be proven to be more suitable for ANDSF, as other entities in the EPC maintain already information that can be used for network discovery and selection process e.g. MME maintains a tracking area for 3GPP accesses.

3GPP desires for PUSH mode to be able to register the UE to ANDSF. Until now, no specification of the procedures has been defined. These procedures presume that ANDSF maintains at least registration state for each UE. It is not yet considered whether this state information contains access network and location information. Also, the mechanism for transmitting policies from ANDSF to UEs is not defined i.e. when, why and how the policies are transmitted to the UE. It is hinted that ANDSF may transmit this information either at specific time intervals or when the Operator Preferred Accesses Policies change.

For PULL mode, the UE transmits its current location and its capabilities e.g. the list of access networks that can be selected. It is not yet clear whether the UE authenticates or whether it receives the selection of the access networks independently of its identity.

As stated in 3GPP standard [102], the ANDSF provides inter-system mobility policies: *"The policies may indicate preference of one access network over another or may restrict inter-system mobility to a particular access network under certain conditions. The ANDSF may also specify validity conditions which indicate when a policy is valid. Such conditions may be based on time duration, location etc."* This definition of the policies is very vague. It can be interpreted as general policies based on location and time periods (e.g. from 10:00 to 12:00). The UE has to execute by itself the attachment procedures accordingly.

In conclusion, ANDSF offers the support for transmitting operator preferences to UEs for network discovery and selection. Although the 3GPP specification introduces the basic concepts and mechanisms necessary, multiple issues still remain open. At this moment, it is not clearly stated the relationship between ANDSF and other network entities, nor its interfaces. Because of lack of features, the access network selection is executing without any relationship to available or required resources, therefore not ensuring the seamless service continuity.

A single interface was defined between the ANDSF and the UE, but this interface lacks the exact technical requirements. For example, it is considered that this interface is able to transmit to UEs the operator preferences, but it is not clear when these preferences are transmitted, whether EPC has any procedures to execute when these policies are transmitted or the impact on the UE.

Through the network assisted mechanism of handovers to non-3GPP accesses, the operator is not aware of the UE's next point of attachment, thus the core network will have to react to the new attachment and determine that a handover occurred. The lack of certainty limits the possibilities of executing preparation phase procedures, thus maintaining a long handover duration which may lead to service interruption.

A similar solution was developed through the European Project EFIPSANS addressing plain IP networks. For access network selection and for routing of data traffic to specific interfaces, an autonomous decision is taken independent of the access networks at the IP level [146], however concentrating on the algorithms for directing specific data flows and not optimizing the communication mechanisms. The work remained in an initial phase without locating the decision points in the network context.

### 4.1.6 Discussion

In this section a set of access network selection solutions were presented. Following, these solutions are compared based on the evaluation criteria which leads to a comprehensive overview of the solved problems, but also to a better understanding of the open issues.

From the perspective of the *Resource Based Access Network Selection*, the analyzed solutions are concentrating on the access network change due to signal strength reduction. An initial concept, in which the target access network has to sustain the mobile device resources appears in the ITU architecture without further detailing. None of the solutions provides the means to execute handovers triggered by modifications in the resources required by the devices or because of the modification of the available resources in the current access network, features which would provide a more flexible adaptation of the communication path to the overall resource needs of the wireless system.

Additionally, none of the solutions considers as part of the network decision the *Subscription Profile and User Preferences*, information which is available in the core network or which can be received dynamically from the mobile devices. Due to this, the network assistance in the selection procedure is limited to static information in relationship to the devices.

From the perspective of *required signalling*, the solutions can be classified as solution for which all the signalling has to occur after the trigger of the handover was initiated such as the MIH or the intra-3GPP mobility support and solutions in which the signalling is not bound to any handover procedure such as ANDSF. In the first case, the signalling is made of a reduced set of information which is transmitted from the network, requiring to fit into a specific delay, but providing the information which enables an appropriate selection decision. In the second case, the signalling can be executed when the network cost is reduced, however not providing the certainty of selection procedures success, due to the lack of momentary network context.

For all the solutions described, a single informed access network selection is not possible, the *adaptation duration* may consist of multiple access network reselection cycles in which the mobile device attaches to different accesses in its vicinity until it is able to connect to the appropriate one. It is assumed that a decision which has all the required parameters is able to drastically reduce the adaptation duration to a single access reselection procedure.

As procedures can not be triggered by the modifications of the resources available in the specific access networks, except for the intra-3GPP access selection one, the current solutions have a minimal *flexibility* in regard to energy efficiency optimizations, fail-over or maintenance procedures.

All the solutions for access network selection require some *modifications of the mobile node* while remaining transparent to correspondent nodes. However, some solutions do not require the modification of the network stack, as they can either be deployed as applications or are part from the initial stages of the access technology specific device drivers. Other solutions such as MIH or P1400.4 require modifications of the network stack of the devices which implies higher deployment costs.

In conclusion, from the solutions analyzed, currently none is able to trigger the access network selection based on the modifications of the required resources of the device or on the changes in momentarily available resources in the network. Based on this, having a minimal flexibility and an increased signalling and adaptation duration.

The non-3GPP access network selection was selected as basis for the exemplification and the evaluation of the concepts proposed here, as it provides the most part of the evaluated qualities and has a high deployment probability. However, this does not imply that the novel concepts can not be applied to other solutions.

## 4.2 Mobility Support Technologies

Network support for mobility has been a very active topic of research for more than a decade, leading to a high number of available solutions. In this section some of the most widely deployed or highly academic considered solutions are analyzed based on a reference scenario illustrated in Figure 4.4. The scenario includes a mobile node (MN) which changes its location between two access points (AP1 and AP2) of the same or different access technology pertaining to the same operator and communicates with a generic correspondent node (CN). The technologies are analyzed from the perspective of obtaining an optimized path between the MN and the CN through a flexible solution while maintaining a minimal impact on the end nodes.

The access network handover was selected as being more intuitive. However, the considered scenario addresses also the case the MN maintains its connectivity to the same access network and the data path through the core network has to be adapted.

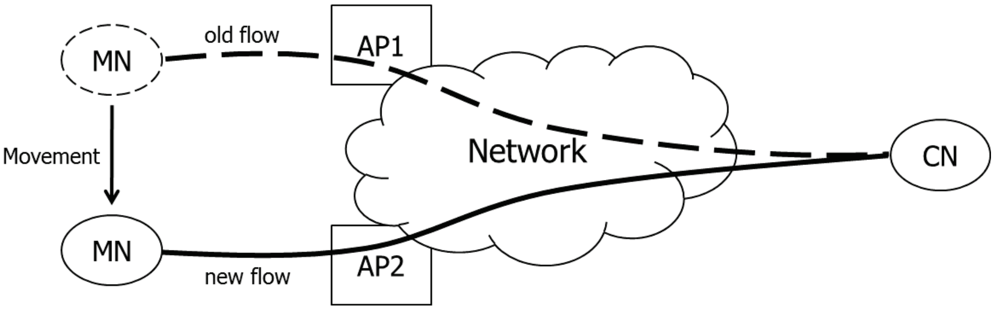


Figure 4.4: Core Network Mobility Support Evaluation Model

### 4.2.1 Mobile IP

Mobile IP (MIP) [147] enables the mobility support over multiple network provider domains transparent to the CN by using an IP layer of redirection. Figure 4.5 illustrates the key aspects of the solution [148], [149]. MIP depends on a globally unique identifier, the Home Address (HoA), which is allocated to each MN by a mobile network operator Home Agent (HA) entity. HA binds the HoA with the current locator of the MN, namely the Care-of-Address (CoA), which is provided by each of the access network (AP1/2) upon connectivity establishment (Step 1). In case of changes in the locator (CoA), MN must inform HA so that the HoA-CoA mapping remains always up-to-date (Step 2). CNs use the HoA as destination IP address for all communication with the MN. Packets destined to MN are first routed to HA, which tunnels them to the current MN CoA (Step 3). At the MN, the MIP implementation decapsulates the received packets and delivers them to the applications containing as destination IP the HoA. Correspondingly, when the MN sends packets to CN, it uses HoA as the source IP address.

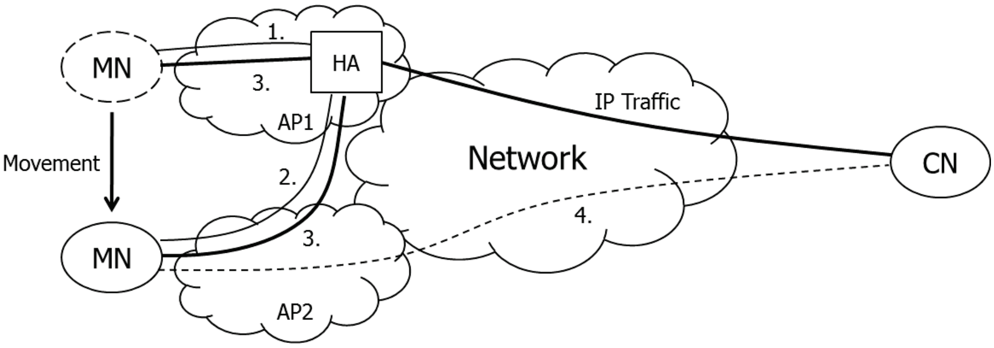


Figure 4.5: Mobile IP

MIP requires no changes to applications, and is not tied to any underlying radio access technology. MIP requires the operator to deploy one or more HAs and all MNs to add MIP support in the operating system. Recent standard updates enable the operator to select which HA handles each MN based on the subscription profile, allowing for more dynamic load balancing in MM operation, however maintaining the same HA for the complete communication of the MN.

If CN supports Mobile IP version 6 (MIPv6), then the route optimization feature can be used, in which case MN signals directly to CN changes in CoA (Step 4). In this case, the CN MIPv6 implementation maintains the HoA-CoA binding, and thus becomes the anchor for the data traffic, which enables optimal routes to be selected also for downlink data packets. However, this implies that all the CNs have to be modified in order to support this feature.

Finally, with respect to multi access support, several solution define MIP support for multiple CoAs, enabling MNs to control which CoA to use for different data flows or automatic using transport layer probing [150]. The specific solution was implemented and further extended through the FP7 EFIPSANS project [151], [152], [153]. Although a higher agility is given to the protocol, it addresses only the case of switching between different connections at network level, thus not enabling a complete data path modification, remaining tributary to the limitations of Mobile IP.

Other research follows the direction of multiple HoA-CoA bindings maintained by the HA [154], [155] enabling a load balancing for multi-homed MNs.

### 4.2.2 3GPP

3GPP has defined two modes of operations for MNs: the active mode in which MNs are connected to the network and able to receive and to transmit data and the idle mode in which MNs which do not require to communicate, even though connected, do not have an established data path and thus consuming less energy [19]. For these two modes different mobility support is offered by the network [78], [24], [101]. In the following only the LTE modes are presented. The 2G and 3G mechanisms are similar.

#### 4.2.2.1 Active Mode

For mobility support in Active State, 3GPP includes a set of solutions for different access technologies core networks, from which only the EPC solution is presented in detail, as it contains all the functionality of the previous optimizations while being the state of the art in standardization of new features.

For active mode, a permanent association is established through a chain of tunnels between the MN and the central anchor, the PDN GW. The tunnels are established using protocols such as GPRS Tunneling Protocol, Proxy Mobile IP (PMIP) or Dual Stack Mobile IP (DSMIP) which are using the same establishing and forwarding context, however, designed with different goals [105], [156], [114], [113].

MNs pertaining to one IP domain receive their IP address from one PDN-GW, selected during network attachment, which makes the IP address space reachable from the Internet. When PDN GW changes a service interruption will be experienced, as a new IP address will have to be allocated by the target PDN GW and this address has to be signaled to the CNs. The selection of the PDN GW is executed by the MME or the SGSN in case of the 3GPP accesses or by the access network specific gateways otherwise.

In all 3GPP mobility support solutions, an intermediary access network specific gateway is introduced to handle mobility inside one part of the wireless domain such as the Serving GW (S-GW) for 3GPP accesses, generic access network gateway (ANGw) and evolved Packet Data Gateway (ePDG) for trusted and untrusted non-3GPP accesses. As part of its functionality, the gateway handles the redirection of the connection to the new location of the MN while keeping it transparent to the central anchor.

Due to mobility, the access network specific gateway may change. In this case, the mapping in the PDN-GW is updated as to redirect the tunnel towards the target gateway.

The model adapted for LTE access is depicted in Figure 4.6. the MME takes the decision that a new gateway has to be selected for the MN. It informs the target S-

GW, which establishes a redirection tunnel for the MN data traffic from the source S-GW. Upon actual execution of the S-GW change, the target S-GW initiates the redirection of the tunnel from the PDN-GW. Due to the redirection tunnel, no packet will be lost during the S-GW change. The redirection at the PDN GW is established only after the MN is able to exchange messages over the target S-GW which enables the downlink data traffic to continue through the source gateway until the connection to the target gateway is completely established.

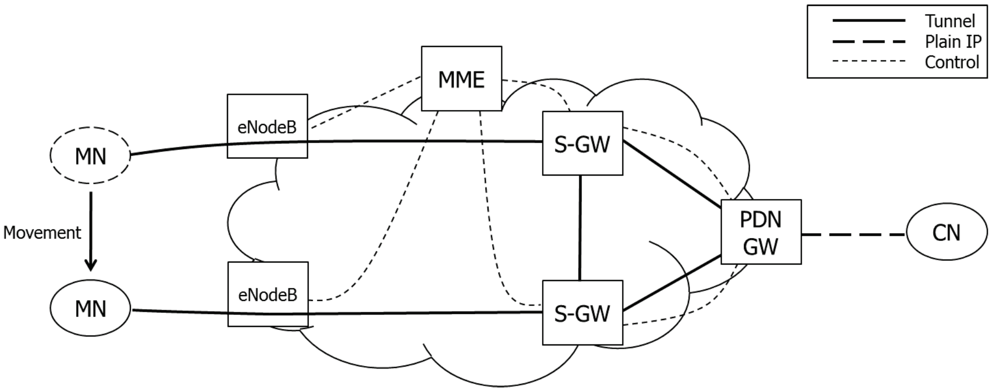


Figure 4.6: 3GPP Core Network Mobility Support

A same procedure is deployed also when a handover between two eNodeBs has to be executed in which the anchor is maintained in the S-GW while the redirection tunnel is established between the target and the source eNodeB.

The redirection is possible due to the control of the handover procedures in the MME. However, this is not realizable when the handover includes one non-3GPP access technology. Therefore using the current standard solution a service interruption will be experienced.

A similar solution based only on Proxy Mobile IP is considered by the Daidalos II project [157] in which the mobility in a single domain is realized through the anchoring in a PMIP Local Mobility Anchor (LMA) ensuring the seamless local mobility support. However, an application level solution is integrated on top, which is further presented in the next sections, enabling the seamless mobility support over multiple accesses.

#### 4.2.2.2 Idle Mode

Devices which connect to 3GPP access networks can enter into Idle Mode state in which no messages are exchanged with any CN, although various registration information is maintained in the network [24], [101].

The MN is tracked through messages transmitted to the base stations of the

3GPP accesses (i.e. eNodeB) resulting in a tracking area which contains multiple adjacent cells. When downlink data traffic becomes available, then the mobile device is announced in all the cells of the tracking area through a paging procedure.

Specifically in 3GPP accesses, for enabling reachability of mobile nodes in Idle State, a Tracking/Routing Area is defined containing the possible entities (e.g. eNodeBs) which might reach the MN and trigger it to become active. The Tracking/Routing Area is stored in the Home Subscriber Server (HSS) database for the subscriber identity. The MN is required to notify the network in case of leaving a Tracking/Routing Area.

When a CN initiates communication with the MN, the packets are received by the PDN GW/S-GW having as destination the IP address allocated to the MN. The PDN GW/S-GW sends a Downlink Data Notification to the MME and/or SGSN which forward it as a Paging Message to the nodes which may trigger the activation of the MN. IP packets are buffered at the PDN GW/S-GW until the MN becomes active and then are forwarded to it. For seamless service continuity this IP address has to be maintained during the new active state of the MN.

If the MN requires transmitting an IP packet to the network, then it changes to Active State, enabling it to communicate to CNs transparently over the previous and actual active state. The IP address allocated to the MN may be replaced with a new one, which enables it to be associated with another gateway, only if the MN has no CNs.

### 4.2.3 Host Identity Protocol

The Host Identity Protocol (HIP) [158], [159] [160] introduces the concept of separating the identity of a host from its location in the network. Moreover, a host can have multiple host identities [161]. In the TCP/IP stack, the IP-address serves also as host identity, which introduces problems for MNs if their IP address changes.

HIP is a new network layer between the transport and IP layer. More formally, HIP introduces a Host Identifier (HI) and a Host Identifier Tag (HIT) which is a 128 bit hash of the HI. TCP connections bind to a HIT instead of an IP-address which enables transport session continuity as long as changes of the IP-address are updated between the peers.

Other solutions consider that the IP address will be split into network locator and node identifier requiring similar architecture and similar interactions [162], [163].

HIP introduces a new node in the network architecture [164], the registration and rendezvous server (RVS), which serves as the initial point of contact for new HIP associations [165], [166]. Cryptographic material is exchanged for creating a secure association between the peers [167].

The MN registers with the RVS before it stores its HIP related information in the DNS like HI, HIT and RVS domain name. The CN performs a lookup for the MN's domain name and receives the HI, HIT and domain name of the MN's RVS [168]. The basic procedure is depicted in Figure 4.7.

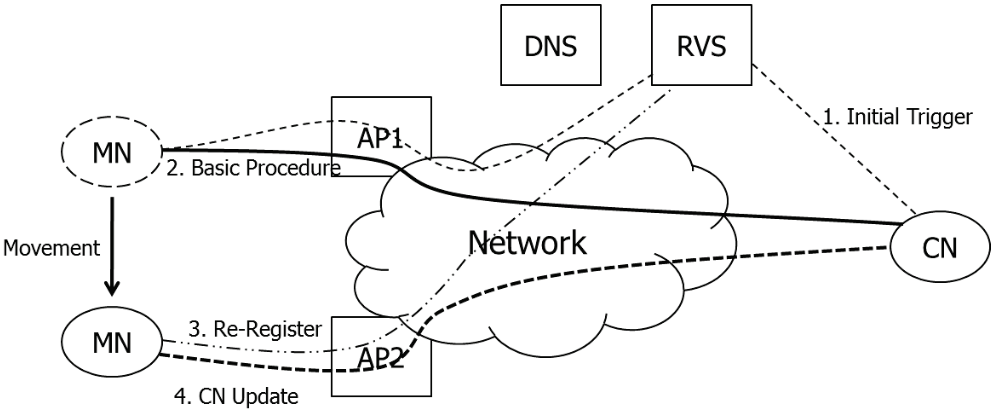


Figure 4.7: Host Identity Protocol (HIP)

The CN will contact the RVS with the initial trigger message, which is relayed to the MN (Step 1). Then the normal basic procedure is applied to establish a HIP association, which consists of a puzzle-solution handshake and the exchange of Diffie-Hellman parameters for establishing a common secret (Step2). Now transport sessions can be created and data transfers can take place. When an IP-address change occurs, the RVS server will be notified (Step 3) and the CN will be updated using a 3-way message exchange(Step 4). Deployed applications do not need to be updated, they can perceive the HIT as IP address. During mobility, transport connections will not break as they are bound to the HIT. Thus, after a short delay for the update, all sessions continue seamlessly.

From the perspective of the communication, HIP represents a high evolutionary step over the Mobile IP route optimization solution especially in the support of security association and of multi-homed devices. However, modifications are required to both the MN and the CNs and the mobility has to be signaled with authentication procedure end-to-end to all CNs, which increases the overall procedure delay as well as the processing at the end nodes.

### 4.2.4 Mobility Support for Locator Identity Split Protocol

Locator Identity Split Protocol (LISP) [169] [170] has as goal to increase routing scalability in the Internet Backbone, therefore being outside of the main scope of this dissertation. However, as it provides a specific mobility support solution [171], it is shortly evaluated here.

LISP presumes that the nodes of a network have assigned an IP address which represents their static Endpoint ID (EID). A local network is advertised and bound to a Router Locator (RLoc) which represents the address to which the Egress Tunneling Router (ETR) may be reached [172], [173], [174].

The uplink packets are forwarded to an Ingres Tunneling Router (ITR) which looks up in a generalized infrastructure for the locator advertised for the EID of the correspondent node, then encapsulates the data packets with an IP header addressed to the ETR which advertised the EID of the correspondent node. The ETR decapsulates the message and forwards it to the CN.

For the mapping infrastructure several approaches are proposed as to be able to scale the EID distribution in the complete IP domain. Through this mechanism, the address space of the network provider core networks is hidden (encapsulated) to the correspondent network providers, enabling a simplified routing scheme to be deployed in the backbone core.

For mobility support [171], the protocol considers that the MN includes the functionality of ITR and ETR as depicted in Figure 4.8. Through this mechanism, it can advertise its static EID to the mapping infrastructure whenever the node changes its network location, making the node reachable from the ITR of the CNs.

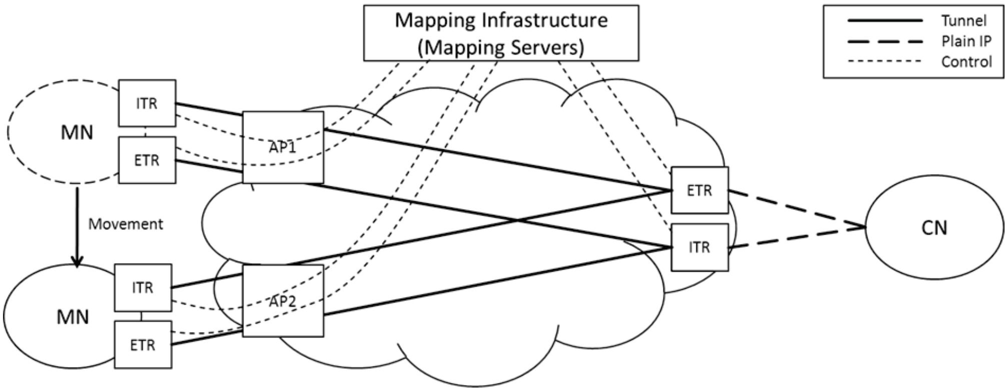


Figure 4.8: Mobility Support in Locator Identity Split Protocol

However, the MN depends on the presence of an ETR in the network of the CN or of a translation entity to encapsulate data packets to/from the MN, therefore requiring modifications in other operator domains. Additionally, the handover procedures have an increased delay as the correspondent domains have to be notified. This requires a signalling proportionate to the number of correspondent domains.

ETR may be replaced by a Proxy ETR (PETR) [175] in case the network of the correspondent node does not deploy LISP, which offers a solution similar to Mobile IP, thus increasing the data path length between the MN and the CN.

The "Routing on Flat Labels" (RoFL) solution [176] proposes that the addresses of the devices represent their identity, without considering any location related semantics. In order to be able to forward data packets to the appropriate destination, each intermediary node maintains a set of hash tables. When devices are changing their network location, all the appropriate hash tables have to be updated resulting in a storm of signalling and high inefficient signalling.

A similar solution to LISP and RoFL was proposed in the EU FP7 4WARD project [177], in which a set of dynamic anchors are placed close to the mobile devices handling both the local mobility and the appropriate forwarding to the correspondent nodes [178]. As LISP, also the 4ward solution requires the cooperation of all the correspondent node domains thus having a large adaptation delay.

Another solution from Eriksson [179] proposes the construction of one or multiple circuits, named Late Locator Bindings, between the MN and the CNs before as part of the communication establishment, thus generalizing LISP for a hierarchical set of ITR and ETR functions. Through this functionality, the service establishment has a larger delay, due to the circuit computations and the system is rather static to data path modifications, while having the large advantage of a faster data forwarding.

A hybrid solution between LISP and the late Locator Bindings is the Mobility and Multihoming Supporting Identifier Locator Split Architecture (MILSA) with specific servers which bridge the different realms and local realm forwarding doubled by a hierarchical identifier system enables the inter-domain forwarding based on node identity and the intra-domain forwarding based on the current connected interfaces [180], [181].

### 4.2.5 Transport and Application Level

The transport and application level solutions are creating a new reachability overlay on top of the underlying IP communication, using the same concepts of reachability, binding to the specific transport or application instance the local IP address.

Multiple options for sustaining mobility at the transport level addressing especially TCP and SCTP are available in the literature such as Multi-Path TCP (MPTCP) [182], [183], [184] and mobility support for SCTP (mSCTP) [185], [186], [187]. Apart from the specific transport level features, such as the freezing of TCP window during handover, these solutions have a common set of characteristics, which will be further described using the mSCTP solution. The SCTP protocol is based on associations between the MN and the CN. Whenever a handover occurs, the MN receives a new target IP address which is added to the SCTP association, then made primary IP address for the communication followed by the deletion of the source IP address. This procedure enables the transparent handover of an SCTP session.

The transport level mobility solutions do not require any network infrastructure. They rely on processing executed at the end nodes. Because of this, the data path length is not affected by mobility support at a cost of signalling the mobility updates for all data sessions established with all CNs.

The application level solutions in which the MN signals directly the application server on the specific solutions are similar to the transport level solutions, however with a reduced impact on the device, requiring only the installation of the specific software of the application on the device. The usage of the same application mobility support for multiple applications using the same service platform highly reduces the signalling as the mobility information has to be updated only one time.

For example, the Back to My Mac solution [188] uses a central DNS server to which the devices register and can be reached for all applications [189], [190]. The solution was designed for a reduced number of devices and is based on anchoring of their data traffic on a network translation node at the device manufacturer premises.

For the applications in which communication is established directly between two peers with server proxy negotiation, such as the multimedia session establishment using Session Initiation Protocol (SIP) [191] either in Internet or in operator deployed IP Multimedia Subsystem (IMS) infrastructures [100], the notification will not be executed only with the CNs, but also to the specific application level proxies.

For example, for SIP/IMS, whenever the IP-address of the MN, the sessions are re-established at the application layer directly or intermediated by service level entities [192], [193]. Figure 4.9 illustrates SIP-based MM in the reference scenario.

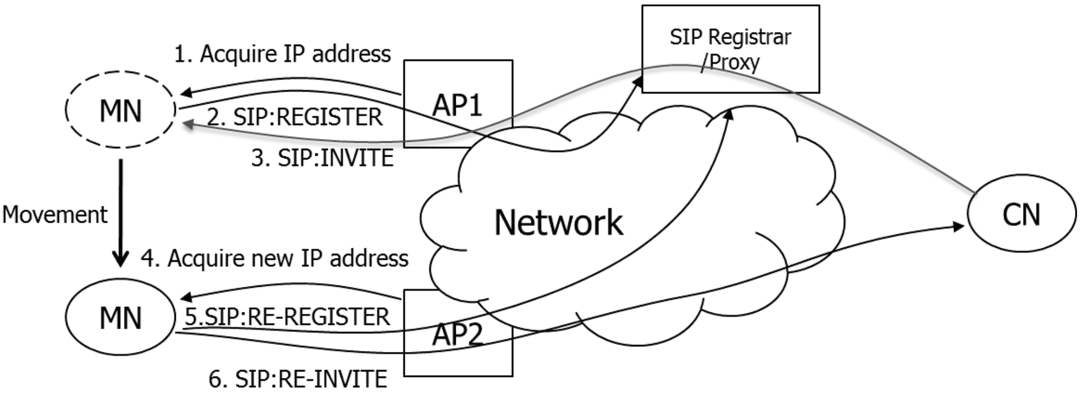


Figure 4.9: SIP/IMS Mobility Support

First, the MN receives an IP address from the first network (Step 1). Then, the MN registers with the SIP service as to be reachable by CNs (Step 2). The registration creates a binding between the MN public URI (e.g. alice@example.org) and the Contact URI (e.g. alice@10.11.12.13:5060) which enables correspondent nodes to reach the MN and to establish sessions (Step 3).

Upon a handover, the MN receives a new IP address from the second network (Step 4). Because of the IP address change, the contact URI changes, and the MN issues a re-invitation to all CNs so that they can update their bindings and re-configure media engines (Step 6). The MN also needs to update its registration in order to be reachable for future sessions (Step 5). Depending on the application, session re-establishment needs to be authenticated, causing additional signalling.

The 3GPP IP Multimedia Subsystem (IMS) adopted SIP for operator carrier grade services. In IMS, same session re-establishment is possible as in Internet SIP, but due to the binding of the session establishment functionality to resource reservation, the re-establishment passes through the Call Session Control Functions (CSCF) as in the case of the session establishment i.e. Step 2 and Step 4 have similar data path. Due to the scalability mechanism internally added to the IMS architecture, the CSCF entities may change during mobility as to better serve the MN, not affecting the number of messages that have to be exchanged.

This mechanism, in which the control of the session re-establishment is left completely to the MN, requires that additional functionality is added into the MN for determining handover events. In order to circumvent this problem and to maintain the MN functionality to the basic IMS functionality, 3GPP developed the Multimedia Session Continuity (MMSC) application server which transfers the control of the session handover to the IMS service provider [81]. MMSC maintains the bindings between the different contact addresses for a subscriber and the active sessions. Whenever one of the contact addresses changes, the MMSC makes a decision which sessions are to be transferred to the new contact address and initiates Third Party Call Control procedures. It creates two re-establishment sessions in which the new location is advertised: one towards the MN and one towards the CN.

Being an application layer mobility solution, MMSC anchors sessions at application layer, in a transparent manner to the other parties. This enables the architecture to cope with terminals which do not have mobility support functionality. The MMSC considers the split and merge functionality for the different data flows of a multimedia session, e.g. to transfer audio first and then video without losing session continuity.

Similarly to IMS, however enabling the communication over multiple service platforms, the Daidalos Project proposes a Virtual Identity Framework in order to reach the subscriber specific services and to seamlessly connect to these service platforms over heterogeneous access networks [157]. This feature transfers the anchoring of mobility to the level of the identity framework similar to the DNS solution, having the same domain and scalability limitations.

### 4.2.6 Discussion

In this section a set of widely deployed and highly academic considered core network mobility support solutions were presented. Following an evaluation of the solutions based on the previously selected criteria is done, leading to an overview of the current status and the challenges facing core network data path handling due to mobility.

Based on their main goal, the solutions can be classified into two categories: some concentrate on how to make mobility transparent to CNs by containing itself in the same operator network while some other are enabling a larger flexibility by integrating the CN or its operator domain into the procedures.

The solutions contained in an operator network such as Mobile IP and 3GPP solutions are maintaining a single anchor node for the data traffic of the subscriber no matter of the network location. While the MN is moving through the wireless environment, the data path is extended to pass through the anchor node. This implies that the data path stretch is longer due to the mobility support.

While maintaining a reduced signalling and adaptation due to the containment to a single operator domain, the anchor node solutions have a single point of failure nodes which reduce the flexibility and the robustness. The transparency to CNs of these mobility solutions enables an easier deployment in carrier grade operator networks. Additionally, solutions which come integrated and do not require further modifications of MNs such as the 3GPP mobility support are even easier to be adopted.

The other class of solutions include routing optimization of the Mobile IP, HIP and the transport and application level mobility solutions. Furthermore, mobility support of LISP is contained in this class of solutions as the role of the CN is taken by its operator domain.

These solutions which involve CNs are offering a data path which is not deterred by the mobility support as the end-to-end signalling enables the end nodes to send the data to the current location of the correspondent party. However, due to the mobility notifications which are send to all CNs, these solutions require an increased signalling and a longer adaptation duration.

As there are no data path entities involved into the mobility support except for the end nodes, these solutions are highly flexible and are providing the robustness required.

All the solutions except for the application level ones do not involve the subscription profile and the user requirements information into the decisions. This is mainly due to the design of the core network mobility support in which the specific procedures are triggered by the the change in MN location. However, core network data path adaptation may be required due to other factors which are currently not considered, such as fail-over or maintenance procedures.

At the moment of writing of this dissertation, no available solution adapts the complete data path through the operator core network transparently to the CNs.

## 4.3 Application Adaptation Current Solutions

Currently a set of mechanisms is available in the operator core network for reserving data path resources. The most pre-eminent paradigms are described in this section from the perspective of the solution's support for the adaptation of the reserved resources for one or multiple flows of the MN depending on the current access network used, with a minimal signalling between the MN, the service platform and the core network.

### 4.3.1 Best Effort Delivery

A great number of service platforms and applications are exchanging data with the mobile subscribers relying on the best effort delivery through the access and the core network. This category generically named Over-The-Top (OTT) [194] includes applications especially coming from the Internet world such as Skype, Yahoo, Google YouTube etc. To be able to deliver the data at the best possible levels for the subscribers, OTT applications are able to communicate at multiple levels of data rates. For example, a video stream may be delivered from the network at a 300kbit rate with a specific level of experience for the subscribers or at a 700kbit rate bringing a better experience.

For OTT, a reactive adaptation mechanism has to be deployed. The active monitoring of the data flows at MN and at CN enables the application to become aware when the level of required resources can not by momentarily satisfied by the available resources. As the monitoring reacts to the received data, awareness when more resources are available in the network can be done only through extending the communication throughput until some delivery failure is notified.

Two categories of actions can be triggered by the data delivery awareness. First, the data traffic may be adapted through data re-transmission, through usage of transport protocols such as TCP and SCTP. This ensures the consistency of exchanged data at the cost of increased delay and jitter. Second, in case one of the communicating parties determines through the received monitoring information, that the communication can not be sustained with the momentary available level of resources, it can trigger an application adaptation procedure. The adaptation procedure implies the signalling of the CN, that a new level of throughput is required.

In order to be able to execute this adaptation, a new profile has to be negotiated between the MN and the CN through signalling at application layer, which is then enforced on the data exchange such as in case of IMS services [100]. In case of mobile communication, this implies that the MN has to signal over the wireless link to the CNs the service adaptation. This solution has two major disadvantages. First, the MN has to negotiate over the wireless link the new session profile with the CNs, which creates a signalling overhead requiring specific functionality to be available at the communicating parties as well as an extended throughput through the wireless and core network link. Second, the negotiation of a new profile is reactive to a data communication deterioration, therefore this method comes with a quality of experience degradation especially because of the increased overall delay due to re transmissions.

The adaptation procedure may be triggered by introducing awareness of the access network currently in use at application level. Several cross-layer communication solutions are available in the literature, concentrating on how and when the applications can be triggered by the link layer to adapt their data sessions. However, these solutions do not concentrate on optimizing the signalling and on reducing the overall adaptation time, relying on the application level negotiation of new parameters which requires at least one round trip over the wireless link.

For OTT applications, the adaptation to available resources is executed reactively based on an estimation of available network resources. However, only the network is aware of the exact level of available resources due to the possibility to observe the communication of all MNs. Therefore multiple adaptation procedures may be required until the correct available resources level is reached.

Furthermore, as for these applications the mobile operator is not ensuring any resource level guarantees, the available resources in the network may change very fast. Therefore, OTT applications may required adaptations even when the MN is not changing the connectivity. By this the reactive adaptation procedures number is furthermore increased, thus the degradation perceived by the users and the signalling overhead in the wireless access network.

### 4.3.2 On Data Path Signaling

One of the attempts to solve the resource reservation issue in the Internet was based on the end-to-end signaling on the data path. The paradigm is named Integrated Services (Int-Serv) [195]. In the Int-Serv framework, the Resource Reservation Protocol (RSVP) [196], [197] allowing the guarantees for bandwidth [198] or for bandwidth and delay [199]. RSVP enables the communicating parties to explicitly announce the network entities about the resource requirements, through this allowing the network nodes to signal each other in order to allocate the resources on the end-to-end data path.

The signalling is executed prior to the session establishment on a single data flow basis, including the service profile negotiation. The Sender of the data flow transmit a message to the Receiver in which the data traffic is described. This enables the establishment of a data path for the packets. The Receiver responds with a selected data traffic profile which is then enforced on the data path. All the messages are received and processed by all the data path entities.

For mobile networks, RSVP was extended to enable data flow differentiation in Mobile IP messages [200]. When a handover occurs, the end-to-end signalling is re-executed enabling the modification of the reserved resources according to the user selected profile over the specific access network. In order to reduce the signalling, a Mobile RSVP [201] was proposed in which the resources are reserved in either all the access networks or only in the networks to which the MN has a high probability to handover too. This solution creates a very high overhead on resources pre-reserved in various access networks, therefore not being suitable for the current data rates required by the mobile users.

Although Int-Serv provides the necessary resource allocation and the notification of the CNs, the number of messages exchanged during the service provisioning and adaptation is directly proportional with the number of unidirectional data flows. With the current number of subscribers and data flows, RSVP is not scalable. For Int-Serv processing logic has to be introduced in all the network nodes for making the service available. Additionally, as the messages are transmitted to all the network nodes before reaching the communicating party, the Int-Serv solution has a high latency.

To overcome the limitations of Int-Serv from the scalability and mobility perspective [202], [203], the Next Steps in Signaling (NSIS) [204] is proposing a new framework which relies on partial data path signalling for the resource reservations. Additionally, it relies on the Differentiated Services Framework (Diff-Serv) [205], [206] for the guaranteeing of the resources reserved between different NSIS capable nodes. The architecture is illustrated in Figure 4.10.

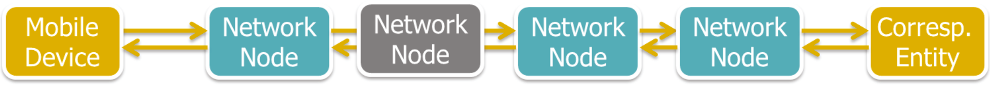


Figure 4.10: On Data Path Signalling with NSIS

NSIS standardized the General Internet Signalling Transport (GIST) [207], [208] which enables the transmission of resource reservation data between different entities in the network which may be intermediary nodes executing resource reservations or the end nodes [209]. The solution relies on applications to provide at one of the intermediary nodes the resource requirements which are then aggregated through the other network entities, thus highly reducing the signalling of RSVP.

Regarding a specific session adaptation based on the current MN's access network, GIST requires the MN to signal the next NSIS capable node over the wireless link including the required adapted resources which then is forwarded on the end-to-end path to the CNs. GIST allows the aggregation of information related to multiple data flows, therefore enables that the MN will send the adaptation resources level a single time.

However, as complexity of processing this solution is not different from RSVP. In this case also MNs have to monitor the link level, be able to determine the access network through the mobile device is currently communicating and make a decision whether the current level of the communication should be adapted. Then this message has to be received and processed by the CN while being communicated also to multiple data path network entities. Additionally, as the communication has to pass through multiple nodes to reach the CN, the adaptation delay is high.

### 4.3.3 Policy Driven Resource Control

A more profile driven method for reserving data path resources is the policy driven solution. A policy is a logical description of a goal and of a set of actions which determines a class of decisions [210]. Typically, policies are expressed as a set of rules which provide the means for administration and control of the network resources [211]. The policies can be used in an automatic response from the network side to the requests of the users and to changes in conditions.

Several solutions are available in the literature on how the policy driven resource control can be integrated into the network provider infrastructure and specifically into the mobile networks. Three of the most pre-eminent ones are described in this section.

#### 4.3.3.1 Basic Policy Control

The basic framework for policy control enables the remote control of network resources. As depicted in Figure 4.11, it consists of two entities: a Policy Decision Point (PDP) and a Policy Enforcement Point (PEP) which exchange information using the Common Open Policy System (COPS) protocol [212]. The PDP is a logical entity which is aware of the network capabilities, of the momentary conditions and of the preferred set of policy rules. The PDP is able to decide which of the rules to be enforced on other network entities denominated as Policy Enforcement Points (PEPs).

Based on the policy related information retrieved from a policy repository, information, the PDP determines which situation rules are effective in the momentary

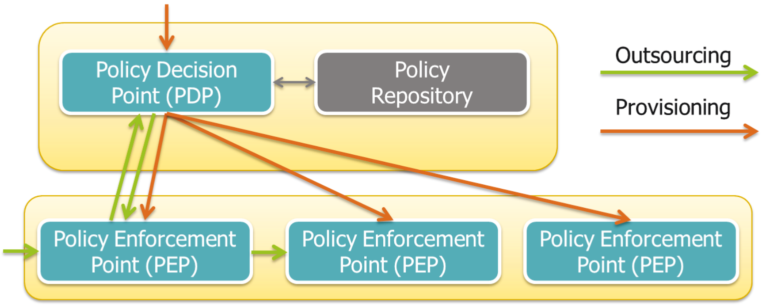


Figure 4.11: Policy Framework

situation and stores them for efficient further processing.

Whenever a PEP is querying about some resource reservation, PDP responds to the query by making a policy decision and responding with the adequate rules. PEP installs the rules on the data path. Additionally, it monitors and reports to PDP the changes in the framework and the exception cases.

Two operation models are considered: "outsourcing" and "provisioning". For "outsourcing" the resource reservation procedures are triggered by external events, for example by the requirement from PEP of a new resource reservation. As the procedure should be executed due to a resource requirement on the data path, for example due to an RSVP request, the policy decision has to be executed fast in order not to deter the data path communication and to further delay the service establishment.

The "provisioning" model relies on the triggering of the policy decisions on the PDP before any resources are required to be reserved on the data path. A decision is requested from the PDP for multiple data flows at the same time, prior to the actual data communication which are then enforced on the PEP. Because of this proactive enforcement, the PEP does not have to query while receiving data path information the PDP, therefore reducing the latency of data path establishment.

The "provisioning" approach is not relying on any other data path signalling for the resource reservation or on reactions to the data traffic received as in the case of the "outsourcing" model. Additionally, it allows the interconnection with service platforms which can transmit the application requirements for the data flows during the negotiation with the mobile devices, through this making the resource reservations transparent to the mobile devices. Also, the provisioning can be done to multiple PEPs using the same decision, therefore requiring less functionality in PDP and a reduced signalling between the PDP and the PEPs.

A new perspective for resources reservations based on classes of connected devices and aggregated data flows description is introduced. However, the framework does not contain any other functionality, constituting as the basis for other solutions, therefore not allowing the comparison on how the two modes would function in case of service session adaptation due to mobility in heterogeneous wireless environments. Following, the description of the extensions of the two modes is presented.

#### 4.3.3.2 Open Flow

Stemming from the basic ”outsourcing” mode of the policy control means, the OpenFlow (OF) is a framework which enables the remote control of data path entities [213]. Designed initially as a solution for large scale testing of new mechanisms for data delivery for new Internet services such as data center cloud computing, OpenFlow entered into a standardization phase through the Open Networking Foundation (ONF) [214].

The architecture of OF is depicted in Figure 4.12. It consists of two entities which can be directly mapped to the basic policy architecture. Located on the data path an OF switch, similar to a PEP, is requesting information on how to handle the data flows from a remote OF controller similar to a PDP [215]. Based on the received policies, the data path is adapted correspondingly and through this the resources are reserved.

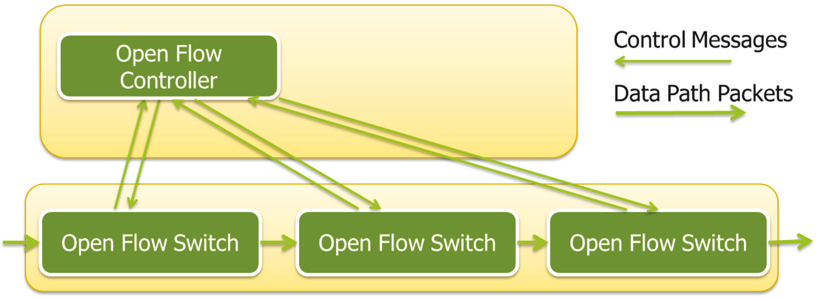


Figure 4.12: Open Flow Framework

Open Flow was designed as reactive ”outsourcing” solution for the policy control of data path entities, enabling the deployment of new concepts without requiring hardware modifications. By separating the decision into a logical entity which may be shared by multiple data path entities, the system can dynamically adapt to network changes, for example to the access network reselection procedures of MNs and to adapt according to the policies the resources reserved.

However, three limitations are foreseen for this solution. First, the triggering of an update of the resources allocated to the MN on a specific access network is reactive to its data traffic. Only when data traffic becomes available over a specific access network, the OF switch is able to query the OF controller on the resources that have to be reserved, followed by a response containing the rules that have to be enforced on the data path.

Second, the installation of the rules on the data path is executed for each of the switches separately upon request. Although this enables a very low granularity control of the network, it requires that the different data path entities are able to react to the data traffic. Also, because of this reactive property, the complete data path resources will be established with an increase delay which may deter the seamless communication.

Finally, Open Flow at the moment of writing of this dissertation does not have an interface through which the users and the application platforms can provision in the controller the requirements for the various policies that have to be executed. Because of this, the Controller has to rely for the policy decisions on a static set of information. Additionally, the identity of the communicating parties remains unknown to the Controller, as the IP address allocated to MNs may change with the change of the point of attachment.

Because of these limitations, Open Flow remains a contained solution to some data path entities and their controller which is not currently able to sustain the complete service adaptation to the access network to which the devices are connected to.

#### 4.3.3.3 3GPP Policy and Charging Control

The basic policy model was extended by 3GPP for the control of the resource reservations over heterogeneous access networks in a convergent manner as to simplify both the service signalling and the actual provisioning.

The 3GPP Policy and Charging Control (PCC) architecture, as illustrated in Figure 4.13 maintains the two major components of the basic policy control: the Policy and Charging Rules Function (PCRF) which has the role of PDP and the Policy and Charging Enforcement Function (PCEF) or the Bearer Binding and Event Rules Function (BBERF) which have the roles of PEP [79]. These functions are augmented by a Subscription Profile Repository (SPR) which contains the MNs' network profile and the dynamic connectivity state and parameters including the current access network used and the IP address [107].

Additional, online and offline charging functionality (OCS and OFCS) is connected to the data path enabling the accounting of the data traffic of the subscriber, the data information harmonization and integration and its transfer to the Billing System [216], [217], [218], [219], [220], [221]. This functionality is not further analyzed.

PCRF is able to communicate with service platforms which are generically named Application Functions (AF). Through this interface PCRF receives re-

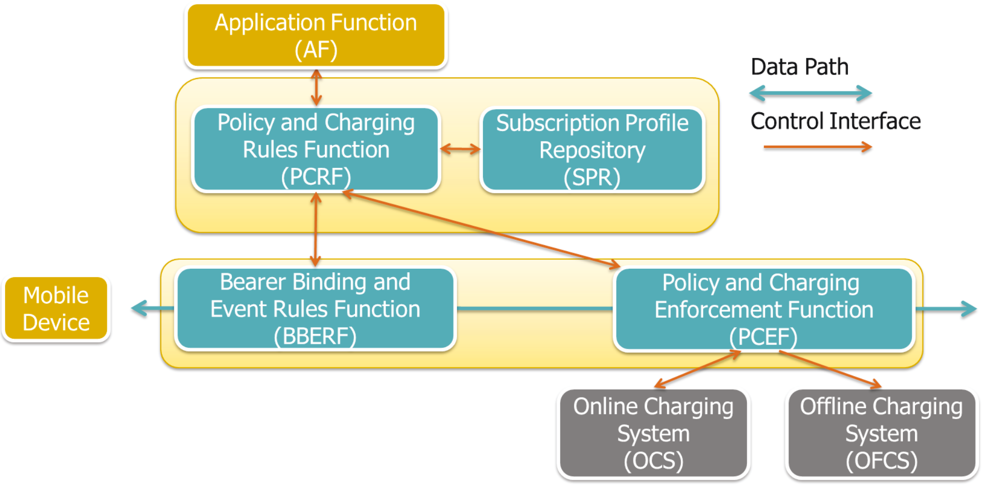


Figure 4.13: 3GPP Policy and Charging Control

quirements and transmits events enabling the resource reservations according to the profile negotiated between the MNs and CNs [109].

PCC relies and functions at the granularity of data bearers. A bearer is a part of the data path allocated to a subscriber of a defined capacity, delay, bit error rate etc. A bearer uniquely identifies packets that receive common QoS and charging treatment between the UE and the PDN GW including scheduling, queuing and rate-shaping. For this, a bearer encapsulates multiple data flows defined by a five-touple based packet filter described as a set of Traffic Flow Templates each containing the source and destination IP address and port and protocol ID [108], [79].

Two types of bearers are defined: default and dedicated bearers. The default bearer is allocated as part of the data connection establishment, providing basic connectivity and default QoS and charging according to the user subscription data. The dedicated bearers are used for allocating a different level of QoS and charging for specific data flows, as requested by the various services. Multiple dedicated bearer can be allocated at the same time, existing as long as a specific data session is active [108].

PCC is functioning in four distinct modes, depending on the action trigger from application platforms, from data path, event based or subscriber information based. The modes are depicted in Figure 4.14.

PULL mode is similar to the ”outsourcing” model of the basic policy functionality, enabling the establishment and termination of default bearers. The data path, entities either BBERF or PCEF, upon determining that the MN has attached or detached from the network, transmits an unsolicited request related to PCRF.

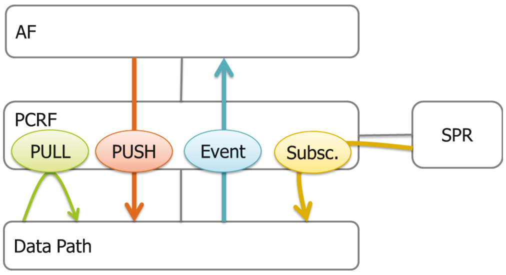


Figure 4.14: Policy and Charging Control Modes

PCRF fetches the subscription profile from SPR and makes a policy decision, resulting in a set of rules, which are then installed on the data path. A specific case of this mode is triggered by the online charging system through PCEF when the subscriber has no more credit, therefore the default and dedicated bearers have to be terminated.

PULL mode operations are also executed in case of handovers which require BBERF change. The target BBERF is not aware that the UE had already attached to the network and transmits a new attachment request to PCRF. PCRF as it is aware that the request represents a handover, sends to the target BBERF the rules as installed on the source BBERF for the default and for all the dedicated bearers. Through this procedure, the resource reservations are transferred between the two BBERFs in case of a handover.

PUSH mode is similar to the ”provisioning” model of the basic policy functionality, enabling the establishment, modification and termination of the dedicated bearers. PCRF receives a request from one of the service platforms represented by an AF and based on the specific data flow descriptions and on the subscription profile from SPR, it makes a policy decision which is then enforced on the data path entities. This mode enables the resource reservation during service establishment, resulting in specific rules installed when the actual data traffic is received, therefore not deterring the communication of the MN.

A specific case of PUSH mode is when the information on a data session establishment is received from a traffic detection function. In this case, the reservation of resources is executed based on the internal operator policies, without any communication with the AFs, enabling the operator to shape the communication only considering its current efficient delivery characteristics.

Two additional modes were added to PCC compared to the basic policy framework: the event and the subscription modes. The event mode, enables PCRF to receive subscription requests from the AFs and to subscribe to data path events. When an event happens on the data path, PCRF makes a policy decision resulting in new rules to be installed on the data path. Additionally, PCRF transmits a notification on these events to AFs which have subscribed. The notification enables service platforms to adapt their communication accordingly, through a new negotiation of the service profile with the MN and then through a new resource reservation using PUSH mode.

The subscription mode enables PCRF to adapt the allocated data path resources to subscription profile modifications. Upon device attachment to the network, PCRF subscribes to SPR for events related to the profile modification. When such a modification occurs, the SPR notifies the PCRF which may make a new policy decision resulting in new rules to be installed on the data path or to transmit notifications to AFs.

Through a combination of these procedures, with a minimal modification, PCC is able to ensure the active data sessions adaptation to the access network to which MNs are connected to. When a handover occurs to a new access network, the specific event is notified to the service platform, which in its turn re-negotiates the session parameters with the MN and then triggers a resource modification on the data path. However, this procedure requires an extensive signalling between the mobile device, the service platform and the core network which influences the seamless characteristic of the handover. As to reduce this effect, PCRF has to first transfer the bearers without any modification from the source to the target access network. Therefore, in case an adaptation is required, the reservation procedures have to be executed twice, once with the level of resources required for the source access network and once with the target level of resources further increasing the signalling and the processing in the core network.

PCC adaptation procedure has a high impact on the delivery of the service and on the service platforms as they have to be able to receive notifications and to trigger, new reservation procedure through the communication with the MN while communicating at the previous levels.

### 4.3.4 Discussion

In this section, a set of core network resource reservation mechanisms were presented, considering their capability of efficiently adapting the resources allocated for the active applications of a MN to the access network to which it is attached to.

One of the main concerns regarding the resource adaptation is the scalability of the given solution reflected in the number of messages that have to be exchanged, the logic that has to be maintained on different network components, and the overall delay of the procedure.

From this perspective, the OTT solutions require the re-negotiation of the service profile between applications and the mobile devices. OTT solutions do not reserve resources on the data path, adapting only the transmission rates.

The solutions which signal on the data path also require that the MN is aware of the access network change. Additionally, they require a large number of data path entities to receive messages and to make decisions on the resources to be reserved, which requires a high level of signaling and processing, increasing the adaptation delay.

The Open Flow solution is able to adapt directly the resource reservation in the network. However, it currently lacks any means for communication with the application platforms. Therefore, even though the resource adaptation may be technically feasible, it can not be currently realized.

The 3GPP Policy and Charging Control is able to adapt the resources reserved for a specific MN depending on the access network to which it is attached to. However, the procedure is composed of two methods apart from the handover procedure: one which includes the signaling between the MNs and the service platform for determining the adapted service profile and another additional one for the resource reservations of the newly negotiated service profile.

At the moment of writing of this dissertation, no solution is available which adapts the MN's resource reservations according to the levels required over specific access networks considering a reduced signaling and processing delay through a single harmonized method.

Chapter 5

# Design and Specification of SelfFit Framework

This chapter includes the design and specification of the SelfFit framework, introduced by this dissertation. Based on the requirements and on the related technologies presented in the previous chapters, a new system is built, enabling flexible adaptation to access and core network characteristics and service resource adaptation to mobility.

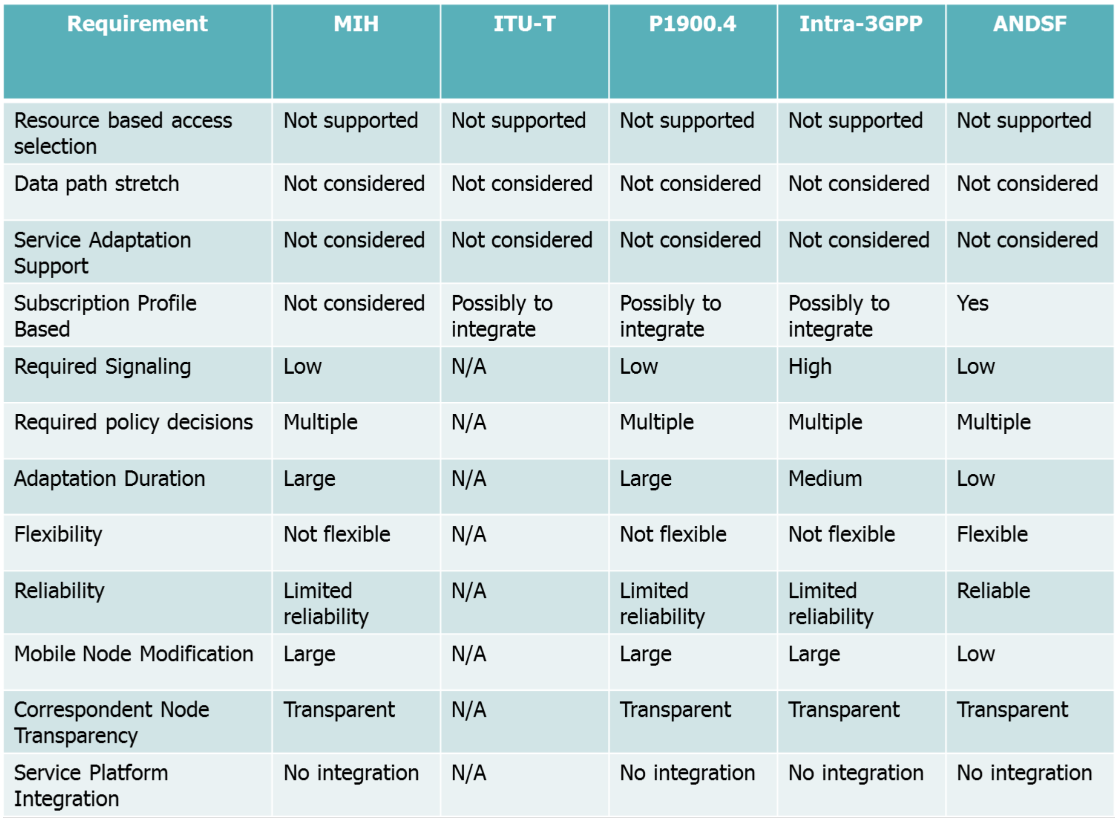
First, the various state of the art technologies are analyzed determining how appropriate is their usage in the SelfFit framework and the functionality gap according to the requirements of the carrier grade operator networks. Based on this a set of technologies is selected and the exact functionality which should be addressed by the SelfFit framework is determined.

Based on the functionality, the high level architecture is presented in the form of a set of generic enablers each exhibiting a specific feature. Then, for each of the enablers, the functionality and the interfaces are designed and specified as new functional additions to a generic carrier grade network, leaving for the next chapters the implementation of the SelfFit framework as part of the current network architectures, including the integration with the state of the art technologies and the prototype realization and validation.

## 5.1 Introduction

In the previous sections, a set of related technologies were presented and individually assessed from the perspective of fulfilling the objectives foreseen for the future carrier grade mobile broadband. This section summarizes and evaluates their usability in a final framework from the perspective on how the requirements are fulfilled.

From the perspective of access network selection, the current available technologies are concentrating on the radio conditions change triggered access network selection. Due to the afferent time duration limitations, they do not integrate resource based access network selection concepts or other consideration on the adaptation of the communication based on subscription profile and user preferences. The comparison of the various solutions, as presented in Section 4.1.6, is summarized in Table 5.1.



| Requirement | MIH | ITU-T | P1900.4 | Intra-3GPP | ANDSF |
|---|---|---|---|---|---|
| Resource based access selection | Not supported | Not supported | Not supported | Not supported | Not supported |
| Data path stretch | Not considered | Not considered | Not considered | Not considered | Not considered |
| Service Adaptation Support | Not considered | Not considered | Not considered | Not considered | Not considered |
| Subscription Profile Based | Not considered | Possibly to integrate | Possibly to integrate | Possibly to integrate | Yes |
| Required Signaling | Low | N/A | Low | High | Low |
| Required policy decisions | Multiple | N/A | Multiple | Multiple | Multiple |
| Adaptation Duration | Large | N/A | Large | Medium | Low |
| Flexibility | Not flexible | N/A | Not flexible | Not flexible | Flexible |
| Reliability | Limited reliability | N/A | Limited reliability | Limited reliability | Reliable |
| Mobile Node Modification | Large | N/A | Large | Large | Low |
| Correspondent Node Transparency | Transparent | N/A | Transparent | Transparent | Transparent |
| Service Platform Integration | No integration | N/A | No integration | No integration | No integration |

Table 5.1: Access Selection Technologies Assessment Summary

Based on the metrics defined for each evaluation criteria, a further quantitative evaluation of the quality of each solution is defined in Appendix A.2.1, Section B.4.

However, with passing towards a massive broadband carrier grade operator environment, in which the coverage of the different access networks is similar, the connectivity continuity feature will be less used as the seamless service continuity through resource based access selection.

A solution for access network selection, which considers the momentary required resources, enables an operator to better orient the communication to the appropriate access networks. Such a solution requires to gather information on the available access networks in the vicinity of the MN, the momentary required resources and subscription profile information.

As observed from the analysis of the current technologies, the only solution with reduced adaptation duration and minimal MN modifications is the 3GPP ANDSF.

The network-located ANDSF is transmitting indications to MNs through a subscriber oriented management layer. The access networks selection itself is executed autonomously by MNs in most of the cases. As the devices do not require any additional information from the network during the critical handover period, the duration of the overall procedure is highly shorten. For this reasons, the solution proposed by this dissertation for access network selection is part of a subscriber oriented management layer. The SelfFit solution practical exemplification is realized as extension of 3GPP ANDSF solution.

From the perspective of the mobility support and data path through the operator core network, the current available technologies are separated into two distinct categories. A first category includes end-to-end solutions which require the modification of CNs or CNs' operator domains and have optimal data stretch at the cost of a high delay and of extended signaling. As these solutions require the modification of a large number of devices, their deployment is not feasible on a medium to long term. The second category of solutions are limited to a single mobility domain and maintain the data path changes transparent to the CNs while having a reduced signaling and a low handover delay. These solutions are considered as the best basis solution for the SelfFit solution. The comparison of the various solutions as presented in Section 4.2.6 is summarized in Table 5.2.

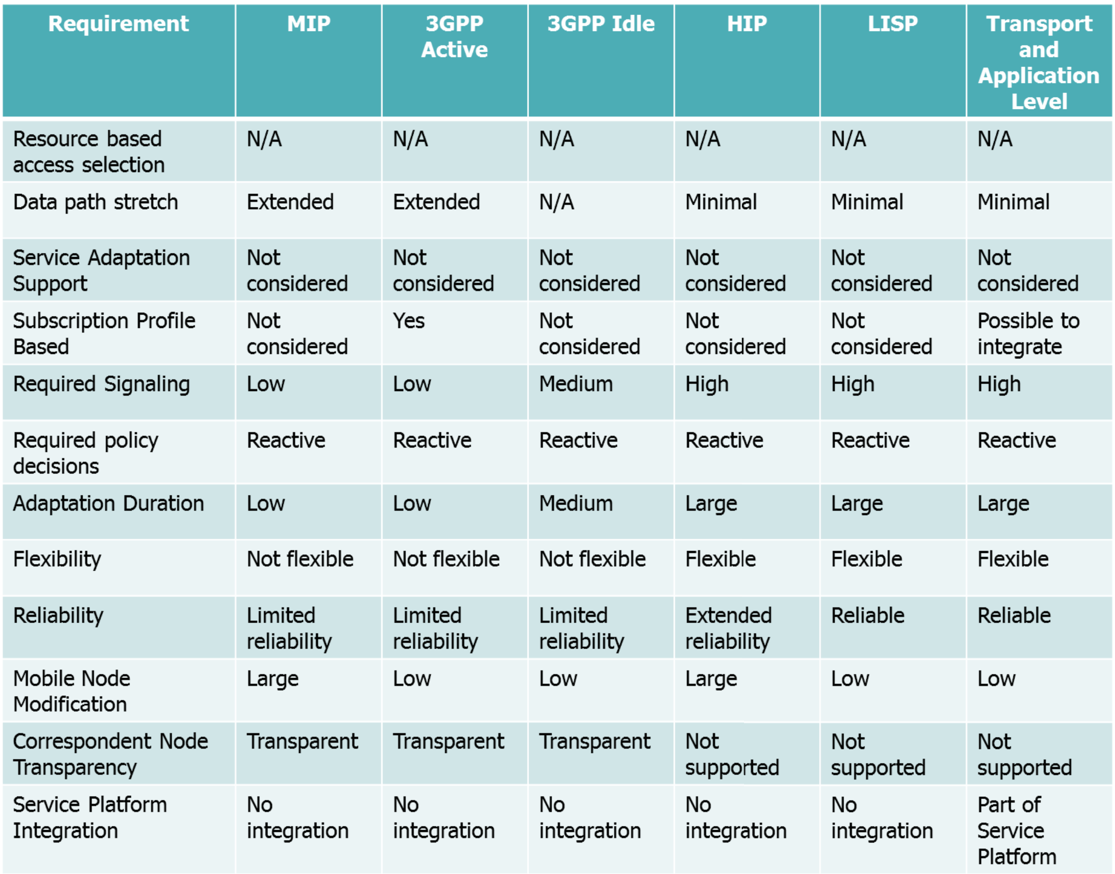
Based on the metrics defined for each evaluation criteria, a further quantitative evaluation of the quality of each solution is defined in Appendix A.2.1, Section B.4.

In the operator contained solutions, the data path stretch is extended as to pass though an anchor node which enables the transparent communication. The extension of the data path is highly detrimental to the network providers especially when deploying flat network architectures such as the Internet, as multiple entities have to be deployed for delivering the same data packets.

A solution for mobility which considers the complete adaptation of the data path has to be able to consider the subscription profile of the MNs as to select the appropriate data path network entities serving the device. Additionally, the solution has to be able to transmit indications to the mobility control layer to adapt the data path. For these reasons, the adaptation solution has to be part of SelfFit subscriber oriented management layer.

From the solutions which are limited to a single mobility domain, the most promising is represented by the 3GPP active mode mobility as it has a larger network coverage as the 3GPP idle mode solution and it has reduced signaling and handover delay compared to the Mobile IP solutions. For this reason, the SelfFit architecture is exemplified using the 3GPP active mode solution with both the GPRS Tunneling Protocol (GTP) and Proxy Mobile IP (PMIP).

From the perspective of the automatic application adaptation to the services to the network, the various QoS reservation solutions provide the specific means



| Requirement | MIP | 3GPP Active | 3GPP Idle | HIP | LISP | Transport and Application Level |
|---|---|---|---|---|---|---|
| Resource based access selection | N/A | N/A | N/A | N/A | N/A | N/A |
| Data path stretch | Extended | Extended | N/A | Minimal | Minimal | Minimal |
| Service Adaptation Support | Not considered | Not considered | Not considered | Not considered | Not considered | Not considered |
| Subscription Profile Based | Not considered | Yes | Not considered | Not considered | Not considered | Possible to integrate |
| Required Signaling | Low | Low | Medium | High | High | High |
| Required policy decisions | Reactive | Reactive | Reactive | Reactive | Reactive | Reactive |
| Adaptation Duration | Low | Low | Medium | Large | Large | Large |
| Flexibility | Not flexible | Not flexible | Not flexible | Flexible | Flexible | Flexible |
| Reliability | Limited reliability | Limited reliability | Limited reliability | Extended reliability | Reliable | Reliable |
| Mobile Node Modification | Large | Low | Low | Large | Low | Low |
| Correspondent Node Transparency | Transparent | Transparent | Transparent | Not supported | Not supported | Not supported |
| Service Platform Integration | No integration | No integration | No integration | No integration | No integration | Part of Service Platform |

Table 5.2: Mobility Support Technologies Assessment Summary

through which this adaptation can be realized. As summarized in Section 4.3.4 and included in Table 5.3, the resource reservation solutions have evolved towards centralized policy based operator controlled architectures.

Based on the metrics defined for each evaluation criteria, a further quantitative evaluation of the quality of each solution is defined in Appendix A.2.1, Section B.4.

All the provided solutions are able to support the adaptation of the resources of the specific applications through a session modification procedure which is similar to the flow establishment procedure, replacing the old parameters with the new ones. However, none of the solutions is able to execute this adaptation without receiving a trigger for the adaptation and the new parameters from the application delivery, thus requiring an extensive signaling.

For autonomously adapting the application resources, with a minimal communication with service platforms, the policy decision entity has to maintain specific management parameters enabling it to make the appropriate decisions. Therefore

| Requirement | Best Effort Delivery | On Data Path Signaling | OpenFlow control | 3GPP Policy control |
|---|---|---|---|---|
| Resource based access selection | N/A | N/A | N/A | N/A |
| Data path stretch | N/A | N/A | N/A | N/A |
| Service Adaptation Support | Not considered | Possible | Not considered | Possible |
| Subscription Profile Based | Not considered | Possible | Possible | Yes |
| Required Signaling | N/A | High | High | Medium |
| Required policy decisions | N/A | Multiple | 1 / adaptation | 1 / adaptation |
| Adaptation Duration | N/A | High | Medium | Medium |
| Flexibility | N/A | Not flexible | Not flexible | Limited |
| Reliability | N/A | Limited reliability | Limited reliability | Extended reliability |
| Mobile Node Modification | None | High | None | None |
| Correspondent Node Transparency | None | High | None | None |
| Service Platform Integration | N/A | No integration | No integration | Possible |

Table 5.3: Resource Reservation Technologies Assessment Summary

it is expected that a subscriber/data service management layer is transmitting these parameters for each data session which requires autonomous adaptation.

As basis for SelfFit solution, 3GPP Policy Control is considered, as it is the only evolved platform having a standardized means to receive control and management parameters for specific sessions. At the moment of writing of this dissertation, the OpenFlow highly promising solution has no interface towards the service platforms, thus not enabling communication with external control entities. Therefore, OpenFlow was not further considered.

In conclusion, the current network architecture does not include all the necessary features responding to the requirements of the future carrier grade operator communication. However, the basic features are available providing the fundamentals on which other new functionality can be developed. From the available technologies a set was selected based on their suitability to answer to the requirements and based on their potential to be further extended.

Through the analysis of the available technologies against the requirements, it is concluded that a subscriber oriented management layer best provides a suitable framework for answering to the remaining challenges. As the main challenges are related to communication customization for each individual subscriber, the functionality represents an extension of the current control layer. However, as such a framework mainly modifies the parameters on which the control level is functioning, enabling it to respond in a customized manner to the requests of MNs and of the service platforms, it represents a management layer functionality. This solution requires a novel approach to management, stemming as an extension of the control plane as it has to modify dynamically, based on the current network status the parameters which enable the control layer decisions.

## 5.2 SelfFit High Level Architecture

This section defines the SelfFit framework architecture. By definition, the SelfFit framework is considered to be an automatic adaptation functionality located in the operator core network relying on the information received from connectivity control functions and transmitting commands which adapt the MN's communication.

For determining the appropriate functionality to be further specified and how this functionality is integrated in the high level architecture, the carrier grade operator environment is first modeled and then the description of the here proposed enablers is following.

### 5.2.1 System Model

A typical communication environment is composed from three main actors as illustrated in Figure 5.1. In order to use the connectivity service, the mobile subscriber uses a MN, which is able to communicate using an operator network to the various service platforms.

From a business and technology perspective, two types of services are established between these actors as considered in the basic NGN principles [15].

Forming the service stratum, the first type includes the services established between the mobile subscriber and the service platforms. For this, data is exchanged between the application on the MN and the application on the service side based on the specific characteristics.

To be able to exchange data with service platforms, the subscriber is using a connectivity service of a core network operator, which service forms the transport stratum. Through signaling between the MN and the operator network, a transport support for data packets is established. The transport support is maintained by the operator for the MN, as long as the communication is required in any geographical location covered.

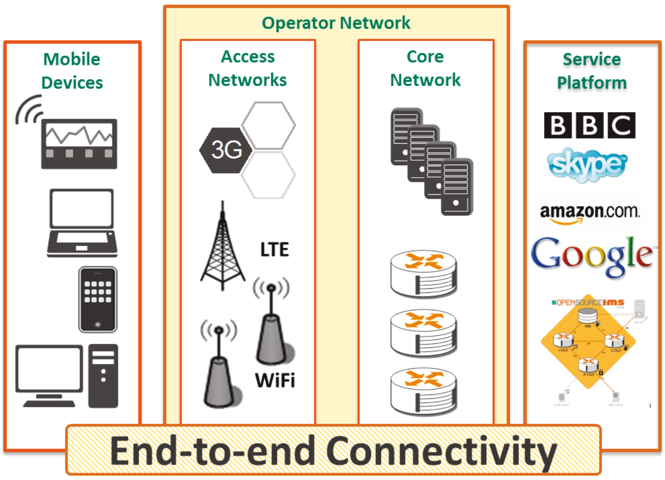


Figure 5.1: Simplified Carrier Grade Connectivity Model

For establishing the connectivity service, the MN is attaching to an access network which pertains to the mobile operator. On its turn, in order to establish a communication channel between the MN and the various service platforms, the operator is establishing a data path through the core network by selecting gateways and control entities.

Additional to access and core network entities selection, forming a complete data path through the network, the operator core network has to reserve a specific default level of resources for the subscriber to enable it to exchange data with the service providers and to supply a default security and privacy level for the communication. Using the default resources, the MN establishes application level services with the service platforms. For these services, special resources may be reserved on the communication path apart from the default resources. The connectivity service is transparent to the service platforms.

The connectivity service is realized through the control functions of the core network including the subscription information, the resource and charging control and the mobility support.

In conclusion, the connectivity service is composed of the following four main features:

- A radio access network which is selected for the mobile device
- A set of core network entities and an afferent data path
- A level of resources reserved on top of the data path enabling the communication with the service platform

- A level of security and privacy for the communication, not in the scope of this dissertation

Each time the MN is changing its network location, a uniform adaptation procedure is executed by the network control entities i.e. the MN executes handovers in the same access network, remains anchored to the same anchor node and receives the same resources over the target network. In order to be able to adapt the operations for each specific subscriber, the control parameters which are used in the default decisions have to be modified according to the specific needs.

The modification of these parameters is mainly done through specific management operations in the operator core network. Thus, a framework which adapts for each subscriber the control parameters is to be implemented at management layer. However, as it is automatically reacting to changes in the environment, this novel management level has to be coupled with the data path and control entities similar to the current available control functions.

A self-adaptation framework for the connectivity through the carrier-grade operator core network has to be able to appropriately adapt the access network, core network path and the application delivery, as illustrated in Figure 5.2.

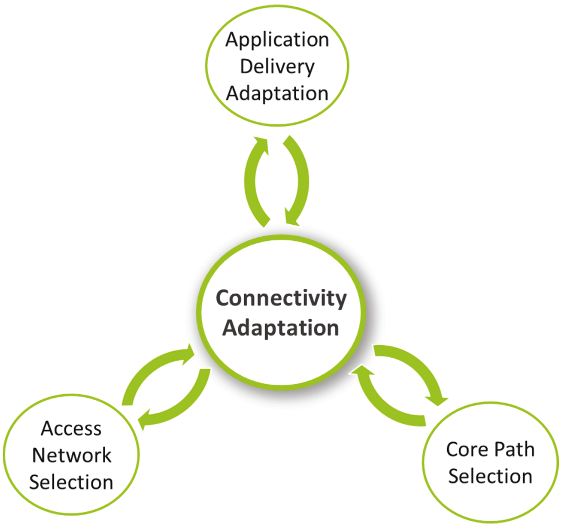


Figure 5.2: Connectivity Adaptation Features

The access network selection is dependent on the access networks available at the subscriber location, on the MN characteristics, and on the currently available resources in the possible target networks. An access network selection decision is transmitted in form of indications to the MN in order to adapt its connectivity through an automatic handover to the decided target access network.

The core path selection is depending mostly on the current status of the core network and on the other subscribers communication, using the system along with the MN's network location. The decision is transmitted to the mobility control entity adapting the data path transparent to all CNs.

The application delivery adaptation is based on the required resources of each of the MN's services established and on the access and core network data path established. The decision is enforced on the core network by the policy and charging control functionality.

Thus, three distinct decisions have to be taken in the core network in order to adapt the connectivity service. Due to the distinct nature of these decisions, due to the different parameters that have to be gathered, and due to the distinct enforcement points and for maintaining the high flexibility of the system, each of the decision points is implemented as a separate enabler. Each enabler may be present or may be missing in a specific deployment.

Following the here presented functional separation principle, the SelfFit self-adaptation framework is composed of three distinct enablers for access network selection, core path adaptation and application delivery adaptation. The high level architecture based on this design model is detailed in the next section.

### 5.2.2 SelfFit Functional Entities

The SelfFit Framework is designed for automatic adaptation of the mobile subscriber data transport through the carrier grade operator networks. It introduces in the transport stratum a subscriber oriented management plane adapting the parameters of the control plane of the operator network in the area of mobility support, resource control and based on the subscriber identity.

The SelfFit Framework is composed of three different enablers addressing the automatic adaptation of the connectivity over the access network, of the core network data path and of the application delivery support parameters. The high level architecture is depicted in Figure 5.3. It includes the interfaces with the other network functions as well as to the MN and the service platform.

The framework relies on the network identity of the subscribers and their afferent MNs, identity which is currently used for the authentication and authorization of the connectivity services. The subscription profile related to the specific identity is extended and retrieved from the same components as used by the control functionality. Through this identity, SelfFit is able to differentiate during decision making and enforcement between the different devices, enabling the independent individual adaptation.

Each of the three enablers is triggered and makes decisions based on parameters received from other functional components in the network. These components include the mobility support control and the resource control. Additionally, the

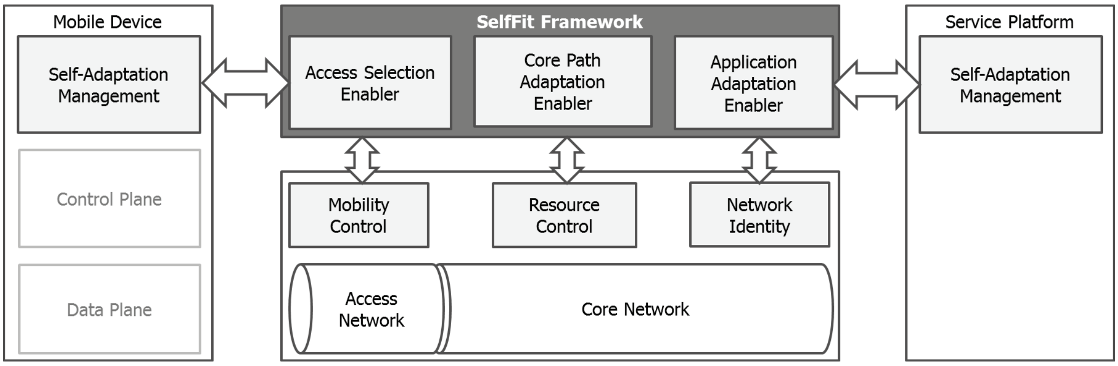


Figure 5.3: SelfFit High Level Architecture

decisions are enforced on the same components through the modification of the control parameters and through the execution of specific procedures. In order to enable this feature, the control plane components have to be extended towards the SelfFit management plane.

Additional triggers and parameters are retrieved from the MN itself and from the service platform. Some part of the decisions may be enforced on these entities, while maintaining as much as possible the communication transparency and minimal modification of functions which are not under the direct control of the carrier grade operator network.

In the next subsections, the functionality of the three enablers will be described separately including the information which is exchanged with the other network and external entities. Based on this information model, specific generic interfaces and their procedures are described. Finally, the integration of the three enablers is discussed as part of a common framework.

## 5.3 Access Selection Enabler (ASE)

The ASE aims at optimizing the MN communication through wireless connectivity adaptation. It concentrates mostly on transparent access network selection procedures addressing devices which can communicate through multiple networks located in their vicinity, using of the same or other radio technology and, through these dynamic means, to obtain a more scalable, carrier-grade operator network.

Additionally, the functionality includes features enabling service continuity and vertical handovers through introducing a new decision axis: A handover is executed due to a change in MN's required resources apart and orthogonal to a change in in radio connectivity.

For reaching its goals, the ASE is triggered and gathers parameters from external entities part of the wireless system. Based on this information, it makes decisions which access network should be considered by the MN in case the wireless connectivity has to be adapted. This information indicates the network operator preferences for the specific conditions of selecting an access network. Additionally, the indications can be transmitted to other network entities enabling the reduction of the handover procedure effects through a specific preparation phase which proceeds the actual handover trigger.

### 5.3.1 Architecture

ASE is composed of a set of functions which enable the communication with other core network and operator management entities for triggering the procedures, gathering parameters and making access selection decisions for each subscriber. Additionally it includes the communication with the MN for enforcing these decisions. ASE maintains a state related to each subscribed MN representing an extension of its network profile.

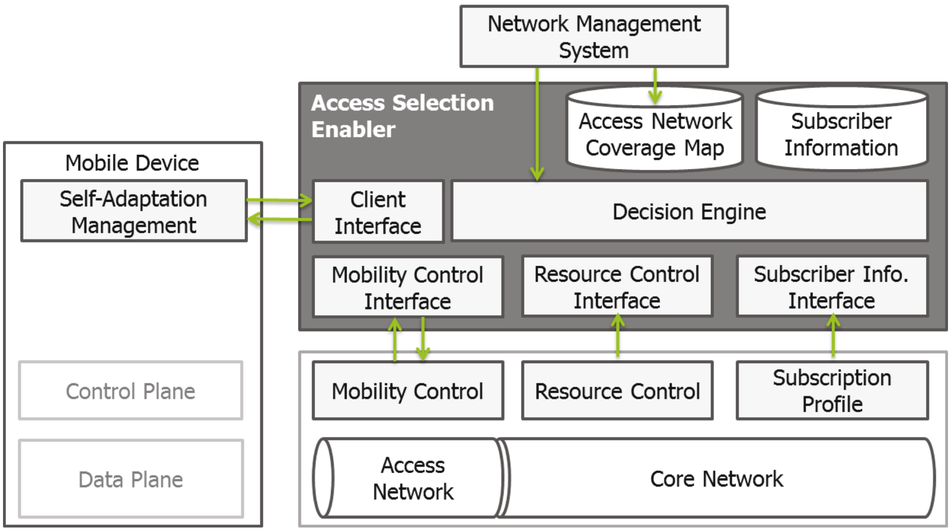


Figure 5.4: Access Selection Enabler Architecture

The ASE contains a set of interface functions enabling it to adapt and to control the communication with the other entities. Additionally, it includes a decision entity and a set of databases storing information and caching the decisions. The architecture is depicted in Figure 5.4. The following functions are included:

- Subscriber Information Interface - through this function, the ASE queries subscription profile related information. Additionally, it may subscribe and receive notifications to subscription information modification events.

- Resource Control Interface - this function retrieves information related to the MN's momentary reserved resources. Through the same interface, the ASE can query the available resources in specific target access networks ensuring that the MN will handover only to access networks which can support the active communication.

  Additionally, the resource control interface receives events on the resource modifications for the mobile device which act as triggers for access selection decisions.

- Mobility Control Interface - this functions is able to receive information from the core network mobility control and to transmit indications on the access network selection decisions.

- Client Interface - this function is enabling the communication to MNs, including authentication and authorization for using the ASE service, parameter transmission for decision requested from the MN, and transmission of access selection decisions to the MN.

- Access Network Coverage Map - this database stores information on the status of the access networks, their location and specific parameters which enable the MN attachment. The database is updated directly by the network management system.

- Subscriber Information - the database stores specific ASE information related to subscribers, such as authentication and authorization information for the communication with the MN's self-adaptation management functionality/

- Decision Engine - the component which is triggered, gathers parameters and makes decisions resulting in access selection indications, and handover preparation indications to the mobility support in the core network. The decision engine may consist of a policy decision function or of an expert system. As the exact capacity of the decision engine is highly dependent on the specific network configuration and elasticity of the deployed system, no further considerations will be made.

The ASE has a functional counterpart in the MN, the recipient of the network decisions. This MN Self-Adaptation Management function is able to receive indications from the network on the most appropriate access networks to be selected in specific conditions. When these conditions are met, then the commands are transmitted to the control plane which executes the handovers through state of the art procedures. Therefore, the ASE does not have a direct handover control, enabling it to transmit to the MN indications, asynchronous to the actual handovers. ASE empowers the MN to execute handovers triggered by parameters which can be measured only at the MN and with no communication with the network during the actual handover execution.

The ASE can be extended to support MNs which are attached to multiple access networks at the same time. The same procedures are considered. However, a more granular differentiation is required, including the prior network identity of the subscriber and adding the description of the data flows for which an access network has to be reselected. This extension requires a MN forwarding extension enabling the routing of specific data flows through the appropriate device interfaces.

### 5.3.2 Information Model

In order to make the appropriate decisions and to transmit them in the form of indications to the MN and to other network entities, the ASE has to be triggered and to communicate with other network entities. In this subsection, the information which is exchanged by the ASE is presented.

To be noted, the ASE does not take the actual selection decisions. These are taken by the mobile device depending on the handover parameters in the specific access network such as the LTE Time-To-Trigger (TTT).

Figure 5.5 includes a high view on the specific triggers, parameters and indications which are considered. The system can function in a sub-optimal manner when part of the information is not available. The information is either fetched from different entities or it may be received as a trigger for a new decision. Therefore depending on how the information is received, it may be used as a trigger, as a condition, or as a parameter in the access selection decision.

The following parameters are received by the ASE from different other entities:

- Location Information - the location information is received directly from the MN which acquires it through other external mechanisms such as GPS or access network specific technologies. It enables the ASE to consider only the access networks located in the vicinity of the UE.

  The system does not require very precise location information, thus partial or low accuracy information is enough for the ASE decisions.

  In a more secure manner, although with a different granularity, this information can be received as part of dynamic subscription information or of the network location information from components which already maintain such information in current carrier grade core networks for other purposes. Obtaining the location information from the network provides a more energy efficient solution with lower delay.

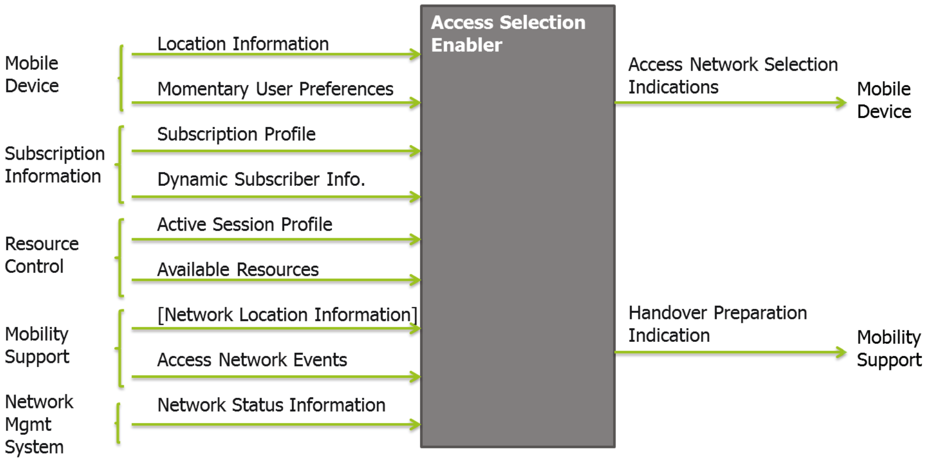


Figure 5.5: Access Selection Triggers, Information and Parameters

- Momentary User Preferences - the MN's user may modify the access selection preferences independently of the subscription information. This information is transmitted directly from the MN as part of the query for new access selection indications or as a result of a requirement from the network. The user preferences include a list of preferred access networks or radio technologies which is to be considered during the decision making.

  This information includes the capabilities of the MN to different access technologies. It enables to limit the decisions only to the access networks to which the MN is able to connect to.

  Additionally, this information can be extended with the Quality of Experience (QoE) perceived by the user for the active services. For example, a decrease in QoE may require a handover to another access network. This feature is not further detailed.

- Subscription Profile - the network subscription profile contains information on the access networks to which the subscriber is restricted or is allowed to connect to the QoS profile the subscriber is allowed to use over different accesses.

- Dynamic Subscriber Information - it contains the dynamic information which is maintained by the network and shared between the different control entities such as MN's momentary IP address or access network. For scalability reasons, this information is usually grouped with the subscription profile.

  Additional to this information, there is dynamic information through which the MN registers to the ASE such as the IP address and the port through which it is able to receive network indications.

- Active Session Profile - for each active service, the resource control maintains a session state which contains the description of the specific data flows enabling their differentiation and the allocated resources such as the QoS class, the guaranteed resources etc. During the decision making, information on the aggregated required resources, reserved over the source access network have to be supported by the available resources on the target access network.

  Additionally, when a change in the resources reserved for a specific subscriber happens, it may require an access network reselection decision. The change of the resources required is a consequence of a change in the state of the MN, like the passing from idle to active mode or a change in the active services through procedures such as the establishment, modification or termination of a service.

- Available resources - the resource control can transmit such information to the ASE for the access networks under consideration. It ensures that the target access network selected can support seamlessly the communication of the MN.

- Network Location Information - the mobility support maintains information on the current network location of the MN, such as the current base station or gateway used for the communication. This information can be transmitted from the mobility support to the ASE. However, it may be also contained in the dynamic subscriber information.

- Access Network Events - for adapting the decisions to the MN's momentary context, the ASE has to maintain information on the MN's reachability statues. However, as the ASE decisions are asynchronous to the actual handover procedures - i.e. the indications are transmitted at a prior moment of the actual handover, with no presumption when this handover will occur - this information is available with a certain delay from the Dynamic Subscriber Information.

- Network Status Information - ASE may require information on the available access networks at MN location. It is assumed that this information is introduced through administrative means independent of the operations of the actual decision and that a trigger for access selection may be pushed for all subscribers in a specific area e.g. for energy or maintenance purposes.

Using the information above, the ASE is making decisions which are then transmitted in the form of indications to MNs and to the mobility support of the core network.

For the MN, it transmits access network indications in the form of policies which have to be processed when a handover happens. Through this, the network operator is able to remotely manage the MN access network adaptation.

For the mobility support, the ASE transmits handover preparation indications which enable the target access network to execute a preparation procedure which reduces the duration of the actual handover.

### 5.3.3 Procedures

In this subsection, the ASE specific procedures are described. First, the internal state machines of the enabler is presented. Then, two exceptional procedures are detailed, targeting the use cases unique to this dissertation.

#### 5.3.3.1 Access Selection Enabler Internal Procedure

The following models for the internal ASE state machine are considered:

1. Simple Access Selection State Machine - the enabler does not maintain state information for the MN. Each time a request is received, it is treated independently of any other information. It can be used only when the MN pulls for information.
2. State based Access Selection State Machine - the enabler maintains a state for each MN. It triggers from other network entities and can push information to MNs. Additionally, the information may be cached until the mobile device queries it.

In the following, only the state based access selection state machine is presented, as depicted in Figure 5.6. The state machine is presented from the perspective of the ASE.

When a new MN contacts the ASE, such as in the case when the MN first attaches to the network, the MN will first select an access network based on its default information. Then it authenticates and receives the authorization to use the ASE service (State 1a). Based on the identity through which the subscriber authenticates, the ASE fetches the subscription profile from the core network subscriber data base, including the dynamic connectivity information (Step 1a-2). Based on this information, a subscriber structure is created within the ASE (Step 1a-3). The rest of the access selection procedures follows, starting from Step 4.

The procedure is also triggered by or for a MN which already authenticated and is authorized to use the ASE service. The trigger can be related to the expiration of previously received indications or due to location or user preferences change. In

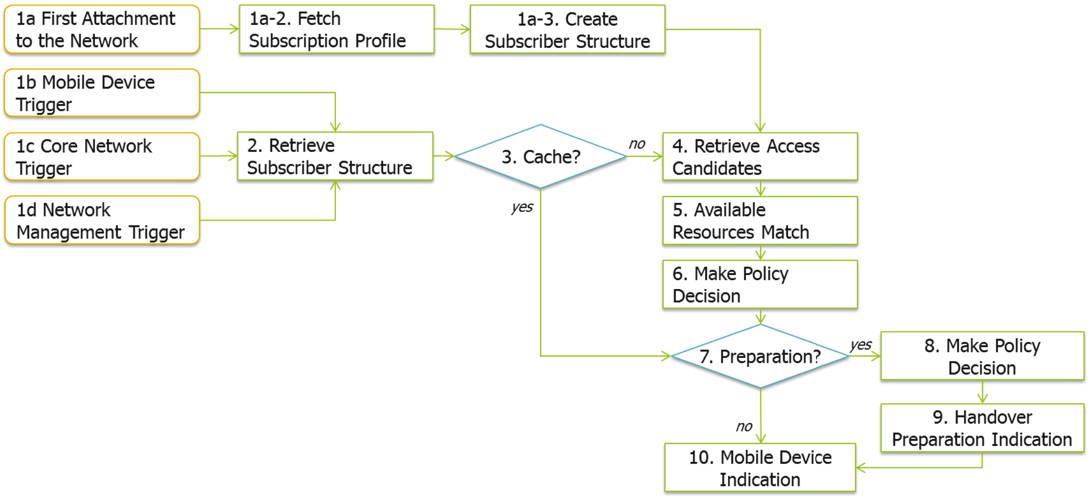


Figure 5.6: Access Selection Enabler State Machine

this case the mobile devices will request a new set of procedures containing the modified information (Step 1b).

The procedure may be also triggered by modifications in the core network, requiring an adaptation for the specific MN, including a modification in the subscription profile, reducing or extending the access networks to which the MN is allowed to attach to or by a modification of the required resources. Additional triggers from the core network include the modification of the available resources in the specific access network, requiring one or multiple devices to handover in order to avoid congestion over the specific access network (Step 1c).

The operator network management system may also initiate the procedures for example in case of failure detection of an access network in a specific area, energy efficiency and other self-optimizing network algorithms or in case of controlled maintenance procedures. In this case, the access network coverage map is modified and then the ASE procedures are initiated for the affected subscribers.

For all these triggers, the structure related to the specific subscriber is retrieved from the Subscriber Info database (Step 2). The matching of the specific information is done based on MN identity.

If the retrieved subscription profile is containing cached information, then the access network selection decision is not made again. Instead the cached information is transmitted in the form of handover preparation indications and as mobile device indications (Step 3).

If no information is cached, or if the cached information has expired, then the ASE retrieves from the coverage map all the possible candidate access networks. This represents a non-ordered list of accesses located in the vicinity of the MN(Step

4).

If the ASE is configured to ensure that the target access network has enough resources available to support the communication of the MN, thus ensuring that the handover is seamless to the active services, then the ASE retrieves the momentary used resources of the MN and based on this information, the list of potential candidates accesses is reduced to only those that can support the communication (Step 5).

Then, based on the remaining candidate access networks, a policy decision is made enabling the ordering of the access networks based on the internal operator parameters. The decision making is highly dependent on the access networks deployment and is not further extended (Step 6).

If the system is configured to execute a preparation phase (Step 7), then a new policy decision is made determining the access networks to which there is a high probability that a handover will be executed to (Step 8). If such an access network is determined, then a handover preparation indication is transmitted to the network entities which in their turn establish a volatile context on this potential target access network, as described in Subsection 5.3.3.3 (Step 9).

After the volatile context is established, then the indications are transmitted to the MN. Using this information, the MN may execute a handover to a target access network in specific conditions (Step 10).

In case the procedure was initiated from a network component and if information can not be pushed to the MN, then the indications are cached until the MN initiates the retrieval procedure. In this case, the handover preparation procedure is initiated only when the device requests the information.

#### 5.3.3.2 Resource Oriented Access Network Selection

The state of the art solutions are considering that the resources for a specific MN have to be reserved over the access network to which it is attached to. Stemming from single access solutions, the resource control functionality is not aware that the MN may attach to another access network at the same location and receive the service.

This dissertation introduces the concept of resource oriented access network selection: a modification on the resources required over the radio network by a specific MN changes the optimal access network over which a device can communicate.

This change presumes that a novel method of automatic adaptation of the radio connectivity through access network reselection is introduced, in which the modification of the resources required by a MN for the communicating data flows is considered as a trigger for an access network selection which may lead to a possible vertical handover decision with immediate execution.

A resource modification for one or multiple data flows in operator carrier grade systems is made through a session establishment, update or release procedure. These modifications are noticed by the Resource Control functionality of the core network.

In order to be able to include the proposed method as part of the operator carrier grade networks, the realization of the concept includes the development of a novel interface between the Access Selection and the Resource Control functions as depicted in Figure 5.7.

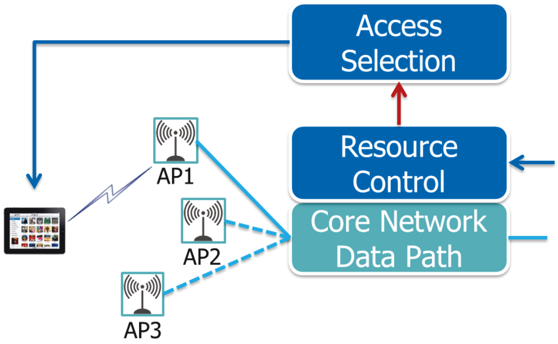


Figure 5.7: Resource Oriented Access Network Selection Concept

Using this novel interface, the following high level method for access network selection is proposed:

- **Step 1:** The Resource Control acquires knowledge of a session modification which requires changes of resources reservations. The change request may be received from the MN, the data path entities through an event notification, specific data traffic detection, or from the various service platforms to which the MN is communicating.
- **Step 2:** Based on the results of specific internal or together with data path entities operations, the Resource Control sends an indication to the Access Selection functionality. The operations may include the decision making and the installation of data traffic handling policies for the specific subscriber of the MN and for the specific sessions which requested the modification i.e. initiation, modification, release, aggregation or split of multiple sessions.

  The indication to the Access Selection may be send by the Resource Control before the initiation of the resource reservation on the access network to which the MN is connected to or after the resource reservation over access network has failed. In the first case, the reselection procedures are executed for all the required resources modification attempts, having a reduced delay than when it is executed only for failure of the resource reservation as in the second case. An intermediary possibility may be also conceived, in which the Resource Control announces the Access Selection to initiate the selection procedures based on some generic filter criteria

For a more informed selection decision, the indication may contain information on the requirements of the session and on the other active sessions of the MN.

- **Step 3:** The Access Selection makes a decision. In case a handover has to be executed, the target access is selected from the accesses located in the vicinity of the MN.

  The decision may be based on the subscription profile of the user of the MN and on the static and dynamic parameters of the access networks in the vicinity of the MN, including the location of the MN and the coverage area of the access networks.

- **Step 4:** In case a handover is required, the access selection indication is sent to the MN containing an immediate execution indication and information on the target access network. Using this information, the MN is able to execute a handover.

- **Step 5:** The Resource Control receives a notification from the data path entities that the MN has attached and is able to communicate over the target access network. Triggered by the notification, the Resource Control initiates the resource reservation procedures and notifies the initial resource triggering entity that the required resources are available. Now, the exchange of data for the modified data flows of the MN may be started or continued with the new required resources available.

The concept here differentiates from other state of the art solutions as it presumes that the resource modifications are triggering the adaptation procedures. The other solutions consider other triggers for the adaptation procedure. They may include as part of the access selection decision the check that the required resources can be reserved on the target access network. These state of the art solutions can be integrated as part of the selection decision from Step 3.

Because in various service architectures such as IP Multimedia Subsystem (IMS), the resource reservation is part of the application level session modification, the procedure here presented can be integrated also into the specific service level procedure. From this perspective, the here presented procedure extends the resources that are available for a MN not only to the level of the resources on the access network to which it is already connected to, but to the available resources located in all the access networks in its vicinity.

#### 5.3.3.3 Volatile-Shallow Context

This dissertation introduces the concept of Volatile-Shallow Context as a means for executing part of the handover related procedures proactive to the handover trigger. The concept relies on the separation between the access network selection and handover procedures and on the execution of the control related procedures before any handover procedure is initiated without installation of resource reservations on the data path.

The current literature considers two types of relationship between the access selection decision and the triggering of the handover procedures which logically follow the same steps as depicted in Figure 5.8.

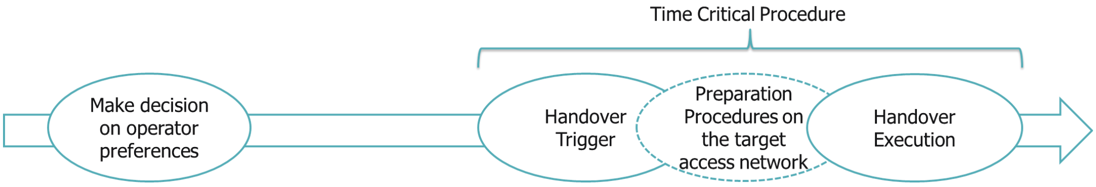


Figure 5.8: State of the Art in Handover Procedures

First class consists of solutions in which the access selection decision is independent of the trigger of the handover, like in 3GPP EPC. The Access Selection functionality makes the decision and sends the indication to the MN. When a handover is triggered in the MN, due to parameters which can be only locally measured i.e. signal strength or internal policies, the handover operation is executed. This implies that no proactive procedure is executed on the target access network before the handover is triggered, either actively or pro-actively.

In the second case, the network makes the access selection decision and transmits it to the MN only after the handover is triggered either by the signal strength measurements in the network or by the MN, such as in IEEE Media Independent Handover. In this case, all the procedures related to the handover are executed after the handover trigger. Even though the handover trigger may be proactive (e.g. based on a rapid decrease of the signal strength), all the procedures related to the handover are executed after the event happens, which from the perspective of this concept is similar to the previous case.

Thus, for both cases, only after the event of an imminent handover is available at the MN, the procedures of handover are executed which may contain a proactive or a preparation phase and an actual handover execution phase. The procedures for the handover have to be faster than the loss of connectivity to the source access network otherwise there is a service interruption. Even though information is already available in the network, it is not used before the handover trigger, especially because of the resources consumed by such a procedure in case the handover is not executed.

This dissertation proposes a novel method for the execution of a proactive phase prior to the handover trigger, by executing a reduced context establishment on the target access network which includes the transfer of all the information required to the specific entities, without enforcing any rules to the data path. This context is denominated as shallow context. It also considers a new interface between the Access Selection functionality and a correspondent entity in the target access network which establishes the shallow context.

The indication for the shallow context establishment in the target access network is transmitted when the access selection decision is made. The solution considers that the context is created due to the new information transmitted to the mobile device; independent of the moment when the handover trigger is received by the mobile device as depicted in Figure 5.9.

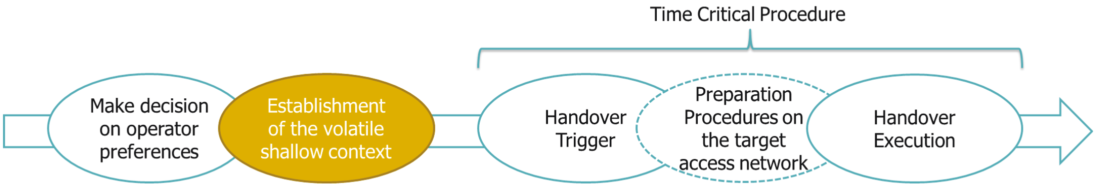


Figure 5.9: Volatile Shallow Context Handover Procedures

As the information of the volatile-shallow context may remain unused when the MN does not execute the handover to the specific access, for a smooth termination, the information has an expiration time, ensuring the volatility of the context. As the volatile-shallow context does not presume any resources to be reserved on the actual data path, the context can be silently removed, without requiring any signalling procedure.

In case the handover happens, the volatile-shallow context becomes the active context and the resources have to be reserved on the data path. As all the entities in the network already signaled each other and already store the required information, the time-critical part of the handover procedure is highly reduced.

- **Pre-authentication and pre-authorization** - The access network entities receive a specific authentication vector and its validation value. The authentication and authorization entity is not involved in the attachment anymore, the procedure being executed locally between MN and access network entity.

- **Location Update** - The dynamic information on the MN location is extended to contain a volatile one on the target access network in case no other location information is available. The volatile location is considered valid until its expiration.
- **Subscriber Data Retrieval** - The subscription profile is retrieved by the control entities, enabling a faster decision during the attachment.
- **Resource Requirements available on the data path entities** - The policy decision is made which resources have to be reserved on the target access network in case of a handover and this information is transmitted to the data path entities. This does not imply the actual reservation of resources which would be executed when the MN attaches to the network.
- **Mobility information available on the data path entities** - The target access data path entities receive information on how to forward the data traffic from and to the MN. However, this redirection to the target network does not become active until the MN attaches to the target network which can be determined through first upload data packet received on the uplink data path.

The procedure of bringing the information to the appropriate network entities on the target data path is reduced from the actual handover procedure. After the handover trigger, the mobile device will execute link layer and network layer attachment procedures. Due to the volatile-shallow context, all the required information is already available in the network entities, thus enabling a fast response to the attachment request and through this a reduced handover duration.

Additionally, the number of access networks where the volatile-shallow context is established is very small depending on the probability of the MN to execute a handover to the specific access network. For example, in case the MN is camped in a small coverage access network, such as a WiFi hotspot, and is allowed to execute handovers to the operator wide area network such as LTE, then the LTE access in the vicinity of the WiFi access may maintain a long duration volatile-shallow context expecting the user mobility and loss of signal in the WiFi access.

## 5.4 Core Path Adaptation Enabler (CPAE)

The CPAE aims at optimizing the communication of MNs through the adaptation of the core network connectivity, specifically between the base station to which the MN is attached to and the border of the carrier grade operator domain. It concentrates in the area of flexible reselection of operator entities and on the actual data path adaptation through the newly selected entities. Through this means, the carrier grade operator network is able to steer the data traffic on a subscriber level basis between the different core network entities at different geographical locations reducing the data path stretch and thus the communication delay.

The CPAE functionality is addressing distributed wireless core networks, in which multiple components having similar functionality are deployed in different geographical areas close to the MN's attachment point to the network. The functions map and extend the mobility support functionality of the operator towards the flexible support of the distribution characteristic through on demand dynamic adaptations. Through this means, a new mobility axis is introduced in the system, in which the MN connectivity is changing its data path characteristics independent or as a result of the actual mobility of the device in the network.

CPAE is triggered by specific operations which occur in the system such as related to the subscriber communication e.g. attachment or detachment from specific access networks, change of network location etc. or management commands such as the scheduled maintenance procedures or fail-over events. The Core Path Adaptation Enabler makes a decision based on the current status of the network and adapts the communication for the subscribers. The indications on the core path adaptation are transmitted to mobility support entities in the control and in the data plane such as base stations and mobility anchors and other forwarding entities which pertain to the operator domain as described in the architecture subsection.

CPAE can support multiple data path for a single device using one or multiple IP addresses and using one or multiple wireless connections. For this, the granularity of the decisions is made at a subscriber level and additionally at a data flow level, which data flows are described by the end-to-end communication five-touple - the source and the destination IP addresses and ports and the protocol number in case of IP communication. In order to be able to enforce the adaptation decisions on the data path entities used by the subscriber, these entities have to be able to forward data traffic at flow granularity level.

### 5.4.1 Architecture

CPAE is composed of a set of functions which enable the communication with the control and operator management functions through which parameters are gathered and the orchestration of the adaptation procedure is executed, while maintaining transparent reachability of the MNs. For each of the subscribers, CPAE maintains a specific state and a list of the entities on the data path of the MN.

Similarly to ASE, CPAE contains a set of interface functions enabling it to

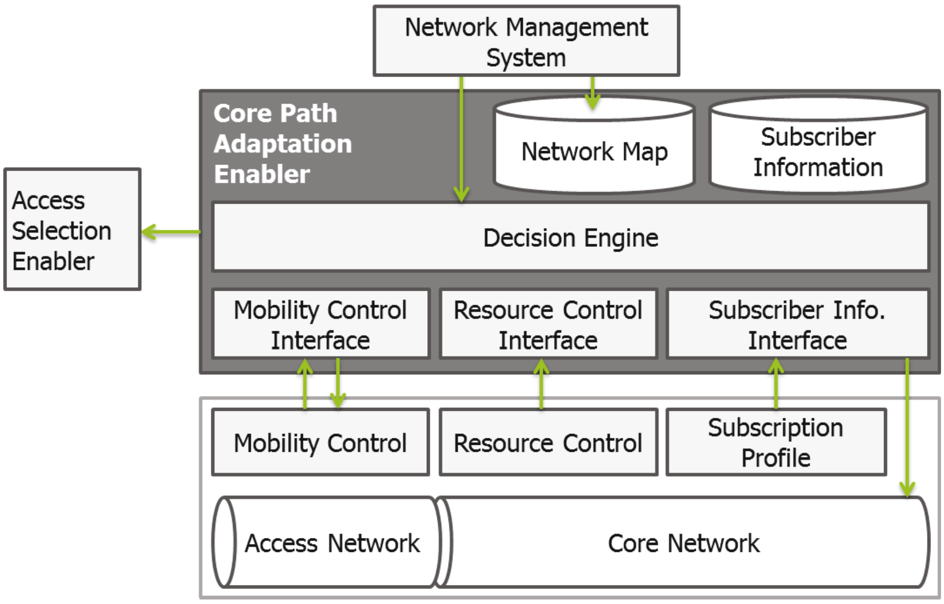


Figure 5.10: Core Path Adaptation Enabler Architecture

adapt and to control the communication with other entities, a set of databases storing subscriber and network related information, and a Decision Entity which is triggered, gather parameters and decides and orchestrates the data path adaptation. The architecture is depicted in Figure 5.10.

For communicating to the core network, three modules are considered: a Subscriber Information Interface for retrieving subscription profile information, a Mobility Control Interface for communication with the core network mobility support including the adaptation enforcement and a Resource Control Interface for the communication with the resource control entities in the network.

Additionally, the CPAE includes a network entities map, enabling the appropriate adaptation decisions to be made and a Subscriber Information maintaining the bindings between the subscribers and the specific data path.

As the core path adaptation is limited in scope to a single operator domain, CPAE operations are transparent to MNs and to CNs and their network provider domains.

The decisions of CPAE are enforced on the mobility control and data path entities. Depicted in Figure 5.11, three different enforcement entities are considered, depending on their relationship with the IP domain and with the MN, as follows:

- Access Point (AP) - the AP represents the point of attachment of the MN to the network. Whenever a device is attached to an access network, there is one AP which is allocated. This can be represented by a Base Station of

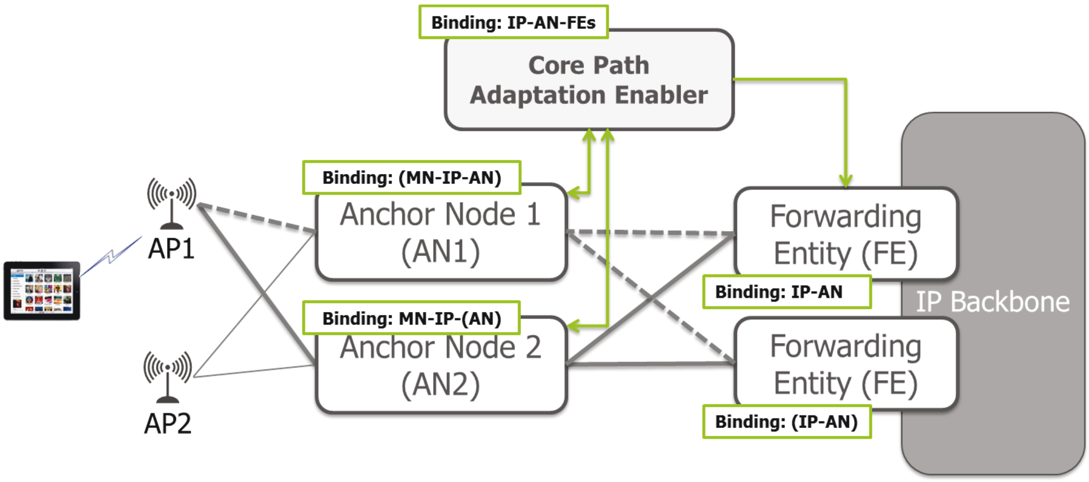


Figure 5.11: Core Path Adaptation Enabler Network Architecture

a specific technology type, aggregation router etc. The AP is not changed during the core network adaptation procedures. If a change of the AP is required, then, prior of the core path adaptation, a command is transmitted to the ASE which triggers the MN to change its access network.

- Anchor Node (AN) - the AN node is handling the mobility of the device in a specific area of the network. The AN may be changed due to a change in the location of the MN, due to the requirement to connect to specific local services or due to administrative procedures such as failure mitigation, energy efficiency or programmed maintenance procedures.
- Forwarding Entity (FE) - the FE is a node located in the IP domain of the operator between the border of the operator domain towards other operators and the anchor nodes. The role of the FE is to direct the data traffic to the appropriate AN of the MN, according to the data path adaptation procedures.

These three functions are topologically located in the operator network and have different deployment characteristics. The nodes located closer to the MNs, such as access points are serving less nodes than the ones which are located closer to the border of the operator domain, towards the service platforms or the Internet. However, due to the physical location changes more handovers are required in the entities located closer to the MN.

The adaptation of the data path consists of the seamless handover of the anchor node of the MN and the modification of the data path through the modification of the forwarding rules in the APs and FEs.

### 5.4.2 Information Model

In this subsection, the information exchanged by CPAE with other entities is presented.

For each of the subscribers the various entities on the data path and the CPAE maintain a binding entry enabling the forwarding of the data traffic towards the appropriate entities. The binding information is depicted in Figure 5.11.

- CAPE maintains all the specific data path locations of the devices for each of the MNs. This information includes location information which may be expressed in the form of the IP address as allocated to the MN and in form of data path entities identities. The data path entities considered are the Anchor Nodes through which a data path is established in the core network.

  Additionally, in specific deployment situations which consider synchronous data path adaptation procedure as described in the next subsections, contains information on the FEs which had to transmit downlink information in a predefined time interval and optionally on the APs to which the device is attached to. This mechanism enables the direct notification of these entities as described in the next section.

- The Anchor Nodes maintain bound to the identity of the MN the forwarding mechanisms towards the MN, through state of the art protocols such as Mobile IP, Proxy Mobile IP or GPRS Tunneling Protocol (GTP).

  An Anchor Node may cache for a specific time duration information on the currently used Anchor Node of a MN. This enables the notifications of the FEs which send downlink information and the redirection of the data traffic towards the current Anchor Node.

- A FE binds the MN's IP address to the current Anchor Node. This information is received from either one of the Anchor Nodes or directly from the CPAE, as presented in the following procedure subsections.

From the information perspective, the data path adaptation procedure represents the efficient update of this information, when required, based on the decision taken by CPAE.

Figure 5.12 includes a high view on the specific triggers, parameters, and the decision propagation mechanisms for CPAE. The system includes as decision parameters the same information in all the situations. In case one of the triggers is not available, then the specific operation is not sustained by the specific installation of the system. For the enforcement of the data path adaptation decision, two solutions are considered: one active and one reactive to the data traffic of the MN.

It is required that only one of them is installed in a specific system. The selection of the most appropriate solution depends on the scale of the system.

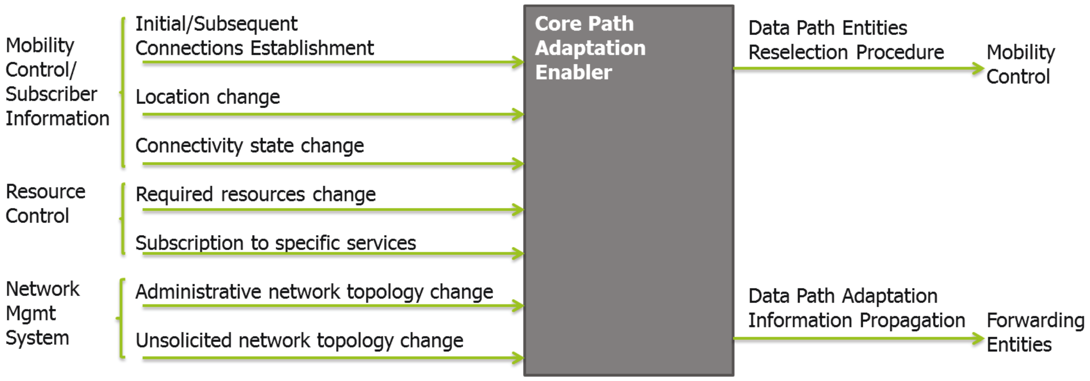


Figure 5.12: Data Path Selection Triggers, Information and Parameters

The following types of triggers and parameters are received by CPAE from the different other entities:

- Initial/Subsequent Connection Establishment Information - when the MN initially attaches to the network, it is assumed that it requires to communicate and to receive messages from CNs. For this, a decision is made and an anchor node is allocated. When the MN requires additional connectivity parameters, such as a second IP address for communicating with other services, then a second data path is allocated through the same or a different Anchor Node. This trigger may be received from the mobility control entities in the operator core network or as notifications on changes of the connectivity part stored together with the subscription profile

- Location Change - when the network location of the MN changes, mainly due to physical mobility, then another anchor node located closer to the new location may serve more suitable the communication of the mobile subscriber. This information is dynamically stored together with the subscription profile and available to the mobility control entities. It is assumed that cell granularity location is required during the anchor node selection.

- Connectivity State Change - when the MN passes from idle to connected state, it signals to the network that a data path is required for the communication. In this case, an anchor node selection procedure is triggered, considering the network location of the MN and other information available in the network subscription profile.

- Required Resources Change - when a device requires resources for one or more data flows the specific communication may be better served through a separate data path with a shorter stretch or with reduced delay or costs for the operator. This information is received from the Resource Control. This type of trigger is used when the data paths used by the MN can be split or merged based on the data flows they are composed of. This functionality requires a high level of network resources and may be missing from specific deployments.

- Subscription to Specific Services - when the MN registers to specific operator dedicated services such as IP Multimedia Subsystem (IMS), then a separate data path may be allocated for optimizing their delivery. This information may be received from the Resource Control when the service platform or the MN initiate the resource reservations for the specific service.

- Administrative Network Topology Change - for executing specific management operations, such as scheduled maintenance or Self-Organizing Network (SON) operations like load balancing or energy efficiency, the CPAE has to first adapt the data path for the affected subscribers. For these specific triggers, the enabler selects new anchor nodes considering the final topology of the network. From this perspective, the anchor node selection is enabling the dynamic execution of SON functions.

- Unsolicited Network Topology Change - in this category enter the unsolicited failure of specific nodes. Based on this trigger, CPAE selects a new data path for the subscribers excluding the affected nodes. This mechanism supports the fail-over mitigation for individual subscribers.

Using the information above, CPAE makes decisions which are then transmitted to the mobility control data path entities and to the IP FEs using specific mechanisms described in the next subsections. Two types of enforcement are considered for the data path adaptation.

First, the data path is adapted through the selection of the appropriate entities in the operator core network supporting the connectivity of the M. Specifically, the data path is adapted in the core network of the operator between the base station to which the device is attached to and a target anchor node. The enforcement operation presumes the notifications of the entities involved into the redirection of the data path through the target anchor node and depend on the architecture deployed.

Secondly, for the downlink data packets to be able to reach the newly selected anchor node, the operator FEs, located between the border of the operator domain towards the Internet and the anchor nodes, have to be informed on the adaptation. Two mechanisms for the propagation of this information are considered:

- Synchronous Data Path Adaptation - when the target anchor node is selected, the information is transmitted to all the FEs which previously requested such information due to downlink data packets.
- Reactive Data Path Adaptation - after the target anchor node is selected, when a FE transmits data packets towards the source anchor node due to having no information update, the source anchor notifies the specific FE, which in its turn adjusts its internal information for forwarding data traffic to the target anchor node.

### 5.4.3 Procedures

This subsection includes three specific procedures of CPAE. First, the adaptation procedures towards the network are presented for synchronous and reactive data path adaptation. Then, the reselection procedure for the components in the core network is described.

### 5.4.4 Synchronous Data Path Adaptation

In order to support the data path adaptation transparent to CNs, the current operator core network solutions consider that for each mobile device a data path is strictly bound to a set of network entities. For example, in case of 3GPP or IETF, when the mobile device receives an IP address during attachment, at the same time a binding is established on the data path through a set of redirection tunnels between the MN and the PDN GW or Home Agent. For the rest of the duration of the connection, the data path has to pass through that specific set of anchor nodes. This feature limits the seamless data path adaptation possibilities as the specific IP address can be reached only through the communication over the specific anchor node.

For transparently changing the anchor node, a new method for flexible on-demand data path selection is introduced, enabling through synchronous signalling the modification of the forwarding rules on the data path downlink towards the MN. The uplink reachability is not considered an issue as the CN may be reached through destination based routing.

As illustrated in Figure 5.13, the abstract network model is formed from two Anchor Nodes (ANs) between which the data path of the MN is adapted, CPAE which makes the decisions and maintains the information on the Anchor Node to which the device is currently communicating through and a set of FEs located at the border of the operator domain.

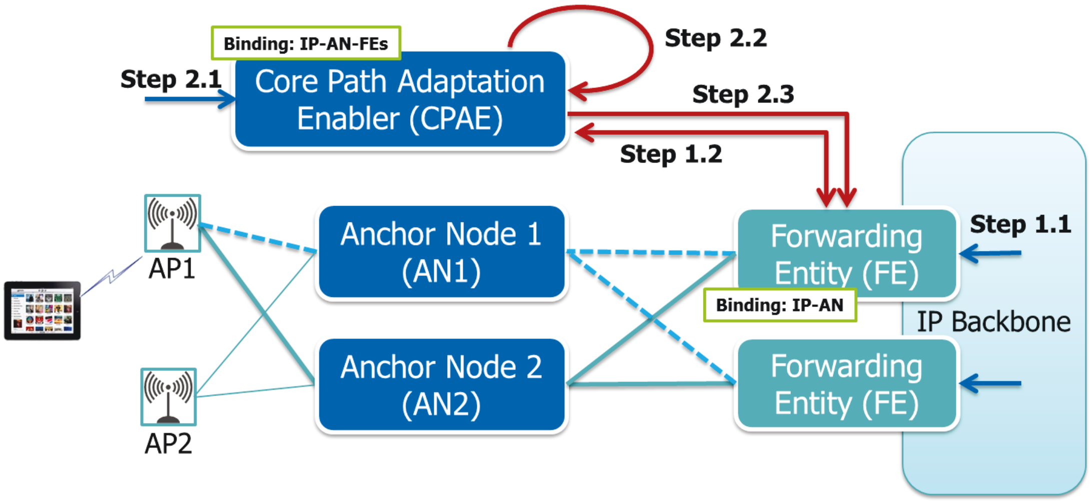


Figure 5.13: Synchronous Data Path Adaptation

It is assumed that the downlink data packets addressed to the MN are forwarded through the Internet Backbone based on inter-operator Service Layer Agreement (SLA) rules with no consideration to the mobility of the specific device. Therefore, a data packet addressed to the current IP address of the MN can be received by any of the FEs of an operator domain.

In order to reach the MN, each FE maintains a binding list, containing bindings between the IP address of the mobile device and its current anchor node. The binding entries are created only when a downlink data packet is received (Step 1.1). Each binding entry contains an expiration timer, in order to be removed without signalling in case it is not needed any more (i.e. no downlink packets for the MN) or to be updated when still necessary.

The information is queried from the CPAE functionality (Step 1.2), which enables in its turn the CPAE to be aware which FEs send in the last time interval downlink packets for the MN. For this information, the CPAE also maintains a binding between the IP addresses of the MN, the FE identities which queried information and the current anchor node.

An anchor node reselection may be required due to mobility reasons - i.e. a shorter data path through the operator core network can be achieved - or from administrative reasons - i.e. failure mitigation, energy consumption optimization or scheduled maintenance procedures (Step 2.1). This information is received by CPAE from either the source or the target Anchor Node or from other network or administrative entities. This concept is not concerned with the data path selection triggers which are integrated into a following subsection.

When the information is received by the CPAE, the binding table is checked and the FEs which had downlink data packets in the last time interval are determined (Step 2.2). An update message is sent only to the determined FEs (Step 2.3), enabling them to update their forwarding table and adapt the downlink data path through binding the new anchor node to the MN identity.

Through the proposed method, there is no strict binding maintained between the MN and a data path entity which enables on-demand data path adaptation based on available network information such as current load, MN mobility etc. The adaptation is transparent to the MN and to its CNs and relies on the notification of the changes only to the number of FEs which have forwarded downlink data.

Additionally, with specific extensions to the CPAE and FE nodes, the simultaneous communication through multiple anchor nodes for different data traffic can be realized.

Considering that the signalling is limited to a set of operator controlled entities, although for a large set of IP addresses, the solution is manageable through the current signalling mechanisms. A further evaluation of the solution is given in the next chapters considering the system design and specification as well as the testbed realization.

### 5.4.5 Reactive Data Path Adaptation

Additional to the synchronous data path adaptation, described in the previous section, a reactive data path adaptation is further presented in this section. The reactive data path adaptation is further constructing on the concept of on-demand mobility anchor selection by further integrating data path forwarding and signalling between the source and target Anchor Nodes (ANs) and the Forwarding Entities (FEs).

As depicted in Figure 5.14, the abstract network model is formed from the two Anchor Nodes (ANs) between which the data path of the MN is adapted and a set of Forwarding Entities (FEs) located at the border of the operator domain.

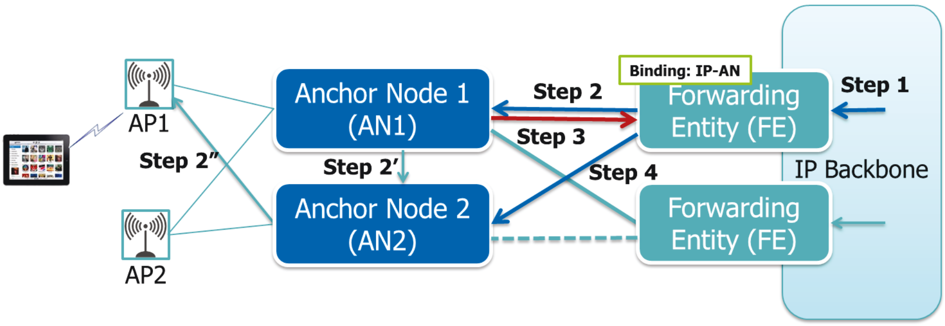


Figure 5.14: Reactive Data Path Adaptation

The same assumptions as for the synchronous data path adaptation are considered here. First, the downlink data packets addressed to the MN are send between the different operator domains with no consideration on the MN's location. Second, the FEs are receiving indiscriminately the data packets addressed to the MN.

Similarly to the previous concept, each FE maintains a binding list containing a binding between the IP address for which downlink data traffic has to be forwarded and the current MN's Anchor Node. Therefore, packets are identified by destination IP address and have as locator the AN identity.

This dissertation introduces a new method of reactive data path adaptation in which the FEs which forward data traffic towards the MN are notified by one of the anchor nodes that a reselection was already executed and that the binding information has to be updated. The update of the binding list is reactive to the forwarding of data traffic towards the MN.

The following reactive adaptation method is proposed. When a data packet addressed to the MN is received an FEs, then the FE checks its internal binding table (Step 1).

As the binding table contains the source Anchor Node identity bound to the IP address of the MN, the FE will forward the data packet to the source Anchor Node (Step 2). It is assumed that all IP addresses in the mobile operator domain receive a default Anchor Node, through a bootstrapping procedure. In case the binding table does not contain any entry, as no binding was previously created, the data packets will be forwarded to the default destination.

The source Anchor Node is not responsible anymore for operator controlled forwarding of the data packets to the MN. It will send the data packet to the target Anchor Node (Step 2') which in its turn will forward it to the mobile device (Step 2"). It may be possible that the source Anchor Node is not aware of the identity of the target Anchor Node. In this case, it has to query CPAE.

As the source Anchor Node is aware that the data path should be adapted through the target Anchor Node, it transmits to the FE which transmitted the initial downlink data packet, a control message containing a binding entry modification for the MN's IP address and the identity of target Anchor Node (Step 3).

Based on this information, the FE is able to update the forwarding table with the current location of the MN's IP address. Then, all the data packets will be forwarded to the target AN (Step 4), which delivers them to the MN.

As the FEs are transmitting the data packets to the ANs towards the MNs, each FE can maintain its own version of the forwarding table. Only the FEs which have sent a downlink data packet, addressed to the specific IP address, will have updated binding tables. The rest will maintain older versions which may be inaccurate.

Also as the FEs are advertising towards the IP backbone the complete set of IP addresses of the mobile operator, the binding list information will not be propagated towards the public domain, limiting the adaptation signalling to the FEs.

The binding entries maintained by the FEs are reactively update upon the existence of the downlink data traffic. Compared to the synchronous method, the number of the FEs which require an update of the binding list is reduced due to the reactivity. In case an FE does not forward any data packet to the MN, then no signalling will occur.

Additionally, the adaptation is maintained with zero packet loss and transparent to the MN and to CNs through the redirection mechanism between the Anchor Nodes.

Two deployment cases are considered: one in which the FEs and the Anchor Nodes are sharing the same local network. In this case the binding list maintains the role of routing table for the downlink data packets from the FE to the Anchor Nodes. Second case, addressing better current network deployments, the forwarding from the FEs to the Anchor Nodes is executed using an overlay which is independent of the IP transport e.g. GTP. OpenFlow, VLANs. In both cases, the impact of the forwarding is limited to the specific entities and does not propagate information in the rest of the IP domain, making the signalling manageable by the overall IP system.

### 5.4.6 Flexible Reselection of Data Path Entities

Using one of the previous two concepts, a carrier grade network operator is able to adapt the downlink data path towards one of the Anchor Nodes serving a MN. It is assumed that by using common IP forwarding mechanism, the data traffic which is generated by the MN will follow a data path which is not extended by the mobility support i.e. data packets are forwarded to the target destination depending on the destination IP address using destination based routing. Therefore, by forwarding the data traffic between the MN and an appropriate anchor node in the operator domain, the data traffic of the subscriber for both the uplink and downlink between the operator gateways and the correspondent parties.

To fully use of the features of data path adaptation, the core network has also to support the reselection of the data path entities, in a controlled manner as to ensure the consistency and the efficient network operation for the resource reservation and charging as well as for accounting and for access network horizontal and vertical handovers.

More specifically, a mechanism in which the anchor node is reselected with zero packet loss and without data traffic duplication while maintaining the adaptation transparent to the communicating entities.

This dissertation introduces two methods for the reselection of the data path entities addressing specifically the anchor node reselection. The two methods are considering two different signalling paths, one through a separate control plane and one parallel to the data plane. The two mechanisms are depicted in Figure 5.15.

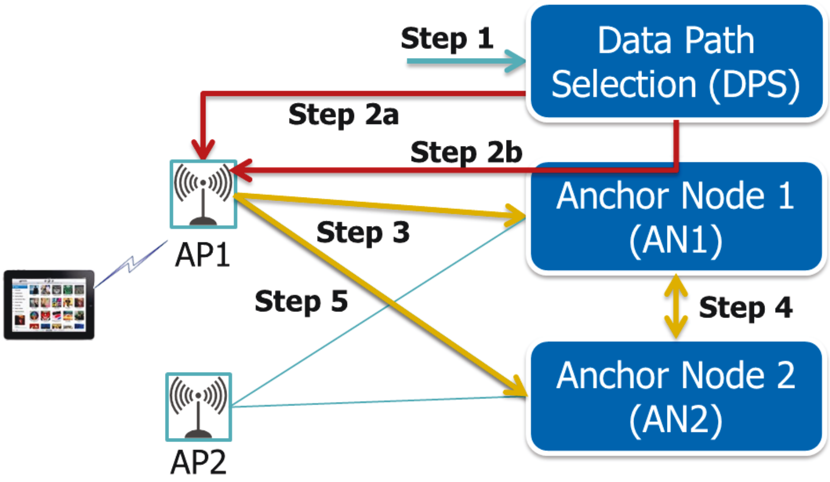


Figure 5.15: Flexible Reselection of Data Path Entities

An anchor gateway reselection decision (Step 1) may be triggered by multiple factors, like the MN or CN initiation of communication, through the MN passing from idle to active mode, or through the notification that a downlink data packet for the MN was received. Additionally, an anchor node may be reselected due to the mobility of the MN. Other triggers include the initiation of administrative or failure mitigation operations.

The specific information is transmitted to the current Radio Access Point node to which the device is connected to e.g. to the eNodeB for LTE or to the SGSN in case of UMTS. These nodes, although being reselected due to mobility, remain the same during the core network data path reselection procedure, as no radio network change is necessarily implied. The transmission of the message can be done through two mechanisms: directly to the Radio Access Point (Step 2a) or through the source anchor node (Step 2b). The first method follows similar path to currently standard mechanisms for initial data path entities selection while the second enables a mobility protocol oriented control. However, in the second case, a mechanism to announce the source anchor node of the reselection is needed(Step 3).

The second method can be used by extending a current mobility support. For example, Step 2b can be represented by a mobility tunnel termination request. However, compared to the current mobility termination requests available in the literature, such as for Mobile IP or GTP, this request includes also information on the redirection towards the target anchor node enabling the radio access network to be informed also on where the new mobility association has to be executed.

For both methods, in order not to lose any data packet, a reselection tunnel is established between the source and the target anchor node similar to the one established between the eNodeBs in case of a horizontal access network selection in LTE (Step 4). This tunnel enables the forwarding of the data packets from the source anchor node to the target access node during the reselection procedure. As described in the reactive data path adaptation, the tunnel may be replaced by a forwarding mechanism which enables a simplified state to be maintained on the target anchor node i.e. it does not require to maintain information on the source anchor node.

When the Radio Access Point entity reselects the target anchor node (Step 5), then the data traffic for the downlink will be received through the new link. When the downlink traffic is addressed to the source access gateway, then the data traffic will be directed through the tunnel established in Step 4.

This mechanism here presented comes to complete the initial selection of data path entities from the current core network architectures. It enables the reselection of the entities when required, transparent to the MN and to its CNs.

The capability of reselection of the anchor nodes, introduced by this concept, enables the core network operator to adapt the data path entities of a MN using parameters such as the subscription profile of the MN, the available and the required resources, mobility patterns and characteristics, etc. The reselection decision is made for each subscriber separately, considering the momentary cost, resources, load and mobility, thus being able to dynamically adapt to the current network context.

The method here presented enables the operator to dynamically control the data traffic enabling advanced solutions of traffic steering. This feature becomes highly important with the increase of the data traffic when the congestion and cost increasing points will be at the border of the operator domain.

## 5.5 Application Adaptation Enabler (AAE)

The Application Adaptation Enabler (AAE) aims at optimizing the core network procedure for adapting to the network context of active applications. The optimization consists in reducing of the number of messages and information exchange between the network entities and over the wireless link with the MN. It includes the elimination of the negotiation procedure between the MN and the specific service platform which was used for setting up of the parameters for the updated communication and the reduction of the messages exchanged in the network for the resource reservation procedure. Through these means, the carrier grade network is able to adapt the resources during the handover to the target access network without requiring MN or service platform interaction which highly reduces the duration of the adaptation procedure while requiring less functionality per subscriber and per application in the network.

The AAE is addressing the applications which are delivered with different communication parameters when connected to different access networks, for example the applications that require different throughput or even service interruption while connected to a specific access network while requiring another throughput on other accesses. The AAE maps and extends the integration between the mobility support and resource control functionality and optimizes the communication with the service platforms.

The role of the AAE is limited to a single operator core network enabling the communication over heterogeneous accesses. It is designed to be used independently by the different service platforms and applications. However, the aggregation of the multiple services may prove further processing optimization.

The AAE maintains information on the multiple levels of resources, which may be reserved for a specific data flow, when connected to different access network. This information is send by the applications platforms through a service provisioning procedure as described in the next sections.

When a handover to another access network happens, the AAE retrieves the provisioned profiles and checks which is the most appropriate to be installed on the specific access network. The resources for the selected profile are then reserved on the core network. An additional notification is sent to the service platform in order to adapt the actual data downlink transmission. It is presumed that the MN is capable by itself to determine the access network to which it is attached to and adapt the uplink communication.

### 5.5.1 Architecture

The AAE is composed of a set of functions allowing the provisioning of the information related to the service delivery and the actual adaptation of the resource reservation parameters based on triggers received from the control plane of the carrier grade operator network. The architecture is depicted in Figure 5.16.

For each of the subscribers and for each of its active services, the AAE maintains an information entry, containing the profiles which can be reserved while communicating over specific access networks, as part of a specific Subscriber Information storage. This information is provisioned from the services through a

Service Platform Interface component. The same functional entity of the AAE is used also for transmitting to the service platform notifications on the modification of the current level of resources reserved for the specific application delivery.

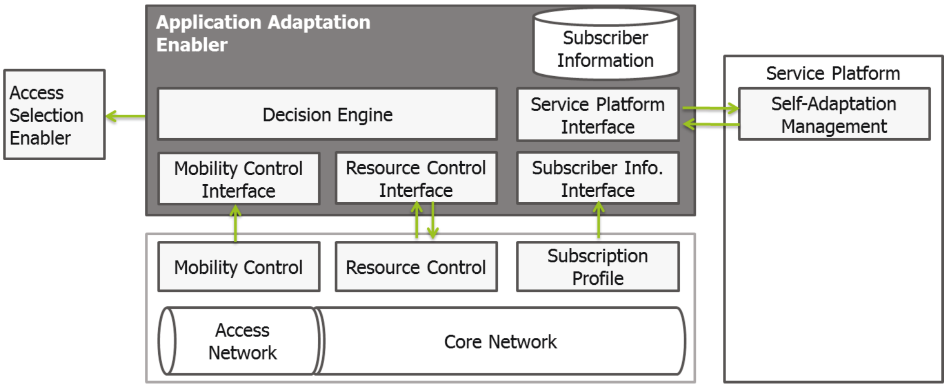


Figure 5.16: Application Adaptation Enabler Architecture

Similarly to the previously described components, the AAE, includes a set of interfaces towards the core network enabling the gathering of information and the adaptation of control plane specific parameters. Three modules are considered. First, a Subscriber Information Interface is used for retrieving subscription profile information enabling the adaptation decision to consider the resources which are allowed for the mobile device over the specific access network. Second, a Mobility Control Interface is used for retrieving the current access network used by the mobile device and handover events. Finally, a resource control interface transmits the currently reserved resources information and receives commands on how the resources have to be adapted when the access network changes. In a practical implementation, these three interfaces may be combined into a single one, as the handover events and the subscription profile are used also as part of the resource reservations.

As central execution of the AAE, a Decision Engine decides which level of resources has to be reserved on a target access network when a handover happens considering the list of possible resource profiles received from the applications and the subscriber profile information from the core network.

### 5.5.2 Procedures

This dissertation introduces the concept of session delivery adaptation based on access network context change, as a set of procedures for the service platforms and the carrier grade core network to dynamically and with reduced signalling and delay adapt the service delivery parameters as well as the reserved resources in the access and core network.

The current state of the art separates the service adaptation from the changes of the access network context. The service modification procedures are now triggered solely by the MNs, either at the request of the user either automatically by re-negotiation of the service parameters. When the MN's access network context changes, e.g. a vertical handover to another access network with different characteristics, the MN initiates on its own the service modification procedures including the resource negotiation with the service provider. This is followed by a new resource reservation procedure in the access and core network.

The concept of session delivery adaptation proposes a novel method for service parameters and resources reserved modification in which the actual service adaptation decision is included in the decision for resource reservation executed during the change in the network context. A special case of the network context change considered is the handover between access networks of various technology types (vertical handovers). As depicted in Figure 5.17, the method presumes a functional addition to the interface between the service providers and the resource reservation functionality in the core network and two procedures: one executed during service establishment and one executed when the network context is modified.

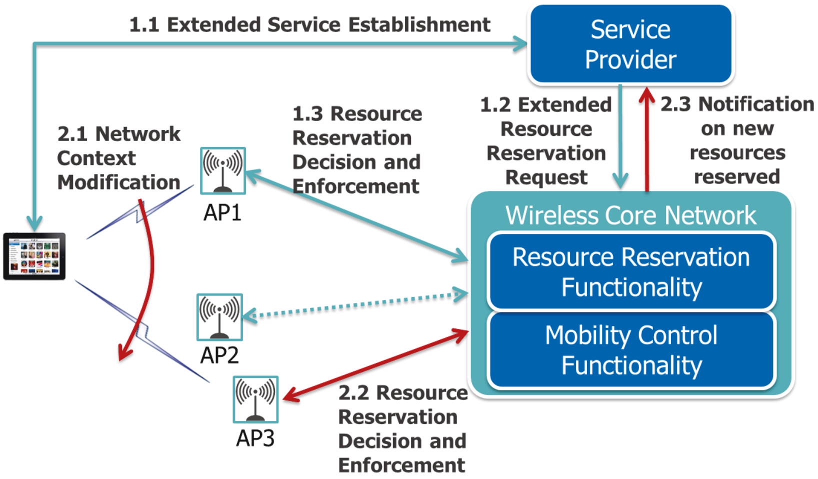


Figure 5.17: Session Delivery Adaptation

Using these extensions to the architecture, the following general procedure for service establishment is considered:

- **Step 1.1** - The MN initiates the service establishment procedures with the service provider. During the negotiation multiple profiles on which the session may be delivered are determined. The profiles of the session may contain different levels of throughput, maximum and guaranteed bit rates, codecs, admissible delays etc. From these profiles one is chosen as a primary profile to be enforced on the current communication path. The rest are considered secondary session profiles.

- **Step 1.2** - The service profiles are transmitted to the resource reservation functionality. Additional to the state of the art, also the secondary session profiles are transmitted to the wireless core network. This enables the resource reservation functionality in addition to the already existing decision making and enforcement of the resources based on the primary session profile, to store the other session profiles which may be used in case the MN's network context changes.

- **Step 1.3** - A policy decision is made whether the primary level of resources can be supported by the communication path followed by the policy enforcement.

  No new messages are exchanged neither between the mobile device and the service provider nor between the core network and the service provider which makes this procedure to introduce a minimal delay to the overall service establishment, due to the storage of more than one session profile in the resource reservation functionality.

When the MN's network context changes, the following novel procedure is executed:

- **Step 2.1** - The MN's network context change can happen because of a vertical handover or because the wireless access or core network become congested and the service cannot be offered anymore using the same session profile. This requires an adaptation of the resources for one or more data flows of the MN's active services.

- **Step 2.2** - When the resource reservation functionality of the core network receives the notification on the context modification event, it selects one of the secondary profiles that can be guaranteed for the service considering the new network context. This session profile is then enforced using the already standardized mechanisms. The selection does not presume a new policy decision which was already done during the service establishment. The decision is enforced on the data path.

  If the procedure refers to the case of an access network reselection, the enforcement on the target access network refers to the replacement of the QoS rules from the source access network directly with the ones referring to the new profile on the target access network, as part of the access network handover procedure. Through this procedure a single step resource reservation is executed during handover.

- **Step 2:3** - The service provider is notified on the level of resources that are currently reserved in order to adapt the active session to it. This notification may happen before the new session profile is enforced in case the resources allocated are lower than the resources of the previous profile or it may happen after the enforcement in case the resources allocated are higher. This allows that the data lost during the session adaptation to be minimal. No notification has to be transmitted for the adaptation of the upload data traffic as the MN is aware that the network context was changed e.g. the mobile device is the one that triggers a vertical handover.

  This notification replaces the service profile negotiation between the MN and the service platform. Through a single message transmitted from the core network to the service platform, the same results as in the case of a service negotiation procedure are obtained.

Although similar results are obtained using current service modification mechanisms, the method here proposed has a highly reduced number of messages exchanged over the network and part of the resource reservation procedures. During the adaptation, no negotiation between MN and service platform is required. Similarly, the number of resource reservation procedures is reduced from two to one, while no policy decision has to be made in the network as it is replaced by an automatic matching of the policies enforced on the data path with the access network to which the MN handed over to.

Through these procedures, the overall service adaptation is reduced while reducing the functionality in the network required for the handling at the cost of an extended information related to the services maintained in the core network. Such a trade off is considered better for the service provider and for the MN as they require less adaptation procedures while having a reduced adaptation delay as well as better for the core network, as the resource reservation procedures are executed for a reduced number of times per subscriber. However, this mechanism addresses only applications with deterministic characteristics over different access networks.

## 5.6 Enabler Interworking

The three different enablers which compose the SelfFit Framework for Access Selection, Core Path Adaptation and Application Adaptation address the adaptation of the different functions in the operator core network. As they all use similar control plane information and adapt specific communication parameters, it is assumed that there is no strict dependency relationship between their functions. However, some interactions are to be considered, as depicted in Figure 5.18.

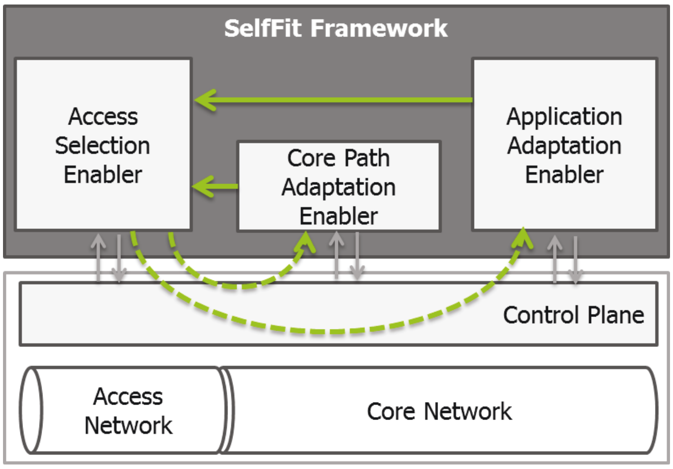


Figure 5.18: Interaction Between Enablers

The interactions can be classified as explicit interaction, in which one enabler transmits an indication or a command to another enabler, and implicit interaction, in which the result of the decision taken by one enabler trigger an operation of another.

Two explicit interactions are considered, as follows:

- The CPAE can require an adaptation of the access network previous or as part of the core network adaptation thus it requires that ASE makes a decision. This is employed in situations in which a more appropriate core path can be selected only for some access networks located in the MN's vicinity. In this case, in order to make a selection of new data path entities, an access network selection has to be executed. The specific trigger is introduced through administrative means in ASE, while the result of the operation will be received through an implicit interaction described underneath.
- The AAE can require in specific conditions, as indicated by the set of user profiles provisioned by the applications, that a level of resources has to be provisioned which can not be allowed by the network operator through the momentary MN's access network, thus requiring an access selection to be executed. In this case, the AAE transmits a reselection command to the ASE, including the required resources, similarly to the resource oriented access selection procedure.

Additional to the explicit interactions, two implicit interactions are considered:

- In specific situations, the access network reselection may be followed by the core path adaptation. The network location is modified as the result of an

ASE decision. This may trigger a CPAE decision. The two procedures are sharing only the location parameter, which is modified by the first enabler while being a trigger for the latter.

- An access network reselection represents the main trigger for the application adaptation. In all situations in which the ASE makes a decision, which is enforced on the connectivity of the device, the AAE has to consider for the specific applications which are registered with the enabler an adaptation procedure. Therefore, all the operations of the first, trigger the operations of the latter.

Through this interactions, the enablers of the SelfFit framework are able to communicate and influence between them their decisions. Other interaction may be possible, however not considered in this dissertation, due to the limited use cases foreseen in the future carrier grade operator networks.

## Chapter 6

# SelfFit Implementation

This chapter describes the prototype implementation of the SelfFit framework.

## 6.1 Introduction

Based on the specification from Chapter 5, the SelfFit concepts are applied on the 3GPP Evolved Packet Core (EPC) architecture, though the appropriate functionality and interfaces mapping and through the adaptation of the different information and mechanisms to the current existing features, such as the access network selection and the resource and mobility control.

Based on the mapping, a reference implementation was realized in the form of three distinct enablers: the Access Selection Enabler (ASE), optimizing the communication through the customization of the connectivity over the wireless links, the Core Path Adaptation Enabler (CPAE), optimizing the connectivity between base stations and border of the operator domain, and the Application Adaptation Enabler (AAE), autonomously customizing the resource reservations according to current access network capabilities.

The three SelfFit enablers are implemented as novel features on top of the OpenEPC toolkit, a software implementation of the 3GPP EPC standard, including the control and data path functionality for supporting connectivity over heterogeneous access networks [72]. For functional interoperability reasons, the SelfFit framework was developed in C language using the same fundamentals as the OpenEPC components. Parts of the features of SelfFit are included in the later versions of OpenEPC as innovation beyond the 3GPP standard, which is further disseminated through testbed deployments in vendor, operator and academia environments and base of research projects, such as FI-PPP FI-WARE [74], Univerself [222] and MobileCloud [223].

3GPP EPC mapping follows the principle of minimal modification. The current standard components are modified only through functional additions and when these functional additions are missing in a specific deployment, the system should function.

Following the same principle, the mapping considered the usage of already existing interfaces if available and suited for the purpose. For these interfaces it was tried, as much as possible, to only modify through addition the information transmitted while in some cases new messages are added.

## 6.2 Implementation Background: OpenEPC

For reducing the realization duration and also for enabling a better exploitation of the results, the SelfFit framework was implemented using as basis the Fraunhofer FOKUS OpenEPC software framework [72]. In this subsection, a very short general presentation of the OpenEPC toolkit is given including the development platform and the specific 3GPP EPC functionality implemented at the moment of writing of this dissertation.

The OpenEPC software is based on a programming framework, named Wharf, which aims at providing a flexible and powerful prototyping platform for developing software based operator core network functionality and components and in general the means to easily create testbeds for communication infrastructures. Wharf architecture is based on the architecture of SIP Express Router [224] featuring a highly modular design with C language written modules, redesigned for supporting a large number of heterogeneous communication protocols.

As illustrated in Figure 6.1, the OpenEPC implementation consists of a set of modules which can be combined in different functions and components customized based on configuration files which are introduced as input at the software initiation.

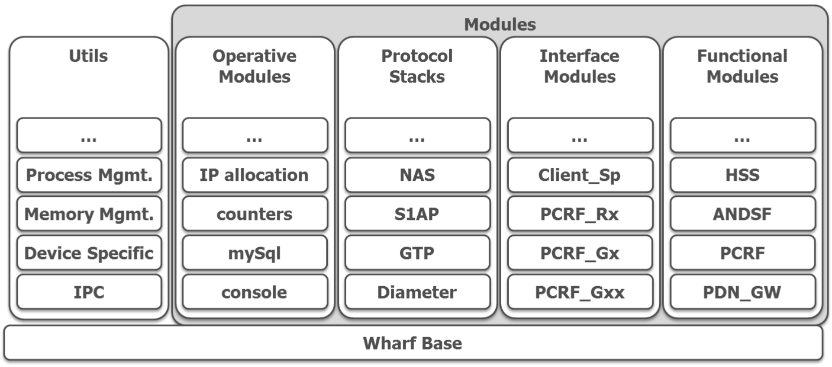


Figure 6.1: OpenEPC Wharf basis and modules

A set of utility modules are forming the basis for a generic type of software platform independent of the actual communication requirements, handling the process and memory control and the inter-process communication, optimizing specific critical features for the specific underlying hardware architectures, such as handling of locks.

Another set of modules enables to construct a communication basis on which the specific EPC functionality is developed. The operative modules include basic counters and communication with the databases, debugging and fast implementation console, and modules enabling basic core network functionality, such as the IP address allocation.

A large amount of the OpenEPC realization resources were spent in developing close to standards protocol stacks which support the 3GPP EPC such as Diameter, GTP, Mobile IP, NAS, S1AP, LLC, SNDCP, NS over IP etc. On top of the protocols a set of interface modules use the specific communication mechanisms to communicate based on 3GPP specified messages. Each of the implemented EPC functionality is represented by such a module enabling the fast updates along to the further development of the 3GPP standards.

In order to compose functions at a higher level and to enable the actual operations of the system, each component is represented by a functional module. From the EPC perspective, it is very easy to merge and to split two components by the unification or the separation of the code in the functional modules.

This base functionality was designed for re-usability and for development of novel protocols, which are or are not directly related to operator core network communication. For installing a specific component a set of modules is selected from the available list, dependencies are checked and the compilation is executed. Additional scripts allow for the automation of the network configuration and software installation, for reducing the complexity associated with OpenEPC deployments.

A large set of modules was developed realizing the OpenEPC toolkit which represents a software implementation of the 3GPP Evolved Packet Core (EPC) standards oriented towards the functionality in the areas of QoS, charging, mobility, security, monitoring and management

As illustrated in Figure 6.2, the current OpenEPC Rel. 3 includes all the components and a major set of the functionality and procedures part of 3GPP EPC standards such as harmonized authentication and authorization, mobility support, resource and charging control.

OpenEPC includes subscription profile based decisions for authentication and authorization over heterogeneous 3GPP (LTE, HSPA, UMTS) and non-3GPP (WiFi, WiMAX) access networks, for resource reservations and charging control, and for subscriber oriented access network discovery and selection. This is realized through a common Home Subscriber Server (HSS) which maintains the static network profile and dynamic information on the UE status, such as the current used IP addresses, location and selected data path entities enabling convergent mobility and resource reservation control.

OpenEPC includes a full solution for policy and charging control, for offline charging, and for access network discovery and selection. The Policy and Charging Rules Function (PCRF) is able to make policy decisions based on the requirements

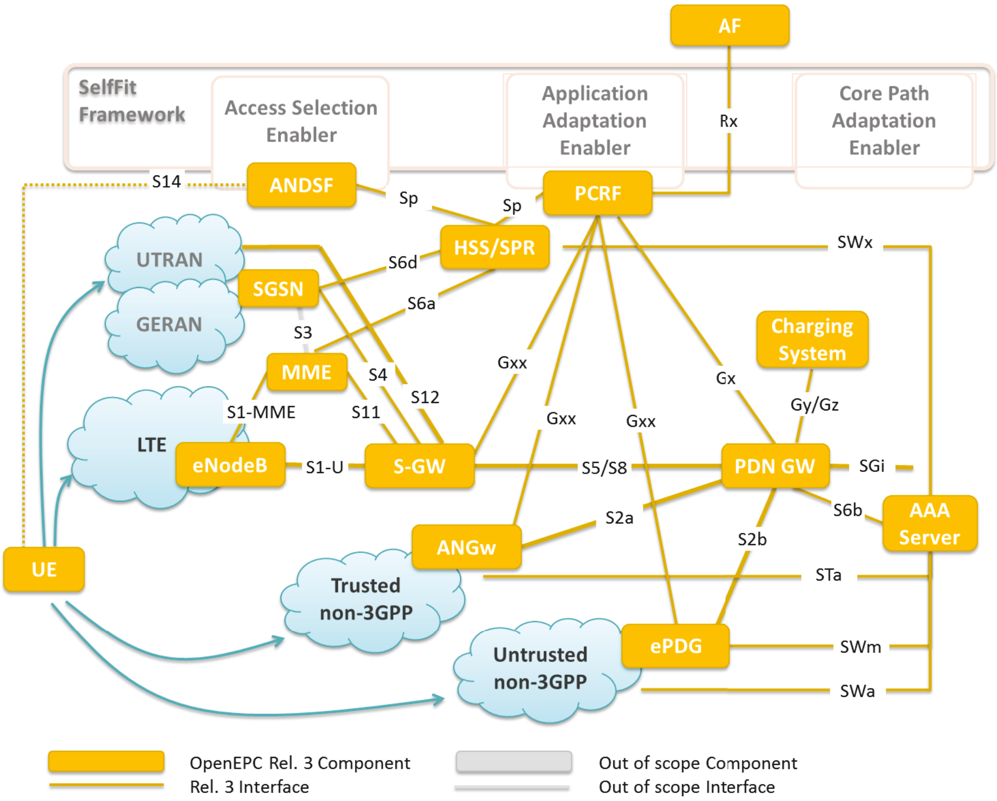


Figure 6.2: OpenEPC Architecture

of the users and of the service providers and to enforce resource reservation and charging rules on the data path. Additionally, the Offline Charging System (OFCS) gathers and correlates event and volume information from the network and from the service platforms, enabling the generation of consolidated subscriber billing. An Access Network Discovery and Selection Function (ANDSF) is transmitting indications to the UE on which access networks to select in order to be best connected from operator perspective. This information is processed by a minimal user space application on the UE enabling also data path steering.

OpenEPC includes a full solution for core network mobility support through both GPRS Tunneling Protocol (GTP) and Proxy Mobile IP (PMIP). The core network mobility support is transparent to mobile devices and application platforms, through dynamic IP address allocation and zero-packet loss handovers between heterogeneous access networks functionality in the Packet Data Network Gateway (PDN GW).

Additionally, OpenEPC contains the core network support for 3GPP and non-3GPP access networks enabling AAA, mobility and resource control procedures.

Specifically, it includes for the LTE access an eNodeB emulation, the Mobility Management Entity (MME) and Serving Gateway (S-GW) functionality and for the UMTS/HSPA access an equivalent NodeB/RNC emulation and a Serving GPRS Support Node (SGSN) functionality. For non-3GPP access networks, the AAA procedures are supported through an evolved Packet Data Gateway (ePDG) and through an AAA Server.

OpenEPC Rel. 3 interconnects directly only with non-3GPP access networks, such as WiFi. For 3GPP accesses, two options are considered. First solution is based on tunneling over a real operator network; including a non-reproducible effect on the demonstrations as the data traffic has to pass through a network not controlled by the testbed and a large data overhead. The second solution is based on emulation of the signalling procedures over WiFi and their translation in the standard core network procedures at the eNodeB and NodeB emulation, resulting in standard signalling, while with different radio network characteristics.

OpenEPC provides the implementation background of a close-to-real operator core network. Its solid basis and its large number of components and functions implemented provide the fundamental features on which novel functionality can be built and tested. The SelfFit Framework prototype implementation was build in parallel and as part of the different stages of the OpenEPC development. As OpenEPC implements a large amount of the 3GPP EPC, it can be considered a reference implementation for the core network standards. In the next sections no distinction is made between EPC and OpenEPC. Whenever the differentiation is required, due to missing features in the implementation or due to other choices taken in the realization of the toolkit, this will be underlined.

## 6.3 Access Selection Enabler

### 6.3.1 Integration in 3GPP EPC Architecture

Following the specification from Section 5.3, the ASE has to be interfaced with the UE, for the enforcing of the access selection decisions. Because of the tight coupling with the communication with the UE, the ASE was mapped to an ANDSF extension, the only EPC component which communicates with the UE at application level and has the goal to transmit access selection policies. The EPC mapping of ASE, implemented as reference prototype, is depicted in Figure 6.3.

The ASE reference implementation was realized based on the know-how, accumulated through the Converged Vertical Handover (ConVeHO) project of Fraunhofer FOKUS and Huawei Technologies. The additional parts are own concept and implementation of Fraunhofer and part of the OpenEPC toolkit.

The communication with the UE was extended in order to support the ASE

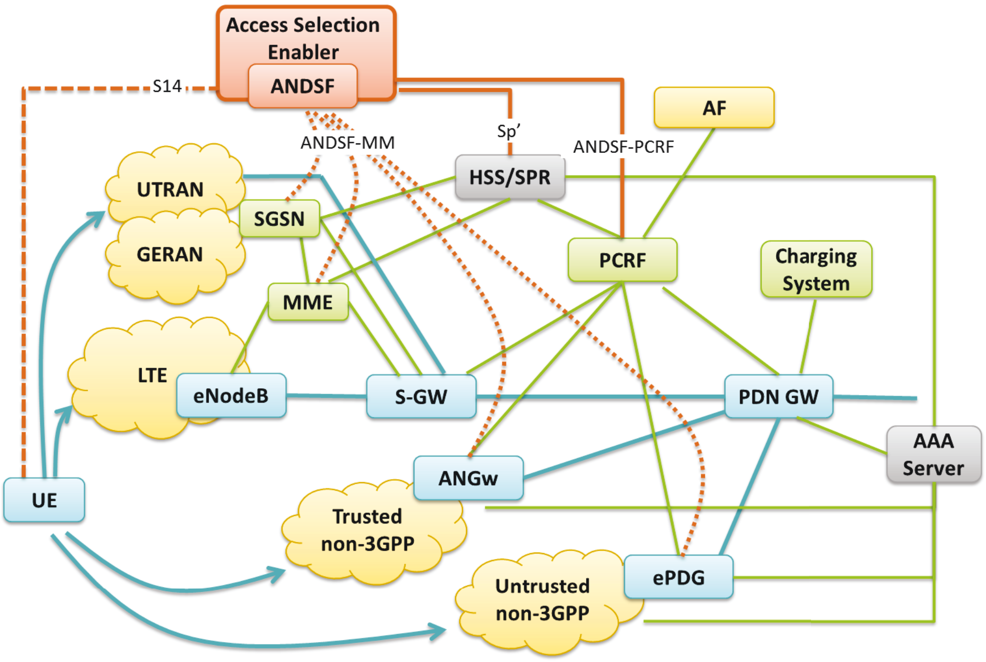


Figure 6.3: Access Selection Enabler Implementation

features. An OMA DM Management Object (MO) send by the UE to the ANDSF, including the current location information and the momentary user preferences. Additionally, in the MO send by the ANDSF to the UE [111], a new set of information was added. Through these additions, the network is able to indicate an immediate action for access network selection or additional information, enabling the selection of an access network based on the various applications established on the UE and on their specific data flow characteristics.

For the exchange of the MOs, no modification to the S14 interface is considered, thus respecting the OMA Device Management (DM) protocol as standardized by 3GPP [111] and realized in a simple manner in OpenEPC.

Apart from the communication with the UE, in order to retrieve additional information and to transmit adaptation indications a set of interfaces to other network entities are considered in the prototype following the specification from Section 5.3.

As ASE uses in its decisions information related to the subscription profile and to the dynamic subscriber information from the core network, which is stored in the 3GPP EPC HSS/SPR, a new interface is considered between the ASE and the HSS/SPR. The same information is requested by the PCRF, thus in order to minimally modify the existing HSS/SPR, the mapping interface is based on the Diameter based on PCRF Sp interface. Through this means, the HSS/SPR communicates with the ASE without any internal modifications.

PCRF aggregates the required information related to available and reserved resources for the subscribers. For retrieving this information, an interface is introduced between ASE and PCRF. The implementation is based on the PCRF Sp interface because PCRF is communicating only over Diameter and because the communication with the ASE does not follow the model of "decision-point enforcement-point" such as the Rx, Gx and Gxx interfaces. The reference prototype implements only the part related to the access network reselection due to resources change, as presented in the following data flows.

For communication with the core network mobility support functions, ASE is using a set of similar interfaces to SGSN for GPRS and UMTS technologies, to MME for LTE, and to ANGw and to ePDG for the non-3GPP accesses. Through this interface, access network related events may be transmitted. However, these events are also available in the PCRF, thus an extension of the ASE-PCRF interface suffices. The interface enables the transmission of handover indications to the access networks in order to prepare for possible handovers. In the implementation a simplified proprietary interface is used for rapid prototyping reasons enabling only the handover indications support.

For notifying ASE on network topology modifications, an additional interface towards the network management system is used. This interface is not longer detailed because it will be deployed in a proprietary form of the vendors. In the SelfFit prototype implementation, the interface is realized in the form of a GUI, enabling the operator of the testbed to control the execution of the SelfFit procedures.

The prototype implements the simple access network state machine described in Section 5.3. For simplification of the implementation it was assumed a static map to the access networks in the area.

### 6.3.2 Procedures and Data Flows

A set of exemplary procedures are described, as implemented on the reference prototype. The most important other procedures which include a large degree of novelty compared to the standards are available in Appendix 8.2.2.3. It is assumed that the ANDSF standard procedures are implemented and available in the system.

#### 6.3.2.1 Initial Attachment Procedure with UE notification

This procedure describes the initial attachment for a UE using the ASE service. For the UE to connect to EPC network and to establish the first connection to ASE over S14, specific network operations have to be executed. The procedure is depicted in Figure 6.4. The following steps are executed:

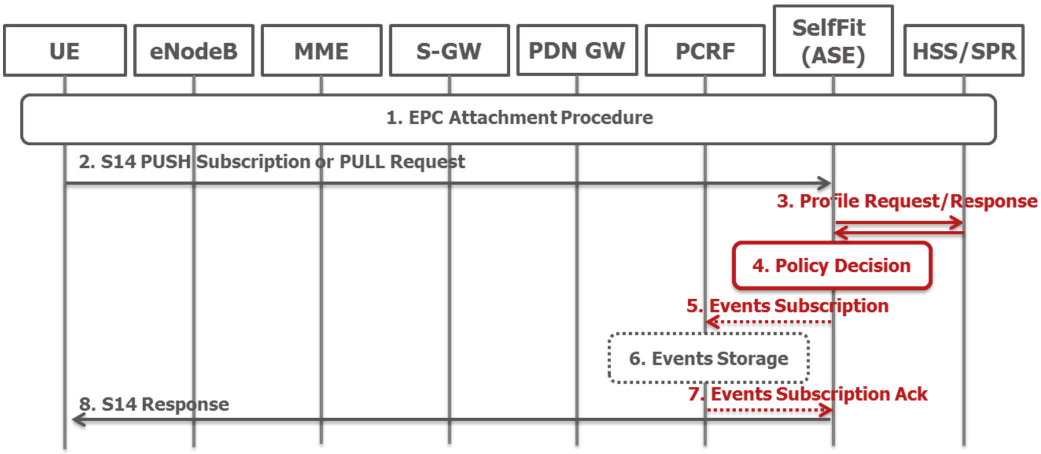


Figure 6.4: Access Selection Enabler - Device Attachment Data Flow

1. The UE attaches to the EPC, as specified in the current 3GPP standards, receiving mobility support, an IP address and a default level of resources. Through this means it can communicate over the network.
2. The UE sends to the ASE a PUSH subscription or a first PULL message. The message is used for authentication of the UE to the ASE ANDSF. The ANDSF component of the ASE is discovered based on pre-provisioned policies or based on DHCP or DNS information, similarly to the standard discovery of other service level components.
3. The ASE requires from the HSS/SPR the profile of the subscriber. A subscription events related to user profile changes can be executed, resulting later in notifications which may trigger other policy decisions. The HSS/SPR sends to the AAE the subscription profile.
4. The ASE makes a policy decision whether the UE is allowed to use the access selection service. Additionally, it decides on the policies to be transmitted to the UE and whether to subscribe to the PCRF for data path events.
5. When ASE subscribes to events it sends to the PCRF an events subscription request for events, such as loss of connectivity or access network change, the resources required and reserved change etc. For this, the ASE has to discover the PCRF allocated to the UE, operation which is currently available in EPC through the Diameter based Service Locator Function (SLF) or Dimater Router Agent (DRA).
6. The PCRF registers the events subscription as part of the application subscriptions for the events similarly to the operations over Rx interface.

7. The PCRF notifies the ASE on the subscription registration completion.
8. A PULL response or a first notification for PUSH is transmitted from the ANDSF component of the ASE to the UE.

For faster prototyping, in the OpenEPC implementation, the Steps 5-7 are implicitly realized. PCRF considers that all the devices have registered for the access network change events as it would be the case when the solution is completely integrated in the network.

The procedure is asynchronous to the attachments of the UE. A handover, executed during the time when the procedure is executed, affects negatively the communication with the PCRF. As the same PCRF and the same ASE are used during all the operation of the UE, the subscription to the PCRF may be followed by a notification with the momentary parameters of the UE (e.g. APN, IP addresses, IP-CAN Type, PDN-Gw address, etc.), as part of subsequent attachment procedures.

Additionally to the Attachment Procedure with UE notification, an alternative Attachment Procedure with PCRF notification was considered, in which the ASE receives the attachment event directly from the PCRF, before the data path is established. This alternative is included in Appendix 8.2.2.3. Although with less parallel processing and thus, with larger delay, the UE notification based procedure is considered more suitable and it was implemented in the SelfFit prototype, due to the registration only of devices capable and willing to use the ASE service.

Subsequent attachments may be executed due to access network selection procedures resulting in horizontal or vertical handovers. For these attachments, the retrieval of the subscription profile is not anymore required. Additionally, there is no need for subscribing to new data path events. Similarly with the initial attachment, two procedures may be considered: with UE and with PCRF notification. From these, the UE notification is preferred in order to maintain consistency of the realization over multiple procedures.

For the complete detachment from the EPC, the same two procedures based on PCRF and UE notifications are considered. However, the detachment may be sudden, therefore in order to be able to maintain in the ASE a consistent state the PCRF notifications are required. The specific procedure is described in Appendix 8.2.2.3 and not further implemented in the reference SelfFit prototype.

#### 6.3.2.2 Resource Oriented Access Network Selection Procedure

This procedure describes the access network selection triggered by a change in the resources required by a subscriber from the network, based on the specification from Chapter 5.3.3.2 applied on the 3GPP EPC architecture. For exemplification, a procedure was selected in which the UE is attached to a UMTS network which is not able to support the required resources. The procedure is depicted in Figure 6.5. The procedure was implemented as proof of concept in the SelfFit prototype enabling handovers between the UMTS and the WiFi access networks. The following steps are executed:

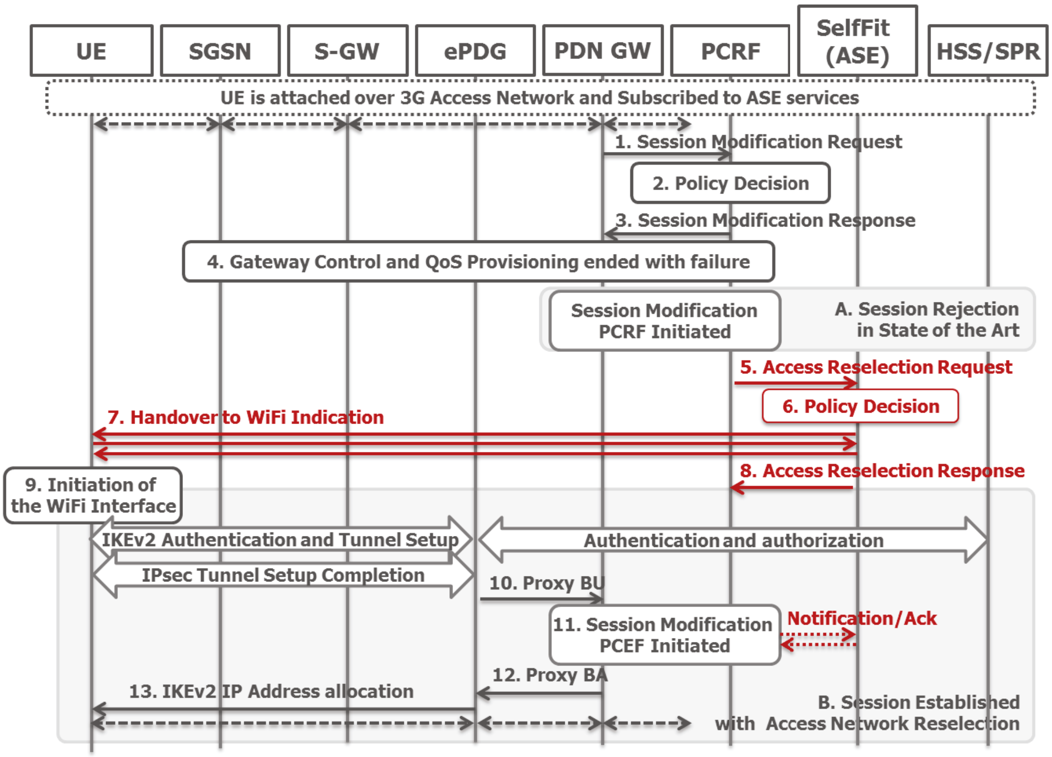


Figure 6.5: Access Selection Enabler - Resource Oriented Data Flow

1. The UE is connected and able to exchange data over the UMTS network. The device is also subscribed to ASE services. It requires additional resources for starting of a new data session. The request is received by PCRF from either the UE through the PDN GW or from the applications. The procedure here presented is initiated by the UE resulting in a "Session Modification Request" message send by the PDN GW to the PCRF.
2. The PCRF makes a decision on the new policies which have to be installed.
3. The session modification is acknowledged to the PDN GW. As considered in the 3GPP EPC, the response is transmitted before the actual reservation of resources takes place.
4. In order to reserve resources over the UMTS network, a Gateway Control and QoS Provisioning Procedure is executed, which is assumed to fail due to

lack of resources

In the state of the art case, this step is followed by an IP-CAN session modification, PCRF Initiated procedure and the session is released due to lack of resources.

5. As described in Section 5.3, the PCRF sends to the ASE a new access reselection message, containing information on the QoS parameters required by the new session that is to be established. Optionally, it should contain information on the prior established sessions of the UE.

6. The ASE makes a decision, according to the information received from the PCRF. Other input parameters should be also considered, like the location of the UE, the load of the access networks located in the area of the UE, etc. In this example, the decision consists on the selection of an access network to which the UE to handover to. It is supposed that at least one WiFi access is preferred by the operator to sustain the continuity of the active sessions of the UE.

7. The decision is transmitted in the form of inter-system mobility policies to the UE according to the S14 standard using the OMA DM PUSH mechanism. The reselection indication presumes an immediate execution from the UE.

8. Optionally, a reselection indication can be transmitted to the PCRF as not to release the session and to wait until the UE connects to the WiFi access network.

9. Based on the indication received from the ASE, the UE initiates the reselection procedures to the WiFi access network. First, it initiates the device interface, then the attachment procedures are executed including the authentication and authorization and the IPsec tunnel establishment.

10. The ePDG requests the establishment of the data path through the PDN GW. In this example a PMIPv6 "Binding Update (BU)" is used.

11. The PDN GW initiates a session modification procedure in which it requires from the PCRF the resources to be reserved over WiFi for the specific UE. The PCRF is informed on the completion of the WiFi attachment. The PCRF may send a similar notification to the ASE, enabling the ASE to know where the device is attached to. As the ASE made the reselection procedure, based on the resources needs of the UE and on the available resources over the access networks, this procedure will succeed.

12. The PDN GW sends to the ePDG a PMIPv6 ”Binding Acknowledgment (BA)”, including the UE IP address, acknowledging the PDN GW attachment completion.
13. The attachment procedure is completed through the transmission of the allocated IP address to the UE. Now the device can communicate directly over the WiFi.

In the reference implementation, this procedure is realized based on session resource modification received from the AF, due to the initiation of a new service. This alternative, in which resources are reserved from the network side, does not change the goal, the handling, or the duration of the overall procedure.

The here described procedure enables the continuation of the UE's service over the WiFi access network, even though the current UMTS network can not sustain them. From the perspective of the services established, the complete procedure remains transparent, however with higher delay, which is in the range of accepted delays for services establishment over the wireless network, as analyzed in the next chapter.

In an optional version of the procedure, the PCRF queries the ASE prior to the resource reservation. However, this introduces a higher delay to all the reservation procedures for UEs which do not require an access network reselection and increases the overall system signaling.

A same procedure may be executed on service modification or on service termination. When a service is changing its resources required or when a service is terminated, it may happen that another access network sustains at a more efficient cost the remaining active services of the UE.

#### 6.3.2.3 Volatile-Shallow Context Procedures

In this section, the implementation in the 3GPP EPC of the volatile shallow context procedures is presented, based on the specification from Section 5.3.3.3. For exemplification, the volatile-shallow context establishment and activation are included here. The volatile-shallow context termination is included in Appendix 8.2.2.3 and not implemented in the SelfFit prototype.

For the volatile-shallow context establishment procedure, depicted in Figure 6.6, it is assumed that the UE is attached to an access network other than LTE and that a set of new access selection indications are transmitted from the ASE to the UE as inter-system mobility policies. Due to the network topology and to the device capabilities, the handover to LTE is considered highly probable, thus a volatile-shallow context is established. The following steps are executed:

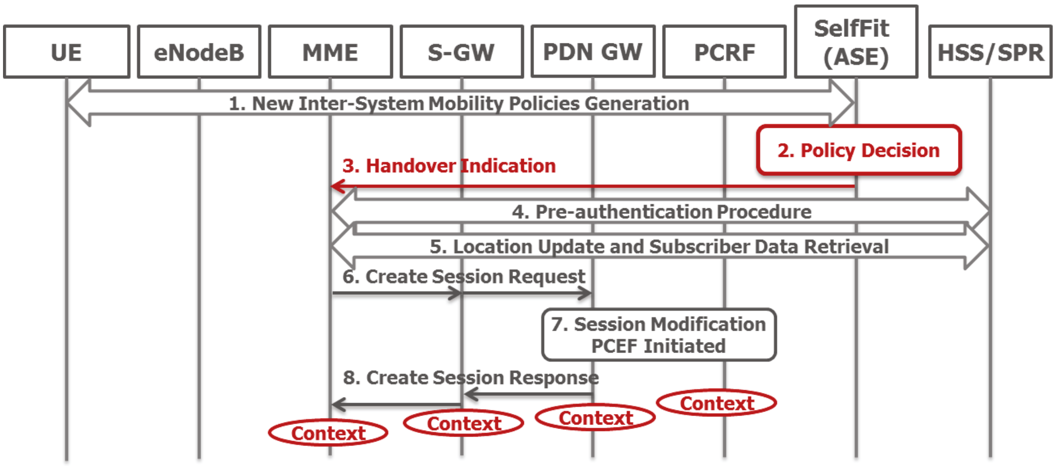


Figure 6.6: Access Selection Enabler - Volatile Shallow Context Establishment

1. New inter-system mobility policies are send by the ASE to the UE through the S14 interface, containing the access networks which the UE can select.
2. The ASE makes a decision for which of the transmitted access networks, a shallow context may be beneficial, based on the access network to which the UE is already connected to, the target access network type, and the probability of a handover. For LTE, ASE will select the MME which would serve the UE in case of a handover.
3. The ASE transmits to the MME a handover indication. It is considered that, by knowing the access network to which the UE may handover to, the ASE is aware also of the MME which will serve the UE in that access network. An MME is expected to cover a large network area encompassing multiple eNodeBs, thus increasing the probability, when a handover to LTE occurs, to be controlled through the ASE selected MME.
4. The MME contacts the HSS and authenticates the UE. The authentication procedure is executed only on the network side. It includes the retrieval of the authentication vectors from the HSS, based on the identifier received from the ASE in the previous step, as presented in 3GPP TS 33.401 Section 6.2.1 [117].
5. The MME retrieves the subscriber data from the HSS, including the PDN GW identity and the information on the PDNs the UE is connected to over the source access network.
6. The MME selects a Serving GW for the case no APN was provided by the UE as described in 3GPP TS 23.401. The MME sends to the selected S-GW a

GTP "Create Session Request", including the information received from the ASE. The GTP "Create Session Request" is based on the standard message adding a "Shallow" flag indicating that the tunnel has only to be created and that the data traffic should not be routed through it.

The message is forwarded by the S-GW to the PDN GW, including the same "Shallow" flag having the same indicative role.

7. The PDN GW is aware due to the "Shallow" flag that a shallow context is created. It executes the mobility and resource reservation procedures as follows:

    - When the GTP "Create Session Request" is received, a new tunnel is created, without routing data traffic to it.
    - An indication to the PCRF is transmitted, which contains the "Shallow" flag using standard modification of the reserved resources procedure due to a handover.
    - The PCRF correlates the request with the identity of the UE and maintains in the subscriber structure the information on the established shallow context.
    - The PCRF makes the authorization and the policy decision, considering that a handover may occur in the near future. A response is sent to the PDN GW, indicating which resources have to be reserved and which event triggers are to be activated in case the connectivity over the target access network is established.
    - The resource reservation rules and the event triggers received from the PCRF are maintained by the PDN GW until the activation of the shallow context. No operations are executed on the data path.

8. The PDN GW responds with a GTP "Create Session Response" including the QoS rules that have to be installed on the S-GW and the IP addresses or prefixes to be assigned to the UE in case of a handover. The GTP "Create Session Response" is forwarded by the S-GW to the MME for completing the shallow context.

Through the establishment of the shallow context, the MME, S-GW, PDN GW and PCRF have all the information necessary for a handover to the target access network. When the UE attaches to the LTE access network, the volatile-shallow context activation procedure is executed with the following steps, as illustrated in Figure 6.7:

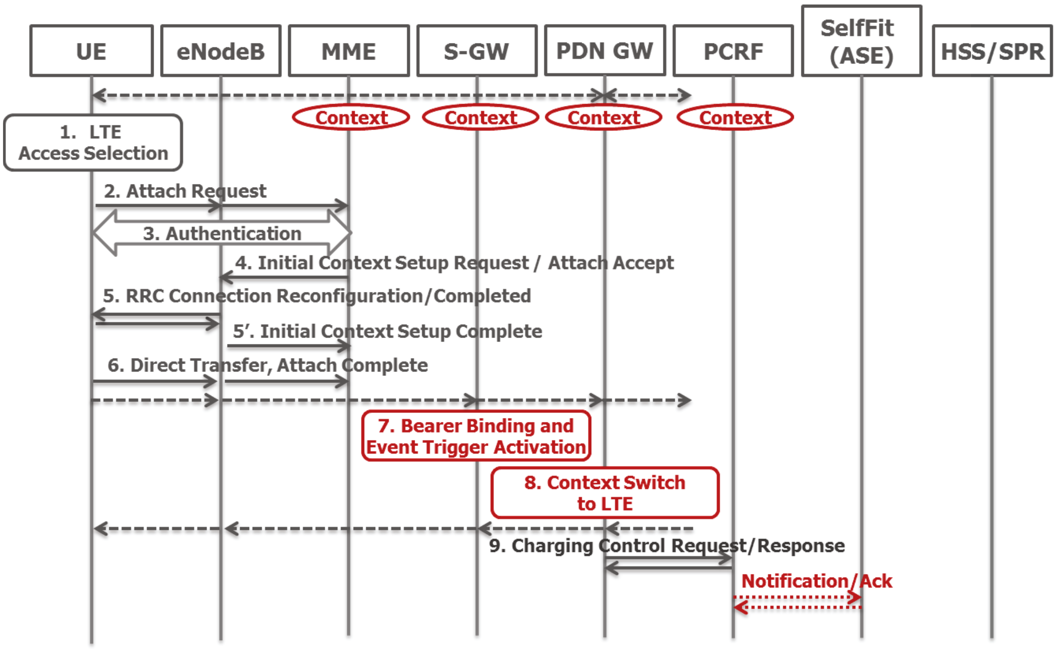


Figure 6.7: Access Selection Enabler - Volatile-Shallow Context Activation

1. The activation of the volatile-shallow context is triggered by the attachment of the UE to the LTE access network. The UE selects an eNodeB, based on radio specific characteristics, such as signal strength levels and on the handover policies received from the ASE.
2. The UE sends to the eNodeB an attachment request which is then forwarded to the MME serving the specific area. The MME maintaining the shallow context is selected due to the area based selection criteria.
3. The authentication procedure is executed. As the MME is aware of the identity of the UE, the procedure is executed only between the UE and the MME and may be entirely skipped, if the confidence level on the ANDSF shallow context establishment is considered enough.
4. The wireless link is configured. The UE may receive its IP address at this step or through the IPv6 Neighbor Discovery mechanism.
5. The MME accepts the attachment and initiates the wireless link context setup. The LTE radio connectivity is established through Radio Resource Control (RRC) procedures and the radio attachment is completed.
6. A notification of the success of the procedure is transmitted to the MME. The MME considers the previous shallow context as active. After this step,

the UE is able to transmit uplink messages over the LTE access network, while the downlink is maintained over the source access network.

7. When the first uplink data packet is received by the S-GW, then it activates the shallow context. The activation includes the enforcement of QoS rules and the activation of the bearer, based on the information stored in the shallow context. No procedure for the data path establishment or for the retrieval on the level of resources to be reserved has to be executed, as in the state of the art case.
8. The first uplink data packet triggers the activation of the context also in the PDN GW. This presumes the activation of the bearer and the installation of the policy and charging control rules as part of the shallow context. Additionally, the downlink data traffic is redirected towards the target S-GW. From this moment on, the data traffic is exchanged bi-direction ally over the target access network.
9. A Resource Modification Procedure is executed between the PDN GW and the PCRF, informing the PCRF that the connectivity was established over LTE. In its turn, the PCRF may inform the ASE on the completion of the activation of the shallow context.

Through the LTE activation procedure, a large number of the steps which have to be included in attachment procedures in a synchronized manner are simplified, through the removal of the remote communication related to authentication and authorization, policy and charging control, and mobility support.

A similar set of establishment and activation procedures can be executed for other access networks, such as UMTS, GPRS, WiMAX or WiFi, considering the specific attachment procedure and the gathering of the required information in the gateways and control entities, without realization on the data path.

This solution is especially beneficial, when a handover has to be executed from an access network to which the signal is fast lost to an access network with attachment procedures with long duration, such as the handovers from WiFi hotspots to a macro-3GPP operator network, like LTE.

As the LTE radio technology integration was not available at the moment of the realization of this feature, a hybrid model was realized in which only the mobility control information is integrated, leaving the pre-authentication and resource reservations to be implemented directly in pre-product versions.

### 6.3.3 Implementation Modules

For implementing the ASE functionality, the practical realization used the following OpenEPC modules and their dependencies, as depicted in Figure 6.8.

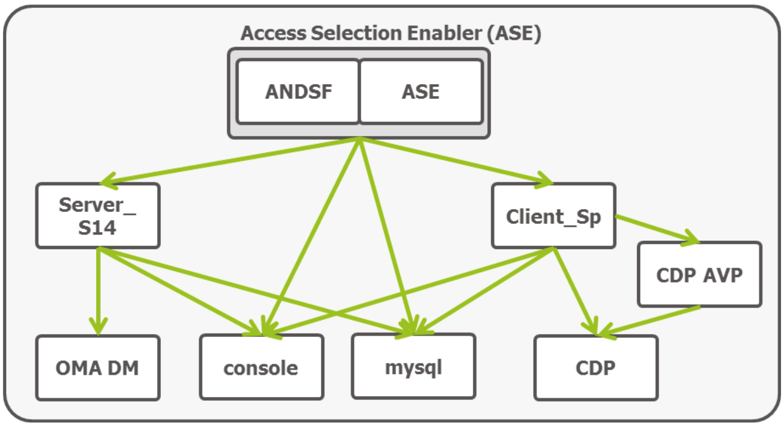


Figure 6.8: OpenEPC Modules for Access Selection Enabler

The main control module for ASE is co-located with the ANDSF control module. This allowed for an efficient prototype, by introducing the afferent ASE states in the ANDSF state machine and thus not requiring a new interface between two modules. The ASE state is completely synchronized in a mysql database. For the communication with the UE, the S14 interface module handles the communication state machine while the OMA DM module encodes and decodes and handles the transmission state machines. For the communication with PCRF and with access network gateways, an OpenEPC console remote procedure call was used. Additionally, the interface to the HSS was implemented using the C Diameter Peer (CDP) stack of OpenEPC including its specific Diameter Attribute Value Pair codec tokens using CDP_AVP module. The basic diameter modules are used by the Client_Sp module implementing the specific interface state machine.

## 6.4 Core Path Adaptation Enabler

Based on the specification from Section 5.4, the CPAE is interfaced with the main control core network entities and with the control functionality in the data path entities, including the anchor nodes and the Forwarding Entities (FE) located at the border of the operator domain. The mobility support of the 3GPP EPC controls only the adaptation of the access and local gateways, while maintaining the same PDN GW for a connectivity session of the UE having the role of the anchor node, while CPAE has a distinct data path adaptation role in orchestrating mobility anchor reselection. For this reasons, CPAE represents a new function which is introduced in the operator network. The CPAE interfaces, as realized in the reference implementation, are depicted in Figure 6.9.

The CPAE reference implementation was based on the know-how from the

Flat Mobility (FlatMore) project of Fraunhofer FOKUS and Huawei Technologies. Several parts are own implementation of Fraunhofer and part of future OpenEPC releases.

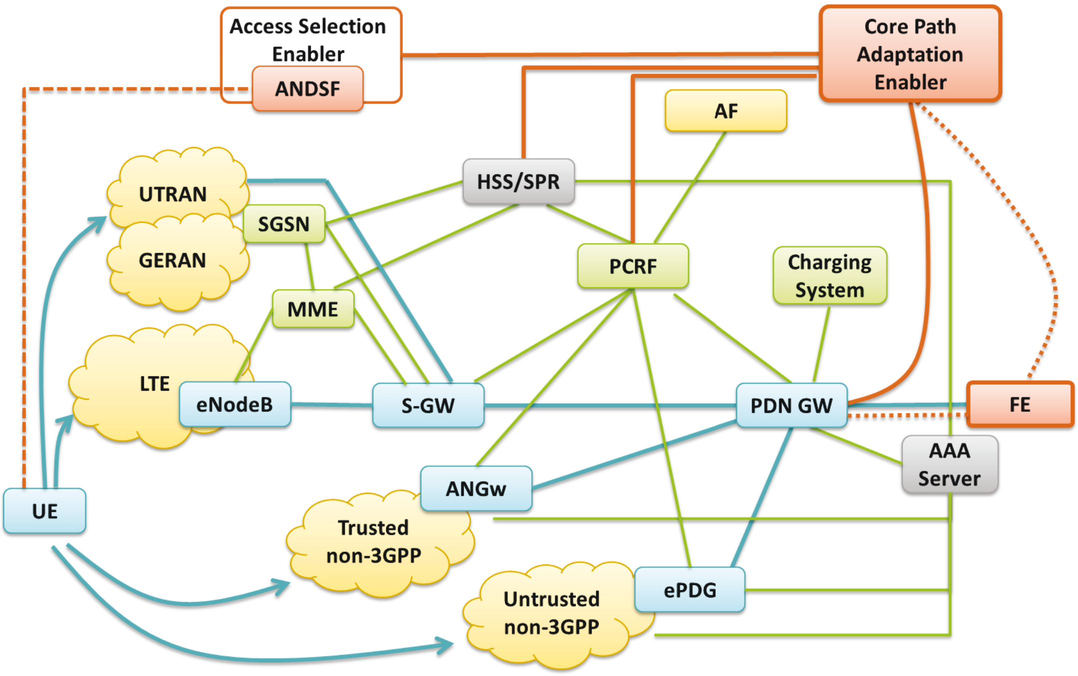


Figure 6.9: Core Network Path Adaptation Enabler

Similarly to ASE, for receiving subscriber information, CPAE is connected to the HSS/SPR database. In order not to modify the HSS, the same extension of the PCRF Sp interface is considered as in the case of ASE.

The operations of CPAE are mainly triggered by changes in network configuration or by subscriber mobility events. For the network configuration changes, an interface is considered between the CPAE and the operator management system, enabling it to receive notifications on solicited and unsolicited network topology changes for maintenance and for fail-over mitigation. This interface is specific to each operator deployment and thus, it was realized in the implementation in the form of an administrative front-end. The subscriber mobility events are best received from the mobility control functionality. As to manage the information in a convergent manner and to reduce the number of interfaces and as all the required information is available already in the PDN GW, a single interface is introduced between CPAE and PDN GW.

From PDN GW, CPAE receives information on the initial and subsequent connections establishment and termination events, through this, being able to determine to which PDN GW UEs are momentarily connected to. Additionally, it enables CPAE to orchestrate the reselection procedure, as described in Section 5.4. Other information, such as the network location of the UE and its connectivity state, can be also transferred on demand or in the form of notifications, when specific events happen. As the notification mechanism enables the customization of the messages over the communication link and thus, only the required events are transmitted to CPAE, this mechanism was preferred during the implementation.

The mobility control information combined with the required resources and the subscription to specific dedicated service which can be acquired from PCRF, enabling CPAE to make accurate decisions on optimizing the core network data path. However, it is assumed that the PDN GWs are capable of accommodating a large number of UEs and even though heterogeneous in capacity, their actual resource consumption ponder normalizes itself due the large scale of communication, this information was not further considered in the realization of CPAE.

Based on this information, PDN GW reselection decisions are made. For CPAE to orchestrate the reselection procedure execution, the interface to the PDN GWs is employed. It enables the source PDN GW to be aware of a reselection and of the identity of the target PDN GW. Additionally, it enables the target PDN GW to determine when the reselection is terminated, as to further continue any additional procedures related to other core network features, such as accounting or charging.

As the PDN GW is changed during UE's connection to the EPC and as for the service continuity, the same IP address has to be maintained, the IP address loses its role of network locator inside the operator domain and keeps only the one of locator between operator domains, while being the identifier of the subscriber in the overall Internet.

For propagating the operator network location of an IP address and thus, to be able to appropriately forward the downlink data traffic, a new loose interface is considered between CPAE and the operator Forwarding Entities (FEs), receiving all the data traffic entering the operator domain. Through this interface, some specific FEs, which require to forward the data traffic of the UE, are notified on the current destination of the data packets. For maintaining this information, the FEs maintain a binding table is updated dynamically for each subscriber, as described in the following subsection.

CPAE maintains a table with a specific entry for each of the subscribers including the IMSI identity of the UE, the IP address allocated, the identity of the current PDN GW and the identities of the FEs.

### 6.4.1 Procedures and Data Flows

In this subsection, a set of example procedures realized in the prototype are described. Other procedures, which enable the functioning of the overall EPC system, implementing CPAE functionality, are described in Appendix 8.2.2.3. The procedures described address the data path adaptation with and without changing the access network to which the UE is attach to.

The procedures for initial attachment and for detachment from the wireless system and the procedures for establishing a downlink data path using synchronous and reactive notifications are included in Appendix 8.2.2.3.

#### 6.4.1.1 Data Path Adaptation with Access Network Change

This procedure is executed when the UE attaches to a target access network for which a more appropriate PDN GW should be selected. The trigger is the handover to a new cell of the same or other access technology. The procedure may be triggered for offloading reasons by CPAE, requiring an access network reselection execution to ASE, in case the target PDN GW selected can not be reached from the current access network of the UE.

In this procedure, two access networks are involved. In order to maintain a simple architecture, the procedure implemented considers that the two access networks do not have a direct redirection data path through the EPC. The procedure is depicted in Figure 6.10.

The procedure includes the mechanisms for flexible reselection of the data path entities, the synchronous data path adaptation and alternatively the reactive data path, as described in Section 5.4, considering the integration with the existing 3GPP features and procedures for authentication, authorization, IP address allocation, mobility, and resource support, as included in the SelfFit reference implementation.

The procedure begins with the UE attached to a source access network and communicating through a source PDN GW. The downlink data traffic is received through one or multiple FEs, which are able to redirect the data traffic to the source PDN GW. The uplink data traffic follows the operator forwarding rules.

1. The procedure may be initiated by a core network data path decision, which can not be realized unless the UE changes the current access network. This decision is taken by CPAE, based on an internal operator trigger, relating to changing of operator connectivity management polices or related to a network topology change.

2. An access network selection request is transmitted by the CPAE to the ASE, requiring that the UE executes a handover to a target access network through

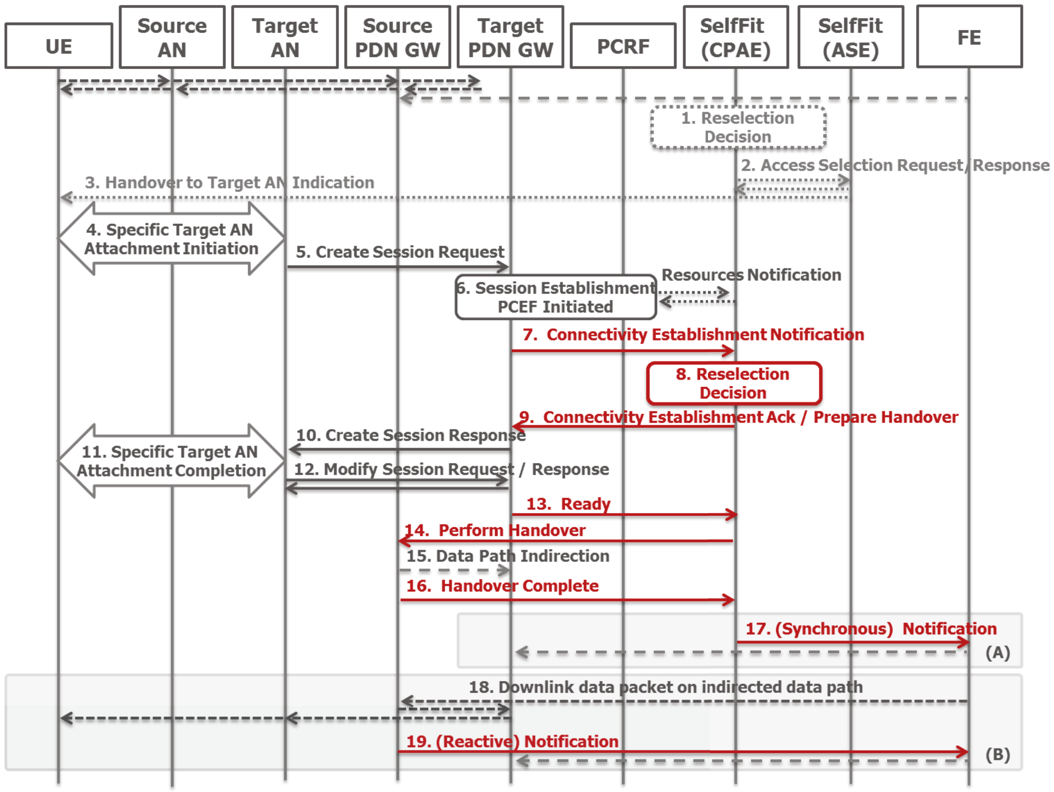


Figure 6.10: Core Path Adaptation Enabler - Data Path Adaptation with Access Network Change

which the core network path adaptation becomes possible.

3. The ASE makes a policy decision and transmits access network selection indications to the UE.

4. Based on the indications from the ASE or triggered by the radio conditions change, due to the UE's mobility through the physical environment, an attachment to the target access network is initiated using specific procedures.

5. In the convergent EPC environment, the attachment procedure will translate independently of the access network into a GTP ”Create Session Request”, which is transmitted from the Target AN to a Target PDN GW.

6. As the Target PDN GW is not aware of any previous connectivity of the UE, it will initiate a ”Session Establishment” procedure with the PCRF for reserving default communication resources. A notification may be send to CPAE, containing the resources allocated to the UE.

7. When the resources are reserved in the target PDN GW, it transmits a "Connectivity Establishment Notification" to CPAE, requesting also the IP address allocated to the UE, if any.

8. CPAE identifies the UE and due to the previous attachment to the Source PDN GW has an IP already allocated. It makes a reselection decision, whether the device is communicating through the appropriate Target PDN GW. In case another PDN GW has to be selected, then the procedure for data path adaptation without access network change is initiated, as described in the next subsection.

9. A "Connectivity Establishment Ack" is transmitted from CPAE to the Target PDN GW, including the IP address allocated to the UE and additional information related to preparing a handover, including the address of the Source PDN GW.

10. The Target PDN GW completes the attachment procedure and transmits a GTP "Create Session Request", including the IP address allocated to the UE and information on the PDN GW end of the mobility tunnel.

11. The procedure of attachment over the target access network is completed. During the operations, the IP address is transmitted to the UE.

12. In order to confirm the termination of the attachment and to modify, if needed, the resources allocated to the UE, a "Modify Session Request/ Response" message exchange is performed between the target access network and the Target PDN GW, as considered by the 3GPP standards.

13. Apart from modifying the session parameters, as considered by the standard attachment procedure, the Target PDN GW also notifies the CPAE that it is ready for a handover from the source access network. At this moment, the UE is able to transmit messages over the Target or the Source PDN GWs, as the two distinct data paths are established, while it may receive data packets over the Source PDN GW, as no redirection to the Target PDN GW was executed.

14. CPAE indicates to the Source PDN GW to perform a handover to the Target PDN GW. The message includes the UE identity and the allocated IP address.

15. Based on the handover indication, the Source PDN GW directs the downlink data traffic to the Target PDN GW, through the introduction of a new forwarding rule for the specific UE.

16. The Source PDN GW announces the CPAE that the handover information was received and that the redirection to the Target PDN GW was realized. The UE attachment context from the Source PDN GW is removed at a later phase when the UE detaches from the Source PDN GW. The information on the redirection towards the Target PDN GW may be maintained longer, depending on the redirection algorithm used and re-created if necessary through the communication with CPAE. The EPC part f the data path adaptation is completed.

17. In case the Synchronous Data Path Adaptation mechanism, described in Section 5.4, is deployed, then, a synchronous notification is transmitted to all the FEs which have to transmit downlink data traffic for the UE, known, as the FEs registered to CPAE when the first downlink data packet had to be forwarded. Through these means of multicasting the information to all and only to the FEs that have downlink data traffic, the data path adaptation is propagated to the next forwarding level in one synchronous step. From this moment, all the FEs will transmit the downlink data to the Target PDN GW, therefore enabling the release of the context over the Source PDN GW.

18. Alternatively, in case of the deployment of the Reactive Data Path Adaptation mechanism, described in Section 5.4, no FE will be notified at the end of the reselection procedure. Instead, when an FE has downlink data packets, it will send them to the Source PDN GW, as being the last one on which downlink forwarding information was available. The Source PDN GW has to forward the received data packets, using the redirection to the PDN GW.

19. Trigger by downlink data packets, the Source PDN GW sends to the FE an indication to direct the data traffic to the Target PDN GW. The other FEs are not updated unless executing Step 18.

The reference implementation was realized considering the synchronous notification of the FEs, due to the reduced deployment considered, with two PDN GWs and two FEs. With a high increase in number of PDN GWs and FEs, the reactive solution is more efficient, as evaluated in the next chapter.

From 3GPP EPC perspective, the procedure here described enables the flexible selection of the last fixed entity on the data path, leading the way to highly distributed core networks in which a major part of the operator core network functionality including the PDN GW is localized close to the served subscribers.

The same procedure can be used for multi-homed devices. In this case, the IMSI and the IP address based identification does not suffice. Instead a set of data flow templates has to be also considered. A template represents a class of active data flows for which the data path can be adapted through a single execution of the procedure. As the information required is higher compared to the case of single interface devices and as potentially the procedure will be executed multiple times i.e. for each data flow template, it requires more signaling support.

#### 6.4.1.2 Data Path Adaptation without Access Network Change

This procedure is executed when an administrative trigger is received, such as for maintenance, energy efficiency or for the modification of the core network topology. Although produced by other systems in the network, in the implementation the trigger is considered to be directly generated in CPAE. As the access network does not change and as the operations remain completely transparent to the UE, this procedure anchors the data path in the access network for the duration of the data path adaptation. The procedure, as realized in the reference implementation, is depicted in Figure 6.11.

Similarly with data path adaptation with access network change, this procedure includes the mechanisms for flexible reselection of the data path entities, the synchronous data path adaptation, and alternatively the reactive data path, as described in Section 5.4 and their integration with standard EPC functionality for the specific use case in which the data path adaptation is not visible to the UEs.

At the beginning of the procedure the UE is attached to an access network and is able to communicate through a Source PDN GW. The downlink data traffic is received through one or multiple FEs which are able to redirect it to the source PDN GW. The uplink data traffic is following the IP routing rules inside the operator network.

1. The procedure is initiated by a reselection decision which is taken by CPAE based on internal operator trigger relating to changing of the operator network topology or subscriber and traffic engineering rules. The decision taken is that without changing the access network to which the UE is connected to, UE's connectivity has to be established over the Target PDN GW.

2. A handover preparation indication is transmitted to the Target PDN GW, including the UE's IP address and IMSI and the information on the Source GW.

3. Based on the information received from CPAE, the Target PDN GW prepares to receive an announcement of connectivity from the access network. As the Target PDN GW does not have to announce any access network, it immediately transmits an indication that it is ready for a handover.

4. CPAE indicates to the Source PDN GW that a handover without access network change has to be executed. The message includes the UE's identity and IP address.

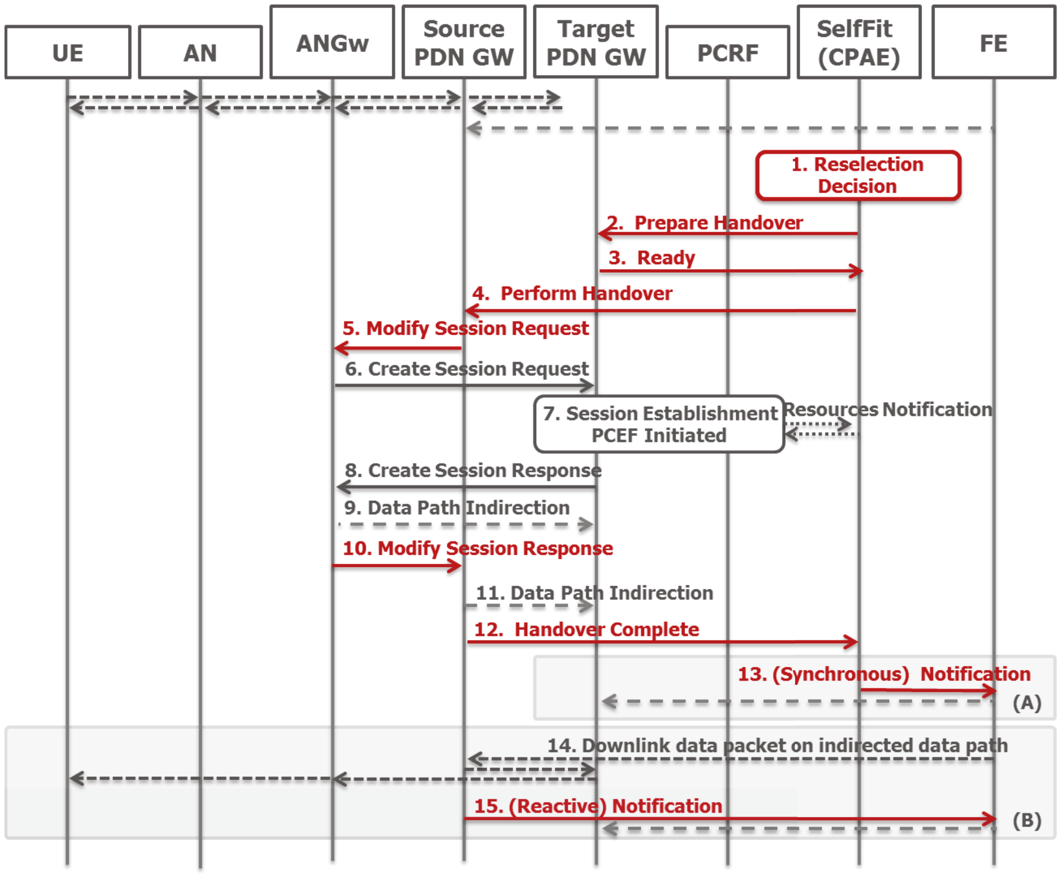


Figure 6.11: Core Path Adaptation Enabler - Data Path Adaptation without Access Network Change

5. The Source PDN GW transmits a GTP "Modify Session Request" to the gateway of the access network to which the UE is attached to. The message contains the information to re-establish the session over the Target PDN GW and to terminate the session over the Source PDN GW.
6. The AN Gw transmits a GTP "Create Session Request" to the Target PDN GW for establishing the new data path.
7. A Session Establishment procedure is initiated between the Target PDN GW and the PCRF, which hands over the resources from the Source PDN GW to the Target PDN GW. A resource notification may be sent to CPAE.
8. A GTP "Create Session Response" is sent from the Target PDN GW to the AN Gw. At this moment the data path through the Target PDN GW is established.

9. The uplink data traffic is forwarded through the Target PDN GW. The downlink data traffic is still received over the Source PDN GW. This connection has to be terminated when the first downlink data packets are received over the Target PDN GW.

10. The AN Gw announces the Source PDN GW that the uplink data traffic redirection through the target PDN GW was completed.

11. The Source PDN GW makes a data path redirection for downlink data traffic through the Target PDN GW. At this moment, all the data traffic passes through the Target PDN GW, however redirected on the downlink.

12. An indication that the handover operations were completed is send from the Source PDN GW to CPAE. The next steps are executed, as described in the procedure for data path adaptation with access network change.

13. In case the Synchronous Data Path Adaptation mechanism described in Section 5.4 is deployed, then a synchronous notification is transmitted to all the FEs which have to transmit downlink data traffic for the UE, known as they have subscribed to changes upon forwarding the first downlink packet.

14. Alternatively, in case of the deployment of the Reactive Data Path Adaptation mechanism, described in Section 5.4, no FE will be notified at the end of the reselection procedure. Instead, when an FE has downlink data packets, it will send them to the Source PDN GW. The Source PDN GW has to forward the received data packets using the redirection to the PDN GW.

15. The Source PDN GW sends to the FE an indication to direct the data traffic to the Target PDN GW. The other FEs are not updated unless executing the Step 14.

Similarly to the procedure for adaptation with access network change, the reference realization is considering only the synchronous notification of the FEs.

This procedure may be the single one which is implemented in a system. The data path with access network selection can be implemented as two consecutive procedures. First, the handover is executed to the target access network using the standard procedure and then the core path reselection without access network change. Even though the duration of the procedure is larger, the delay does not affect the subscriber data exchange and the EPC system remains more robust.

From the EPC perspective, this procedure brings the flexibility of the data path without requiring any communication with the UE, similarly to the changes of the data path in the IP communications, providing additionally resource reservations, micro-mobility control, and the stability of the connectivity through using same data path for the same data flows.

This procedure enables a simple mechanism for data traffic offloading, independent of the UE. By differentiating data flows based on traffic flow templates in both the ANGw and in the redirection information transmitted to FEs, the system can maintain two network paths through two PDN GWs from which one can be representing the offloading connection. Additionally, the flows can be separated and merged using the procedure here described.

### 6.4.2 Implementation Modules

For realizing the functionality described here, the internal structure of CPAE is based on the OpenEPC modules and their dependencies illustrated in Figure 6.12.

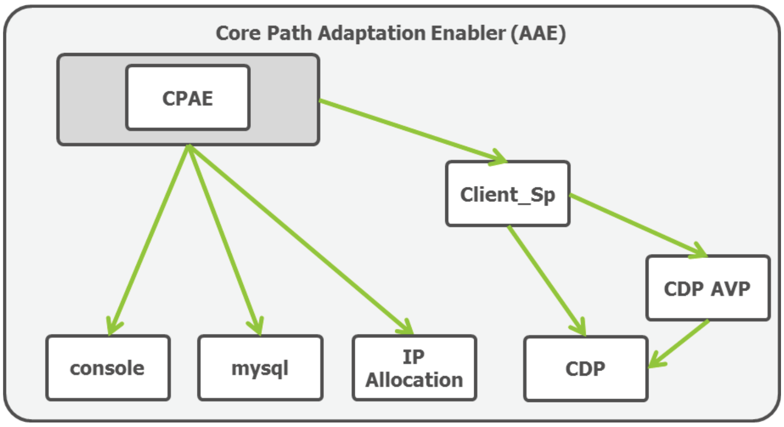


Figure 6.12: OpenEPC Modules for Core Path Adaptation Enabler

Similarly to the ASE, for the communication with the HSS, it uses the Client_Sp module implementing the interface specific state machine, the CDP_AVP implementing the interface specific communication tokens and the CDP module for the Diameter state machine. Additionally, for the communication with the PDN GWs and the FEs, CPAE includes the console module through which remote procedure calls can be executed. The CPAE state is synchronized in a mysql database

## 6.5 Application Adaptation Enabler

Based on the specification from Section 5.5, the Application Adaptation Enabler (AAE) is a network entity which is able to make decisions related to the resources that have to be reserved over the different access networks for adapting service delivery. As the current function for control in the 3GPP EPC core network is the

PCRF, making decisions synchronous with the signalling of the service establishment, modification or termination and as the AAE requires similar interfaces as the PCRF, for communication with the subscription entities and to the mobility and charging control in the operator core network, the AAE is considered to include the PCRF functionality in 3GPP EPC exemplification. The EPC interfaces of the AAE function, as realized in the reference implementation, are depicted in Figure 6.9.

The AAE reference implementation was designed as a main research feature available in the Fraunhofer FOKUS OpenEPC toolkit and disseminated through multiple license acquisitions by operators, vendors and academia.

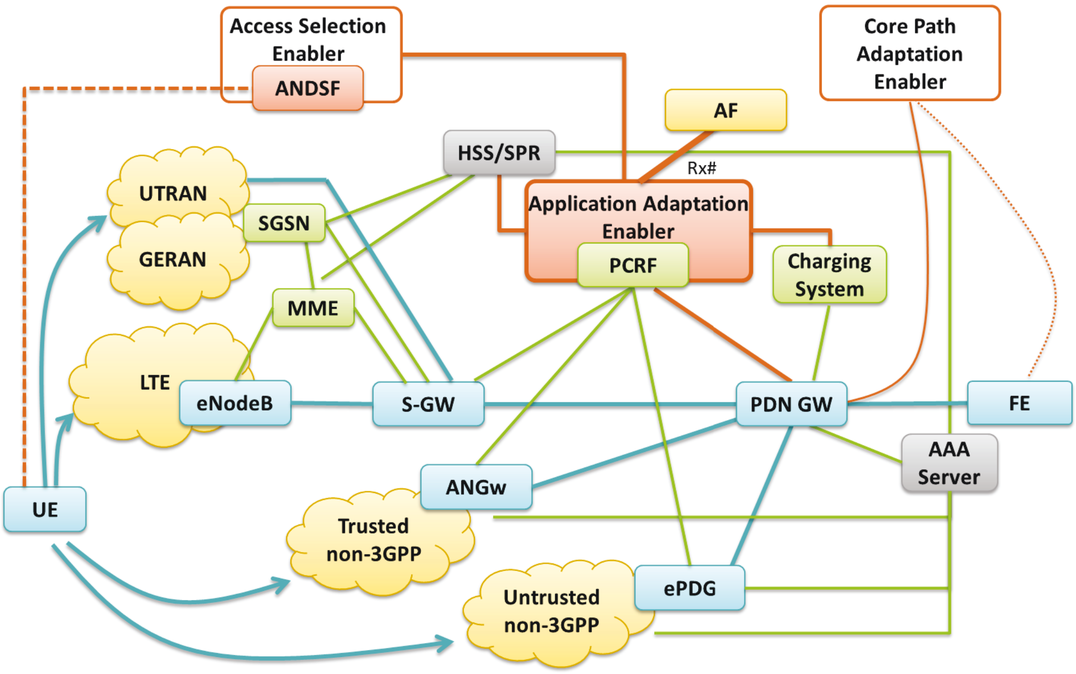


Figure 6.13: Application Adaptation Enabler

A large addition to the PCRF, brought by AAE is the management and provisioning interface offered towards the different service platforms and applications. This interface enables the service platforms to interact with the core network in a more flexible manner, enabling Future Internet applications to transmit their requirements and to receive data path events in a simplified manner.

The interface exposed does not require a strict binding between the different Application Functions and the core network, specifically the AAE, thus making it an appropriate choice for the current Over-The-Top applications, provided by the mobile network operators or by other third parties. A minimal self-adaptation functionality is expected to be deployed in the service platforms or in third party components, representing the service platforms, which enables the provisioning of the services at core network level. The interface is based on the 3GPP EPC Rx interface. The new Rx# interface enables the provisioning of multiple service profiles for specific network conditions, by this, removing the requirement that at each service establishment, modification or termination, the service platform has to communicate with the core network.

For the rest of the interfaces, the AAE relies on the existing PCRF interface. For the communication with the mobility control functionality, the 3GPP EPC Gx interface between the PCRF and the PDN GW is used. Through this interface, the AAE subscribes and receives notifications related to data path events, as considered by EPC standards. For retrieving and updating the network subscription profile, the AAE uses the 3GPP EPC Sp interface between PCRF and HSS/SPR repository.

As AAE includes the functionality of PCRF for functional efficiency, there is no need for a resource control interface. Otherwise, if the AAE is separated from the PCRF, the Rx interface would be used for the transmission of resource requirements and for receiving of data path events.

The internal AAE subscriber information maintains multiple service profiles for the same service from which one may be active. Similarly, the PCRF subscriber information maintains information related only to the active profile. For efficient handling of the subscriber information, the PCRF database is extended with a new "Active" flag for each of the service profiles and with a set of conditions when the profile is to become active. Based on this additional information, the current behavior can be maintained, while supporting the AAE.

### 6.5.1 Procedures and Data Flows

In this subsection, two example procedures, realized also in form of the prototype, are described: a procedure for service profile information provisioning in the AAE and a procedure for the data path resource adaptation execution. For the other procedures, the AAE relies on the existing Policy and Charging Control functionality and they are not further extended.

#### 6.5.1.1 Service Provisioning

This procedure is executed when an application or a service platform requests to provision multiple service profiles for a subscriber, addressing the delivery of the same data flows in different network conditions. The procedure is depicted in Figure 6.14.

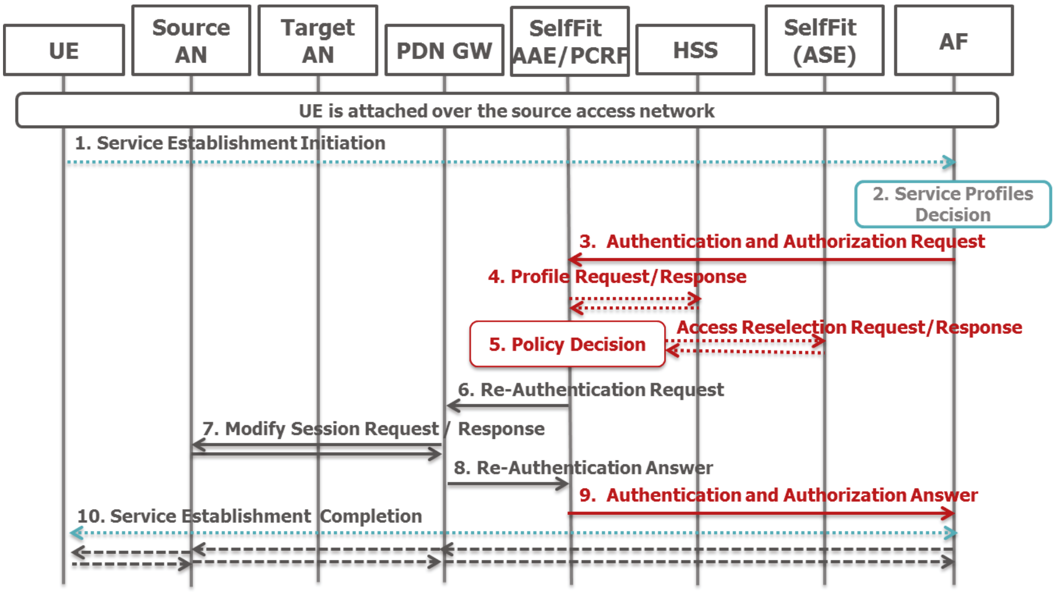


Figure 6.14: Application Adaptation Enabler - Service Profiles Provisioning

The procedure is triggered by the AF, due to the service initiation or due to the deployment of new UEs or of new services in the network.

When the procedure is triggered by the service establishment, the UE has to be attached to the network and has to be able to communicate with the service platform. In this case, the procedure has the role to optimize the signalling and the service during service modifications, due to network conditions.

When the procedure is triggered by the deployment of new UEs or of a new service, it is executed independently of the connectivity status of the UE. In this case, it additionally optimizes the signalling during the initial service establishment. However, it is required that the communication is executed over data flows having the same traffic flow template, thus that IP addresses and ports are fixed.

1. A service establishment procedure is initiated between the UE and the AF. The messages exchanged are specific for each application level protocol and are not further detailed here.

2. The AF makes a decision on which service profiles are to be used on the different network conditions, including the guaranteed and the maximum resources required and priority and pre-emption classes.

3. The service profiles are send by the AF to the AAE, using the standard Diameter Authentication-Authorization Request (AAR), containing multiple Media-Subcomponents each one describing a possible session profile, ordered

by the preference of the service provider. It also contains a registration to an event on the modifications on the network context of the UE, as for example an "IP-CAN Type", for receiving a notification when the UE connectivity changes. This addition to the AAR message enables the AAE to receive the multiple session profiles.

4. The AAE fetches the subscription profile information from the HSS, if not yet available.
5. The AAE makes a standard policy decision for the session parameters to be enforced on the data path. The other secondary service profiles are saved for further usage. Additionally, an access network reselection request may be transmitted to the SelfFit ASE for selecting a better access network for the subscriber from the perspective of the provisioned service.
6. The AAE transmits a standard Diameter Re-Authentication Request (RAR) to the PDN GW for reserving resources on the specific access network for the active services.
7. The resource reservation over the current access network is executed through the GTP "Modify Session Request/Response".
8. In order to notify the AAE of the completion of the reservation procedures, the PDN GW responds with a Re-Authentication Answer (RAA) to the PCRF. At this moment resources can be used by the specific service using the active profile.
9. An Authentication-Authorization Answer (AAA) is transmitted to the AF in order to notify it of the completion of the procedure. This step may be executed asynchronous to the resource reservation completion immediately after the policy decision is taken in Step 5. In this case, no acknowledgement of the resource reservations is received by the application. Additionally, in case the resource reservations fail, the service termination procedure network initiated has to be executed.
10. The AF and the UE complete the service establishment procedure and they exchange data with the resources available in the active profile

In case the service provisioning is made independent of the service establishment, no resources have to be reserved. The role of the procedure is only to bring service profiles to the network. The Steps 1 and 6-10 do not have to be executed.

### 6.5.1.2 Automatic Application Adaptation

This procedure automatically adapts the reserved resources for a specific service based on the service profiles stored at a previous provisioning phase. The procedure is triggered by a change in connectivity of the specific UE, for example the attachment to a new access network, as initial connectivity or as part of a handover.

For an initial connectivity, at the end of the procedure, the resources reserved include the default bearer and dedicated bearers for the specific services for which the provisioning procedure was executed.

For the handover, at the end of the procedure, the resources reserved include the default bearer, the adapted dedicated bearers for the specific services for which previously a provisioning procedure was executed and the other dedicated bearers for the other applications with the same level of resources as over the source access network. The handover procedure for a specific service which is adapted is depicted in Figure 6.15 and presented in the following.

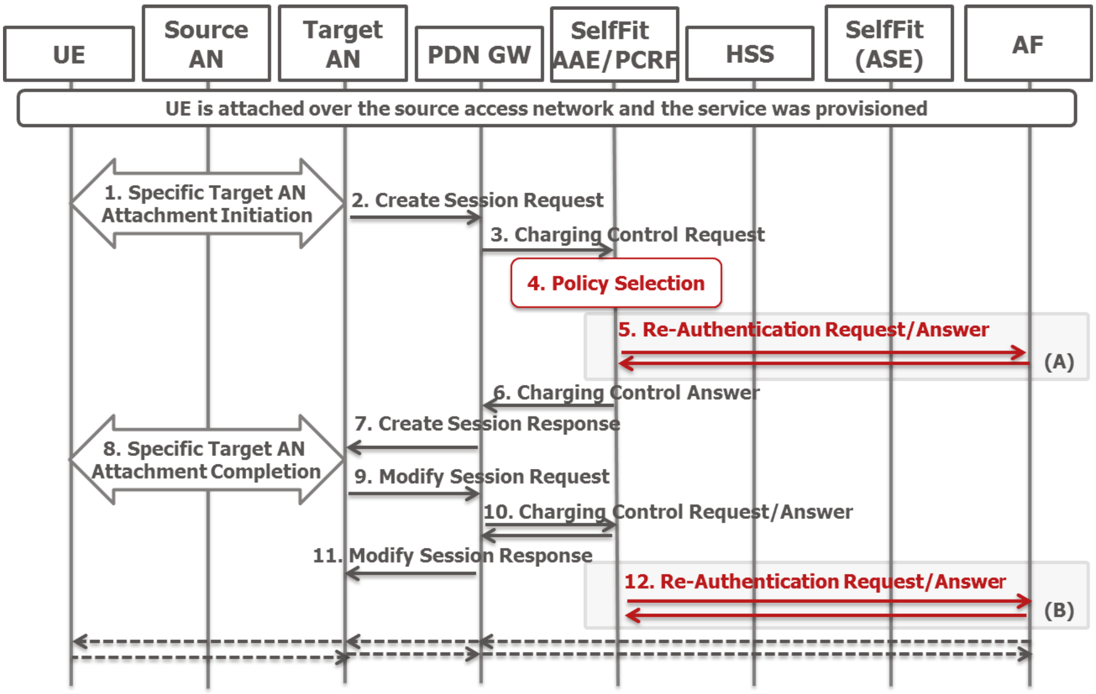


Figure 6.15: Application Adaptation Enabler - Automatic Application Adaptation

The UE does not have to be remotely notified on the change in service profile as it receives this information during the provisioning phase directly from the AF. This information, correlated with the notifications on access network changes which may be received from the operating system, enable the UE to adapt its part in the service delivery without any signaling to the network.

1. The UE executes a standard attachment procedure to the target network.
2. The gateway of the specific access network transmits a GTP "Create Session Request" to the PDN GW requiring the establishment of a data path context including mobility and resource control support.
3. The PDN GW announces through a Diameter Charging Control Request (CCR) that the UE executed a handover to the target access network and queries which resources have to be reserved over the target access network.
4. For the services for which a provisioning procedure was executed, the PCRF selects the most appropriate service profile to be used over the target access network. Using this service profile, it makes a policy decision on the resources that have to be reserved. The resulting policies are grouped with the policies for the other adapted services, with the unmodified policies for the non-adaptable services, and the default bearer.
5. A Re-authentication Request (RAR) is transmitted to the AF in order to notify it that the resources reserved were adapted. The AF acknowledges the event notification with a Re-authentication Answer (RAA).

   The operation is optimal to be executed before the resource reservation when the resources of the service are downgraded, as the resource reservation not to affect delivery. For other services, the same operation will be executed after the resources reservation.
6. The PCRF responds to the PDN GW with a Charging Control Answer (CCA) including the combined policies to be reserved on the data path.
7. The PDN GW forwards the policy information to be installed to the target access network as part of the GTP "Create Session Response".
8. The specific target access network attachment is completed and the resources are reserved on the radio and on the access network specific gateway.
9. A GTP "Modify Bearer Request" is transmitted by the target access network to modify the reserved resources and to notify the PDN GW that the attachment procedure was completed and that the resources were reserved.
10. On its turn the PDN GW transmits a Charging Control Request (CCR) to the AAE in order to notify the completion of the attachment and of the resource reservation procedure which is responded with a Charging Control Answer (CCA).

11. A GTP "Modify Bearer Response" is transmitted from the PDN GW to the target access network to notify the receive of the attachment completion message.
12. When the AAE receives the notification that the attachment was completed, then a notification may be send to the AFs which required service adaptation as to adapt the data delivery. This notification may be send instead of the one in Step 5 for the services in which the adapted resources are higher as the ones reserved over the source access network. At this moment the communication is executed with the adapted level of resources.

The reference implementation was realized for adaptation of active services enabling the usage of multiple session profiles while handing over between heterogeneous access networks such as emulated LTE, public operator UMTS and WiFi.

The adaptation procedure reduces the number of steps used in the 3GPP standard though the execution of a provisioning phase. In the current standard version, all the steps of the here described procedure are executed except for the selection of a service profile. However, this does not result in a service adaptation. For this Steps 2-10 of the previous procedure have to be executed considering a single service profile, doubling the number of exchanged messages, policy decisions and resource reservation procedures executed by the data path entities.

This procedure addresses specific deployments in which the adaptation is required for the specific services at least one time during their runtime. It includes the cases in which the UE can connect to more than one access network and for services which are flexible to be adapted, such as video streams.

### 6.5.2 Implementation Modules

The AAE functionality is based on the communication interfaces of the PCRF, thus similar OpenEPC modules are considered, as illustrated in Figure 6.16.

The AAE module is co-located with the PCRF module, allowing the inline extension of the PCRF state machine by introducing the afferent AAE states. For the communication with the other network entities, the Diameter protocol stack is used based on the CDP Diameter protocol module and the CDP_AVP coded module. Additional PCRF modules including the management of the state machines for the interfaces are included: PCRF_Rx, PCRF_Gx, PCRF_Gxx and Client_Sp. Additionally, the console module is available for debugging purposes and the mysql for state synchronization.

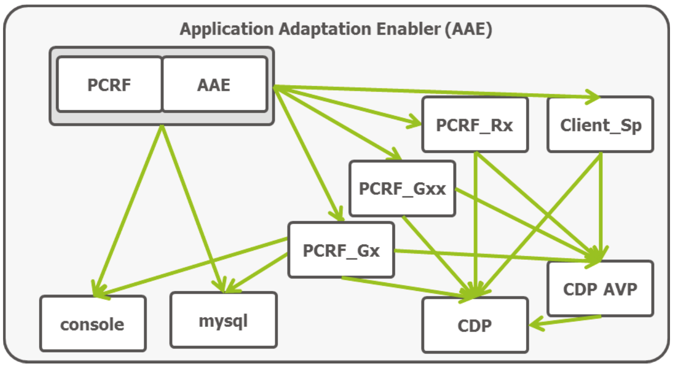


Figure 6.16: OpenEPC Modules for Application Adaptation Enabler

## 6.6 Discussion

Following the specification of the SelfFit framework from Chapter 5, in this chapter a prototype solution addressing the 3GPP EPC architecture was proposed. The prototype consists of three enablers, tightly coupled to the EPC architecture and loosely coupled between themselves. For a specific deployment to use the adaptation only for a specific part of the communication of the device, it can include a single enabler.

The enablers are mapped to the 3GPP EPC architecture, depending on their main functionality and depending on the synergistic integration with a selected control entity already available in the core network. The ASE was associated with the ANDSF functionality, due to the communication with the UEs which is currently possible in the EPC only through the ANDSF. The AAE uses the functionality of the PCRF for executing resource reservations during service establishment and during handovers. For the CPAE, it is considered optimal to develop a new function and then through large scale trials and pilots to determine the feasibility of its placing in another network entity. For this association perspective, the SelfFit management functionality here proposed is an extension of the control functionality towards introducing a new level of flexibility for the parameters, which usually are static to the policy decisions.

From the operator core network management perspective, the SelfFit enablers present three mechanisms of interpreting the modifications in the network topology. For orchestrating the three mechanisms, a set of explicit and implicit interconnections is considered, as specified in Chapter 5. The explicit mechanisms are illustrated through the procedures deployed in the SelfFit realization.

The realization is presented mainly from the perspective of the supported messages and procedures, making possible distinct realizations of the enablers. The practical implementation here presented is based on the Fraunhofer OpenEPC toolkit and on the already implemented interfaces, adapted to the specific requirements of the SelfFit framework.

In the next chapter, the SelfFit framework is validated in different deployment scenarios and used for performance evaluation through comparison with the current architecture, which includes only the 3GPP EPC.

Chapter 7

# Validation and Evaluation

The following sections describe the validation and the evaluation of the SelfFit solution.

## 7.1 Introduction

This chapter includes the validation and the evaluation of the SelfFit system specified in Chapter 5 and based on the implementation described in Chapter 6. For this, a series of testbed deployments were considered in various vendor and operator environments as well as in the Fraunhofer laboratories.

First a set of performance evaluations are presented as conducted under different system conditions comparing current available solutions with the newly implemented SelfFit features and the duration of the proposed procedures. Based on these evaluations, a general impression can be grasped on the value of the various features.

Secondly, the SelfFit framework is compared with the other available solutions both from standardization and from academia. The evaluation considers a complete system, even though some of the solutions available are covering only part of the required functionality. As the SelfFit framework is covering a large research area, the solutions which are very similar or are bringing optimization to a specific standard are considered as a whole considering the best parameters possible. The qualitative evaluation is based on the metrics derived from the requirements as described in Chapter 3.

## 7.2 Proof-of Concept Validation and Performance Evaluation

In this section the results obtained from the practical implementation of the SelfFit enablers on top of the Fraunhofer FOKUS OpenEPC are assessed.

In order to grasp the SelfFit specific features, the measurements were performed as much as possible separately per enabler and further even more separated per feature. Different metrics were used in order to evaluate the delay of the specific procedures. The same structure is also considered for this section. As the evaluation was done using proof of concept testbeds and as no automatic network capacity loading mechanism was considered, a limited number of tests were performed.

As different features were developed via different projects having specific limitations in the additional EPC functionality deployed, for each of the test scenario the considered deployment is separately described.

### 7.2.1 Access Selection Enabler Evaluation

The proof-of concept implementation of the SelfFit ASE was realized in the form of a complete operator core network testbed targeting the feasibility of new features for standardization in the area of core networks.

The testbed implemented the ASE functionality as designed and specified in Chapter 5 and realizes the procedures as described in Chapter 6 in the section addressing the ASE implementation.

As specific technology and OpenEPC development features were missing at the moment of the proof-of concept testbed realization, a reduced carrier grade operator software realization was considered as depicted in Figure 7.1.

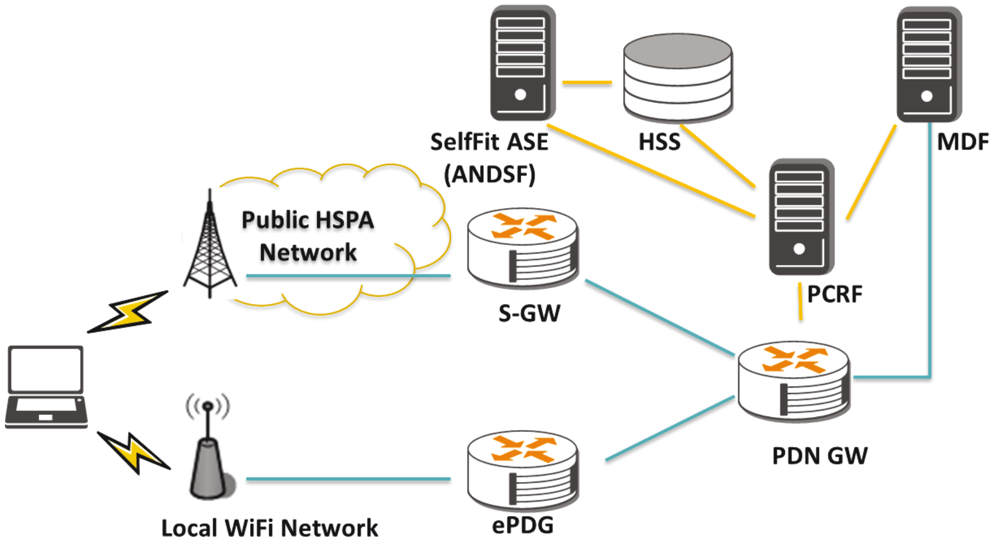


Figure 7.1: Access Selection Enabler Validation Testbed

As no access to controllable 3GPP radio access technologies was possible, the testbed uses WiFi as trusted non-3GPP access and a public HSPA network as untrusted non-3GPP access. For obtaining a proof-of concept, instead of the standard IPsec, OpenVPN over UDP was used. This functionality change in the testbed does not affect the communication except for the MTU values. In order to avoid IP packet fragmentation, the MTU of the data packets exchanged from the MN the network was reduced as to fit together with the additional VPN header through the public operator network.

Two gateways were deployed: a "S-GW" for the HSPA access network having the functionality of a trusted non-3GPP gateway with additional VPN server capabilities and an "ePDG" for the WiFi access network also having the functionality of a trusted non-3GPP access network.

A PDN GW was deployed, enabling the anchoring of the data sessions. Between the access network gateways and the PDN GW, PMIPv6 protocol was employed. Additionally, as no real 3GPP networks were used, the attachment procedures were triggered by using DHCP requests from the MN. This operation makes transparent to the evaluation the link layer attachments.

The testbed includes a PCRF, connected to an HSS and a minimal Media Delivery Function (MDF), enabling the establishment of a video service with the MN.

Apart from these components, which were integrated from OpenEPC toolkit, the SelfFit ASE was introduced into the system, as described in Chapter 6.

#### 7.2.1.1 Resource Oriented Access Network Selection

For the evaluation of the Resource Oriented Access Network Selection procedure, described in Section 6.3.2.2, it is considered as initial setup that the UE is already connected to the HSPA network without having an active service. The resources over the HSPA network are administratively limited as not to support an active video session. Using the Session Initiation Protocol (SIP) INVITE transaction, the UE initiates the establishment of a video session with the MDF.

Upon receiving the information, the PCRF checks the available resources over the HSPA network, administratively controlled in the test environment. This operation is equivalent with a resource reservation procedure, which is expected in a real deployment to have a longer duration than obtained through the evaluation.

An indication is transmitted to the SelfFit ASE which takes the decision to select the WiFi access network for the UE. When the UE attaches to the target WiFi, the video session is initiated.

Based on this minimal scenario, at the following entities, the different procedure delays were measured:

- UE - the duration of all the procedures was measured from the perspective of the mobile user including the time intervals for the establishment of the video session, the communication with the SelfFit ASE and the WiFi network attachment.

- PCRF - the duration of the internal procedures and the communication with the SelfFit ASE.
- ASE - the duration of the decision making and of the communication procedures with the UE and with the PCRF.

The measurement results, split per procedure step, are depicted in Figure 7.2. A previous version of the measurements, using a simplified testbed, is available at [51]. The results slightly vary, due to a closer-to-reality functional separation and due to the development of the testbed based on the OpenEPC toolkit.

A set of 50 measurements were executed independently for each of the two testbeds in similar network conditions. A single device was used and no artificial congestion was introduced, thus no limitation of the testbed resources in computation or storage was reached.

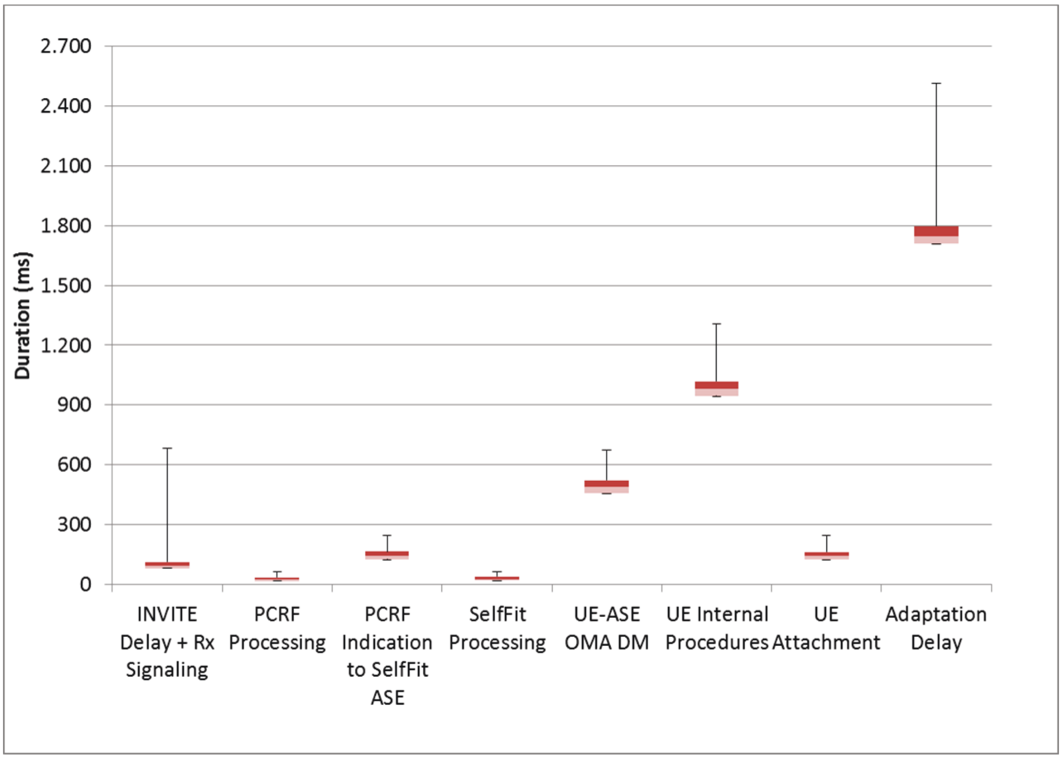


Figure 7.2: Resource Oriented Access Network Selection - Measured Values

The service establishment procedure including the access network selection had a median delay of 1.7s in an interval between 1.4s and 2.5s, highly varying due to the wireless environment, both over the HSPA and the WiFi network. As depicted, the internal network operations consume less than 10% of the complete procedure, while a major part is taken by the communication between the ASE and the UE and by the attachment to the WiFi network.

The 10% of the network operations include the SIP session signaling over the HSPA network, the signaling between the video server and the EPC, the PCRF and the SelfFit communication and processing. In a carrier grade operator network, having a high number of MNs connected, these operations would be slightly longer, even though specialized hardware will be used.

The rest of the procedure includes the communication between ASE and UE and the UE attachment to the target access network. The communication between the UE and the ASE has 28% of the overall procedure. This is due to the three-way communication of the OMA DM for the PUSH mode which is executed over the HSPA network. Using an own controlled network would highly reduce the duration of this operation, as the data path does not include the Internet backbone anymore.

The major part of the delay is due to the internal processing in the UE, having a proportion of 56%. This is mainly due to the processing of the policies and to the update of the routing tables. All the respective operations were serialized as the UE was not in the main focus of the proof-of concept.

The rest of the delay is due to the attachment to the target access network. It includes the DHCP over WiFi procedure, the PMIP, and the resource reservations.

In these conditions, it is expected that for each of the procedures a Gauss distribution will be obtained. An overall procedure standard deviation of 180ms was obtained, which is rather reduced due to the stable network conditions during experimentation.

As a software testbed was used on very similar network conditions, the measured results have a small variation in the core network, mainly due to the retransmission of the Diameter messages, resulting in a standard deviation of maximum 32ms per procedure step.

A larger variation was obtained in the mobile device and in the communication over the wireless network as the resources available were not completely controlled during the experiments. A standard deviation of maximum 130ms per procedure step was obtained.

The maximum acceptable delay for IMS session establishment is of 4 seconds [225]. Thus the maximum delay obtained through the measurements is less than half of the standard benchmark for IMS services, making the solution feasible to be deployed in real-life systems. Although the authentication and resource reservation operations were not included, it is expected that their duration, as well as the duration of the operations measured will be reduced through the deployment of specialized operator hardware.

Additionally, the procedure does not affect the quality of the other active services of the mobile device.

To be noted, that through the EPC functionality, the vertical handover procedure is transparent to CNs, thus not affecting the delivery of the other active services of the MN.

The procedure enables the establishment of the session which otherwise would not be possible when using state of the art technologies. Therefore, this procedure should be regarded as a feature added to the core network architecture which at the cost of an acceptable delay and processing power consumption, enables a lower number of sessions to be terminated due to lack of resources.

#### 7.2.1.2 Volatile-Shallow Context

For the evaluation of the effects of the volatile-shallow context establishment, procedure described in Section 5.3.3.3, it was considered that the UE is connected to the WiFi network. An indication is transmitted by the ASE to the UE that, in case an application is started to use the HSPA network instead. At the same time, an indication is transmitted to the S-GW to establish a shallow context.

For this evaluation, a shallow context contains information on the IP address of the UE and the establishment of an inactive PMIPv6 tunnel between S-GW and PDN GW. The shallow context does not include resource reservation and authentication information.

When the application is initiated, the UE attaches to the HSPA access network as part of a handover. Due to the volatile-shallow context, the attachment duration is highly reduced.

Based on this scenario, measurements were performed in the UE and the S-GW for the attachment delay, with and without having a shallow context pre-installed.

A set of 50 measurements were executed independently in similar network conditions. A single device was used and no artificial congestion was introduced, thus no limitation of the testbed resources in computation or storage was reached.

The measurement results, per procedure and per entity, are depicted in Figure 7.3. Similar results were obtained at a previous evaluation, as available from [50].

The values obtained show a deduction of the procedure delay of 100ms, representing around 90% on the network side and 45% on the UE side, including also the delivery over the wireless access system.

For the measured procedures it is expected that a Gauss distribution will be obtained. For the state of the art case a standard deviation of 30ms was obtained with a 22ms deviation for the optimized procedure, which is reduced due to the stable radio and core network conditions during experimentation.

Although the procedure was measured for a single MN and for a limited set of operations, the obtained values prove the optimization proposed. A complete shal-

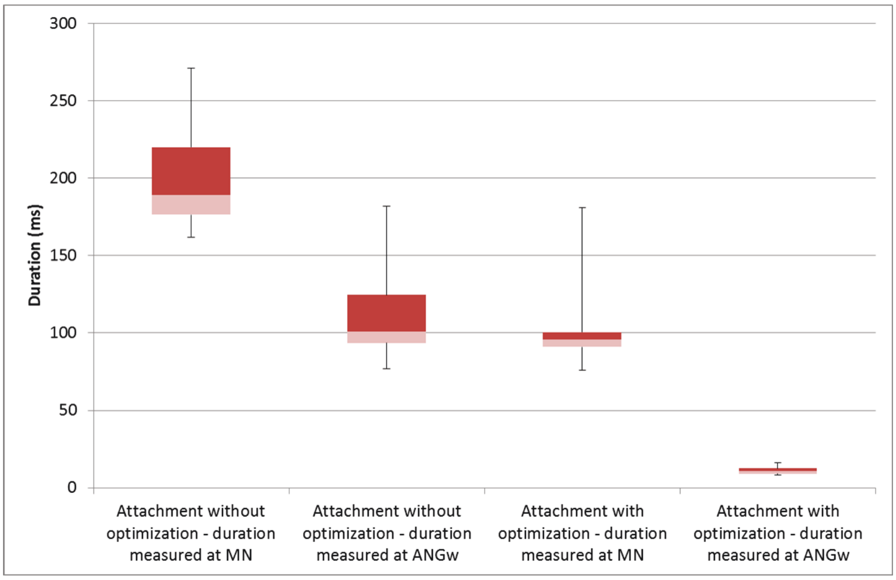


Figure 7.3: Volatile Shallow Context - Measured Values

low context would also include the pre-authentication, subscriber data retrieval, and resource information retrieval in the access network gateways.

When adding the other operations, it is expected that similar optimization will be obtained, with a percent around 50% of the overall attachment time.

This brings the opportunity to use the shallow-context for handovers between the access networks where the communication is suddenly lost to access networks where the attachment duration is high.

Therefore, even though it reduces the overall procedure, the shallow context requires a pre-establishment procedure, which has to be executed. The opportunity of using a shallow context in real deployments depends mainly on the probability that the UE will execute a handover to the specific access network and on the characteristics of the target radio technology.

### 7.2.2 Core Path Adaptation Enabler Evaluation

The proof-of concept implementation of the SelfFit CPAE was realized in the form of a complete operator core network testbed, targeting the feasibility of deploying transparent mobility of subscribers between multiple PDN GWs.

The testbed implements the CPAE functionality as designed and specified in Chapter 5 and realizes the procedures as described in Chapter 6.

As specific technology and OpenEPC features were missing and as the CPAE is data path oriented, for the proof-of concept testbed realization a reduced carrier grade operator network was realized as depicted in Figure 7.4.

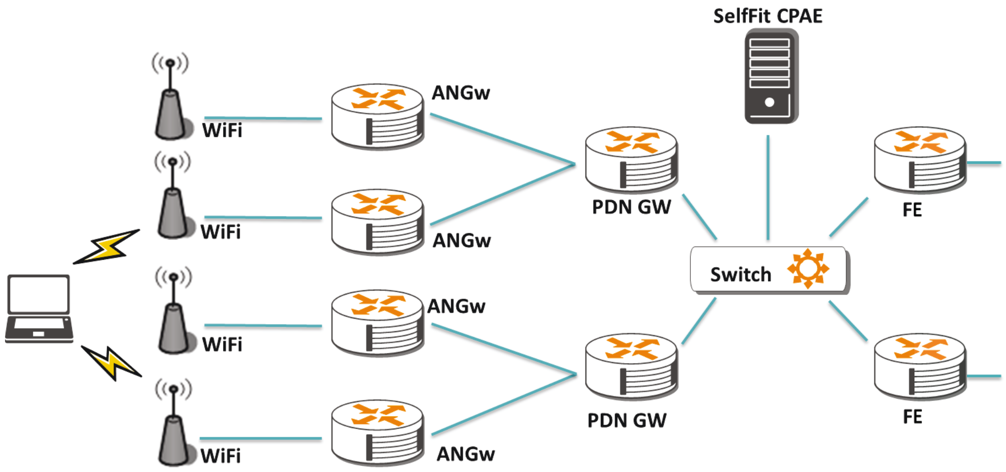


Figure 7.4: Core Path Adaptation Testbed

As the core network adaptation is independent of the radio access networks deployed, the testbed uses four WiFi access networks, which are connected each to an own access network gateway. Through this a large network operator is emulated enabling the establishment of a physical topology considering network distances.

Two PDN GWs are included in the testbed. The first two ANs are considered better served by the first PDN GW, while the other two ANs are considered better served by the second PDN GW.

Apart from this OpenEPC functionality, specific to this testbed, two FEs were deployed in the testbed, inside the operator network and beyond the PDN GWs. The FEs are able to redirect the downlink data traffic to the appropriate PDN GW.

For controlling the redirection, the SelfFit CPAE was deployed in a separate component, as considered by the SelfFit mapping from Chapter 6. The SelfFit CPAE component was enabled to support the procedures for flexible PDN GW selection, with and without access network change, as described in Section 6.4.1.1 and in Section 6.4.1.2. For notifying the FEs, the synchronous mechanism was deployed.

As the functionality proposed by this dissertation is addressing the data path adaptation, the policy and charging control even though present in the testbed was not further developed and synchronized i.e. to enable the transfer of resource reservations between the PDN GWs. Instead, a default bearer was used for all the devices.

The following scenarios were validated and measured:

- PDN GW reselection with access network change - the UE is handing over between the second and the third WiFi access networks. A data path reselection is automatically triggered.

- PDN GW reselection without access network change - the PDN GW is reselected for a specific subscriber transparent to the UE and to its radio communication characteristics.

For the evaluation of the core path adaptation, the procedure of PDN GW reselection without access network change as described in Section 6.4.1.2 is further assessed. Similar delay was observed for the PDN GW reselection with access network change. When remarkable differences occurred, they are underlined in the following.

The measured scenario considers that the UE has a connection through an access network and through the source PDN GW. An administrative trigger is given to the SelfFit CPAE to reselect the PDN GW for the specific UE. Based on this trigger, the SelfFit CPAE orchestrates the data path adaptation through the communication with the source and target PDN GWs. The reselection is notified to the FEs.

Based on this minimal scenario, the different procedure delays were measured at the SelfFit CPAE. To enable the measurements, an acknowledgment message was introduced for the "Perform Handover" and the notifications of the FEs messages. It is assumed that a delay of less than 10ms was introduced to the overall procedure by these synchronization messages. The measurement results split per procedure step are depicted in Figure 7.5.

A set of 30 measurements were executed independently in similar network conditions. A single device was used and no artificial congestion was introduced, thus no limitation of the testbed resources in computation or storage was reached.

The "Prepare Handover/Ready" message exchange between the CPAE and the Target PDN GW has a median of 16ms including the communication and the introduction of the binding between the UE and the IP address as part of the information in the Target PDN GW.

The "Perform Handover" and the synchronization acknowledgement had a median of 10ms due to the communication between the CPAE and the Source PDN GW and due to the information modification in the Source PDN GW.

The Source PDN GW retrieves the information on the AN GW of the UE and transmits a "Proxy Binding Update" modified message to terminate adapt the data path in the AN Gw. A new PBU is transmitted to the Target PDN GW which synchronizes with the SelfFit CPAE. The duration of this data path adaptation has a median of 50ms, due to the multiple operations that have to be executed.

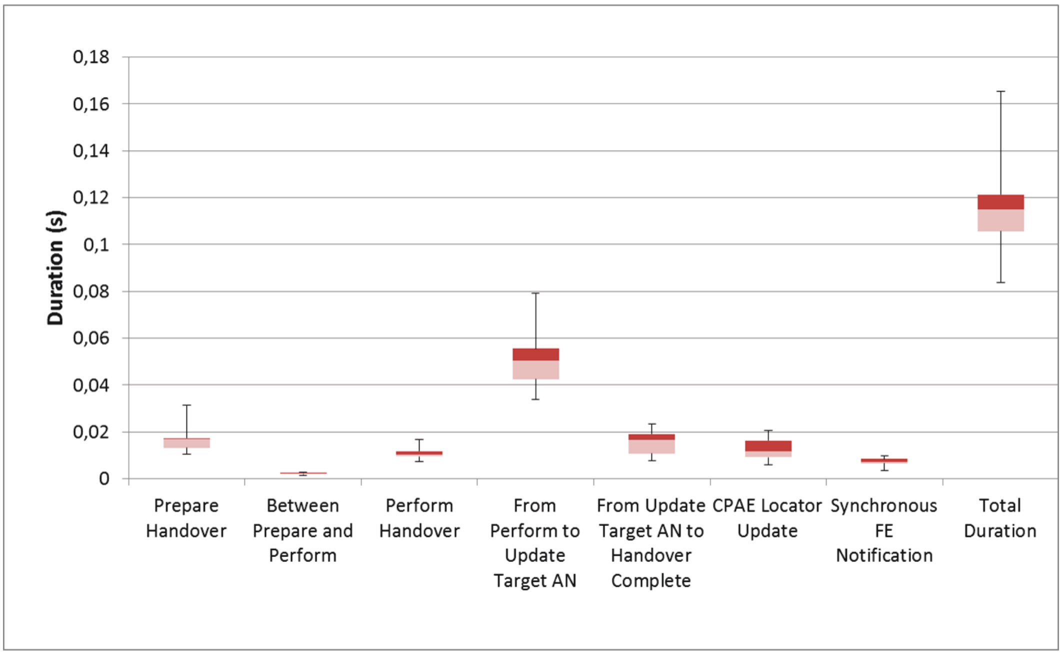


Figure 7.5: Core Path Adaptation Enabler - Measured Values

The completion of the adaptation has a median of 16ms, followed by the update of the UE state information inside CPAE with a median of 11ms.

The PDN GW reselection orchestration has in total a duration median of 107ms, including the internal operations in the various entities and the communication between them. This duration represents an estimation of the software implementation on top of the OpenEPC Wharf platform. For a real-life vendor deployment, the proportions of this median delay may vary, due to the implementation on dedicated hardware, which brings specific optimizations, due to the parallel support of a large number of subscribers, and due to larger network transmission delays between the different sites of the operator.

On top of this delay, the notification of the FEs should be considered with a median of 7ms. The low duration is due to the implementation using a simple communication protocol, as recommended also for real deployments, in order to mitigate the large amount of notifications which have to be exchanged.

As the data path through different PDN GWs has different delay, this factor may be also visible in the communication of the UE. Otherwise, no packets are lost, due to the redirection of the data traffic between the Source and the Target PDN GW.

For the measured procedures it is expected that a Gauss distribution is obtained. As only core network software procedures and communication over Ethernet connections are measured, the standard deviation is low up to 10ms and 24ms for the complete procedure.

For the state of the art case a standard deviation of 30ms was obtained with a 22ms deviation for the optimized procedure, which is reduced due to the stable radio and core network conditions during experimentation.

The procedure for PDN GW reselection with access network change has similar duration for each of the respective operations. However, as the Source PDN GW does not have to announce the Target ANGw to initiate an redirection procedure, the overall median delay is reduced with approximately 10ms. This delay does not consider the attachment to the target access network.

In case of a vertical handover, the overall delay can be tolerated as the communication of the source access network is maintained until the path through the target access network is completed.

In case of horizontal handovers, considering the delay of the here described evaluation, it is recommended to first execute the attachment to the target access network gateway, using the source PDN GW and then to transparently execute the core path reselection.

### 7.2.3 Application Adaptation Enabler Evaluation

The proof-of concept implementation of the SelfFit AAE was realized as an extension of the OpenEPC toolkit available and testbed by all the OpenEPC partners since Rel. 1 and further maintained by Fraunhofer FOKUS. For AAE evaluation, a typical OpenEPC testbed was deployed as described also in [52] and its last version depicted in Figure 7.6.

The testbed implements the AAE functionality on top of the OpenEPC PCRF, as designed and specified in Chapter 5 and realizes the procedures as described in Chapter 6.

The testbed includes the same components as a typical OpenEPC testbed as described in Section 6.2.

Additionally, the Media Delivery Function (MDF) was extended to support multiple service profiles for the downlink video streaming service and to receive notifications on events happening at the data path level. However, in order to be able to develop fast such a service and to make it available as part of OpenEPC, the state of the art case was not developed. Thus, here an estimation of its delay is considered.

The AAE functionality follows the specification description from Chapter 5. The AAE was enabled to support the service provisioning during service establishment and the automatic application adaptation support procedures as described in Section 6.5.1.1 and Section 6.5.1.2.

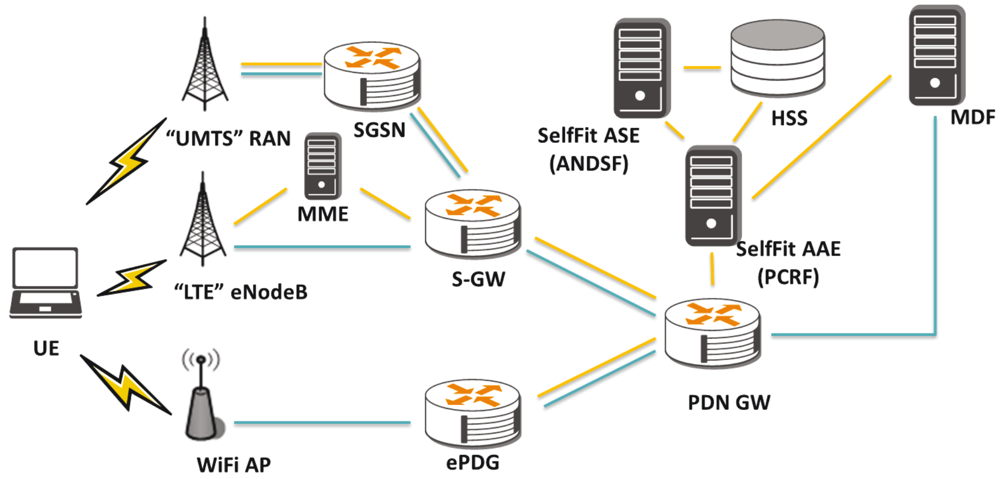


Figure 7.6: Application Adaptation Testbed

For adapting the video streaming resource reservations, the SelfFit AAE was developed on top of the OpenEPC PCRF, considering two video profiles of the MDF. As the actual selection of a profile depends on the internal implementation of the policy engine in the PCRF and as the target of the AAE is the adaptation of the video stream itself, the information on the two profiles was pre-provisioned in the PCRF.

The following scenarios were validated and measured:

- Service Provisioning during Service Establishment - during the service establishment, multiple service profiles are sent to the network. The first one is enforced while the rest are considered secondary.
- Automatic Application Adaptation - during a handover, the resources allocated to the UE are adapted, based on the application context. An event notification is transmitted to the application, enabling it to adapt the video stream.

Based on the video service scenario, the different procedure delays were measured at the PCRF and at the correspondent parties. As the communication between the entities has a median delay of less than 2ms in the testbed, for simplicity only the measurements in the PCRF are presented. The measurement results, split per procedure step, are depicted in Figure 7.7.

A set of 20 measurements were executed independently in similar network conditions. A single device was used and no artificial congestion was introduced, thus no limitation of the testbed resources in computation or storage was reached.

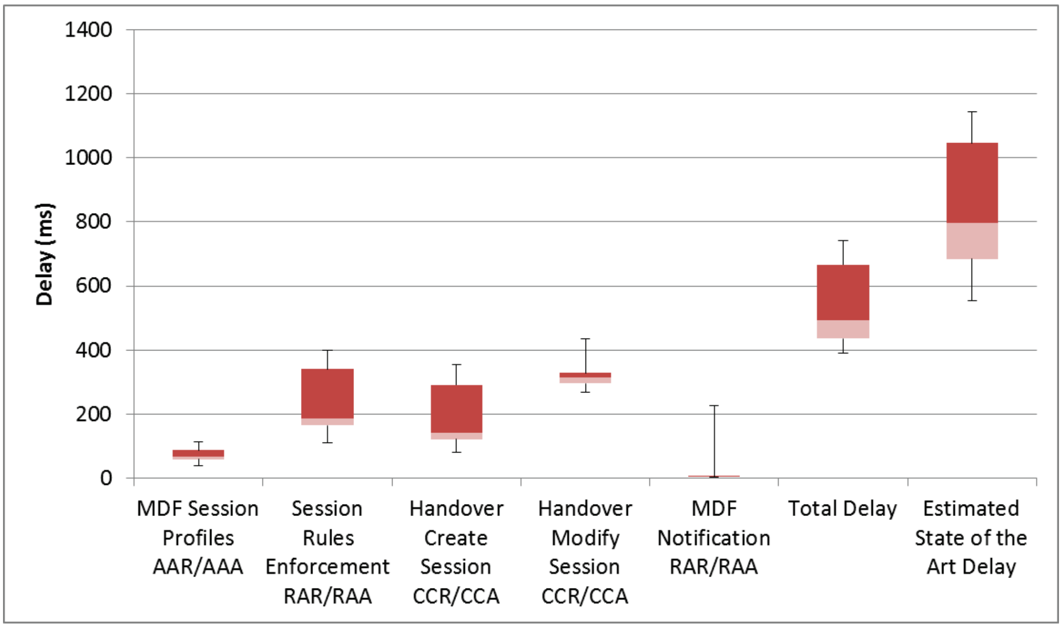


Figure 7.7: Application Adaptation Enabler - Measured Values

As the same number of messages is considered, the communication over the Rx# interface of the extended session profiles during the establishment of the service has similar duration to the communication over the Rx in case of state of the art service establishment with a median 68ms. This includes only the communication over the Rx interface and does not include the enforcement of the specific bearers.

The enforcement of the specific bearers on the PCEF has a median of 187ms. This includes the installation of the PCC rules in the PDN GW. It is assumed that a similar delay is obtained by the resource reservation procedures on the BBERF function of the ANGw. These two operations are executed in parallel. This operation was not modified by the AAE, thus the same delay is assumed in the current standard solution.

For the measured procedures it is expected that a Gauss distribution is obtained. The standard deviation per procedure step is up to 105ms with a standard deviation of and 124ms for the complete procedure while the state of the art evaluation results in 202ms.

From the perspective of the service establishment, the here proposed solution and the state of the art one do not significantly differ in delay having a median of 200ms. Thus, it is assumed that the provisioning of multiple profiles during the session initiation is not introducing a high overhead, except for the extended stored information in the AAE enabler.

During the handover, the PDN GW will query for the resources to be reserved for the UE twice: once for the creation of the data path with a median of 141ms and once for the modification of the bearer with a median of 315ms. The high difference is due to the specific feature implementation in the OpenEPC PCRF and not related to the AAE functionality. In these steps, the modified QoS rules are transmitted to the PCRF. With a similar delay, in the state of the art solution, the QoS from the source access network are translated to the target access network.

Additionally, an event notification is transmitted to the MDF, operation which has a median of 6ms. However, due to the underlying Diameter protocol stack used, implemented as software, this operation may have longer delays up to 200ms.

These operations conclude the automatic adaptation of the resources leading to a median duration of 492ms.

For the state of the art solution, after receiving the event from the PCRF, the MDF has to make a decision and transmit a new set of the resource requirements which are then translated into QoS policies by the PCRF and installed on the data path. It is estimated that these two steps are equivalent in duration with the service establishment procedure with a median delay of 255ms. Thus, the estimated service in the state of the art case has a median of 747ms.

The automatic adaptation procedure here presented, among with the reduced signaling and decisions at the application and network level, brings a delay optimization of 33%, which makes it highly desirable to be deployed in carrier grade operator environments.

## 7.3 Comparison with other Solutions

In this section, the SelfFit framework is critically assessed against the developments proposed by other researchers as part of individual solutions, standard solutions, and standard extension solutions. The approaches are classified based on the core network functionality they address and compared with the equivalent SelfFit enabler.

For each of the technologies, a qualitative assessment is presented, based on the evaluation criteria decanted from the requirements for future communication as summarized in Section 3.4. For the comparison of the heterogeneous architecture designs, for each evaluation criterion a metric was created enabling a quantitative representation. Based on these metrics, an expert evaluation was performed. The results of this evaluation are presented in this section. In this evaluation the subjective factor should not be underestimated, as it was realized by the same person which realized the SelfFit solution.

### 7.3.1 Access Network Selection Comparison

In this section, the SelfFit ASE is compared with the other solutions from the literature, having similar goals. The comparison is based on the evaluation from Section 5.1, the evaluation criteria from Chapter 3 and the metrics defined in Appendix A.2.1. The qualitative comparison is graphically illustrated in Figure 7.8 based on the data from Appendix A.2.1.

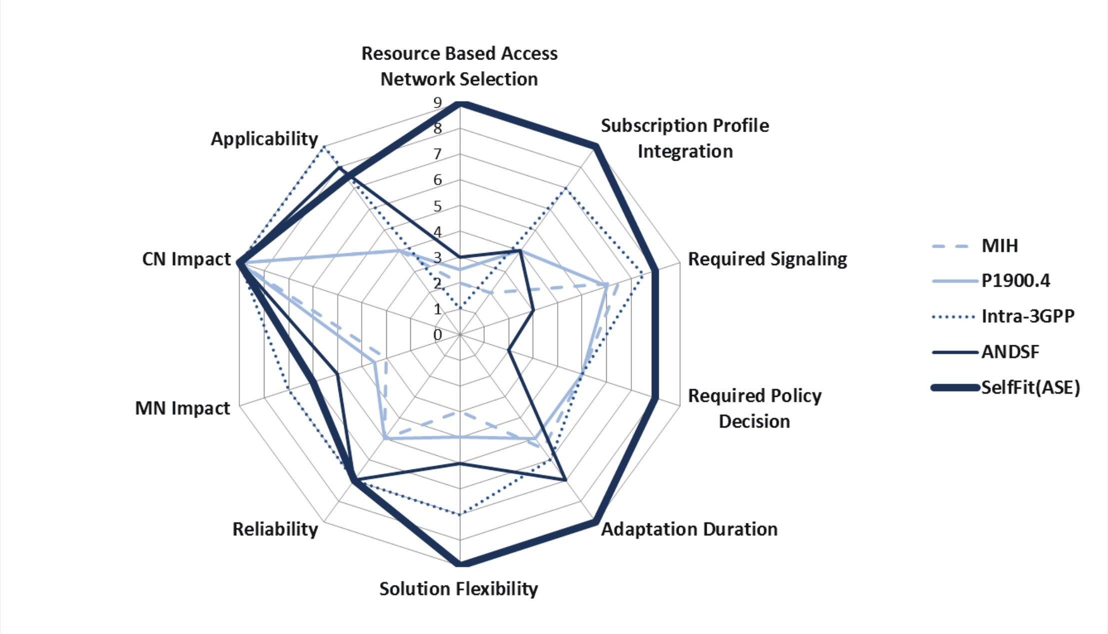


Figure 7.8: Access Selection Enabler Qualitative Comparison Representation

#### 7.3.1.1 Media Independent Handover (MIH)

MIH offers a solution for access network detection and selection based on link layer events and commands and on exchange of information with the network, as presented in Section 4.1.1.

Compared to MIH, Selfit ASE is interconnected with network subscription, network mobility, and resources support functions, enabling it to adapt the access network based on the required and available resources in the networks and based on the user requirements, which is not possible with MIH mainly due to its limitation to link layer information communication.

The various extensions available in the literature concentrate exclusively on the availability of the required resources in the target access networks, not considering that an access network reselection may be triggered or required by a modification in the resources required by the MN.

Additionally, compared to ASE, it requires a rather large signaling and number of policy decisions, for each access network selection a decision has to be taken by the network and a new set of discovery and selection information has to be transmitted to the MN. This also increases the duration of the procedure.

As the ASE acts at application level, it can achieve a larger resilience and flexibility and can be easily deployed. Compared, MIH requires an extensive MN integration.

#### 7.3.1.2 IEEE P1900.4

Similar to MIH, IEEE P1900.4 offers a link layer solution for communication adaptation, directed at distributed decisions for heterogeneous radio usage, as described in Section 4.1.4.

Compared to ASE, IEEE 1900.4 has the same characteristics as MIH compared to the ASE solution, including the lack of interaction with the mobility and resources control core network functionality. However, it can be integrated with the subscription profile for the decisions.

It also requires a high integration in the link layer of the MNs and has a limited resilience and flexibility due to its strict radio integration.

#### 7.3.1.3 Intra-3GPP Selection

The intra-3GPP access selection is controlled by network management entities, which are aware of the parameters of the MN and of the momentary network characteristics, as described in Section 4.1.5.1.

Compared to the intra-3GPP selection, ASE requires a low number of decisions and a highly reduced signaling, as the communication and the policy decisions do not have to be taken for each handover separately. Due to this, ASE has a lower duration for the reselection procedures.

ASE is realized at application level and thus it is flexible and more reliable than the network level solutions, which require the integration with the radio network management functions, such as the intra-3GPP reselection mechanism.

Due to the integration with the specific radio technologies, the intra-3GPP reselection provides a minimal mechanism for resource based access network selection. However, due to the large number of decisions to be taken, the intra-3GPP reselection can not be further extended towards SelfFit ASE functionality.

#### 7.3.1.4 ANDSF

The non-3GPP access network discovery and selection mechanism in the 3GPP architecture is based on the ANDSF network entity which makes decisions and transmits individual policies to the MNs, for appropriately selecting the next access network.

Compared to the ANDSF solution, SelfFit ASE is highly increasing the interaction with the other network entities, enabling the triggering of access network selection procedures based on the modifications of the required resources for each device through a reduced signaling procedure.

Being an application level solution, similar to ANDSF, ASE has the same reliability level. However, being tightly integrated with the subscription profile at network level, including the allowed access networks for the devices, and the current dynamic network properties, ASE sis able to make more complex decisions.

These complex decisions which consider the subscription information, the resource requirements and information on mobility and resource control are resulting in better access network selection decisions, by ensuring also that the resource requirements can be satisfied.

ASE decisions are taken individually for each subscriber, considering the modifications in the required resources as a trigger for access network reselection. Through the complete integration of the resource information, ASE is able to select the appropriate access network, without requiring multiple trials, as in the case of ANDSF decisions.

The less access network selection procedures come with an additional reduced signaling during the critical time of the reselection procedures, through the usage of the volatile-shallow contexts. Through these means, the duration of the adaptation is reduced at the cost of propagating information through the core network.

ASE requires the MN to make policy decisions and to select access networks, based on the network received policies, same as in the ANDSF solution. The MN functionality required is a user space application which can handle the application level communication with the network and orchestrates the device connectivity over the available network interfaces. Thus, no modifications are required in the operating system or network stack of the device making it easy to be deployed also on legacy devices.

#### 7.3.1.5 Conclusions

In conclusion, SelfFit ASE provides though a high integration with the network entities, especially with mobility and resources control, a more accurate decision mechanism and an extended support for the afferent network related procedures, make it more appropriate for dense network environments.

Additionally, by relying on an application level communication with the MN, it provides a high flexibility to the different deployments while not requiring deep integration with the specific MNs.

### 7.3.2 Core Network Path Adaptation Comparison

In this section, the SelfFit CPAE is compared with the other solutions from the literature having similar goals. The comparison is based on the evaluation from Section 5.1, the evaluation criteria from Chapter 3 and the metrics defined in Appendix A.2.1. The qualitative comparison is graphically illustrated in Figure 7.9, based on the data from Appendix A.2.1.

In a later Section, the solution is compared with the current Distributed Mobility Management solutions from IETF [226], developed in parallel with the SelfFit framework.

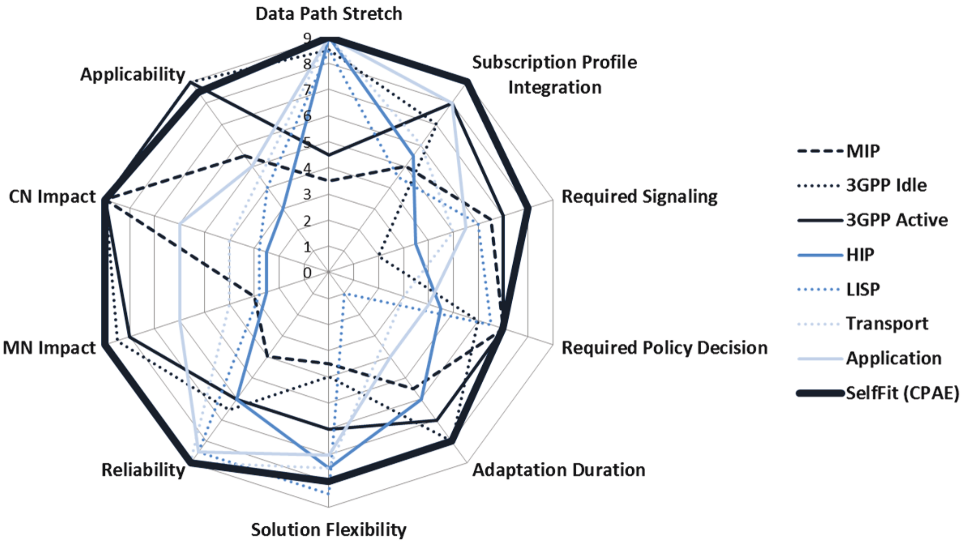


Figure 7.9: Core Network Path Adaptation Qualitative Comparison Representation

#### 7.3.2.1 3GPP Idle Mode Solution

When forwarding data traffic from CNs, the 3GPP Idle mode solution relies on paging over a specific tracking/routing area, as a mechanism to determine the current location of the MN which is in Idle Mode.

Compared to 3GPP Idle Mode, SelfFit CPAE has the advantage of addressing to a more generalized mobility, independent of the access technologies and of the underlying forwarding mechanism, thus addressing more than 3GPP accesses only.

Also, compared with the IP paging mechanism, CPAE is not duplicating data packets to an entire routing area in order to reach the MN. However, it either has to query the CPAE for the current location of the MN, which may be time consuming or it has to forward the data traffic to a default anchor node, which takes care of forwarding it to the appropriate destination.

It is assumed that data packets duplication will further extend the scalability problem, as in the case of IP paging, instead of reducing it by shortest data path stretch forwarding as in case of CPAE.

#### 7.3.2.2 Mobile IP and Variants

Mobile IP solutions are standardized for a long period of time. They enable transparent mobility support over multiple network providers through redirection of the data traffic to the home domain, specifically to a Home Agent entity, which is representing the MN location to the IP domain.

Compared to Mobile IP, SelfFit CPAE has the following advantages:

- CPAE supports multiple anchor nodes, which are dynamically selected for each MN. CPAE enables the data traffic to be forwarded on a shorter data path stretch, as data traffic does not have to be redirected to a fixed home node.
- The data path adaptation signaling is highly reduced, as it is executed only between CPAE and selected anchor nodes, thus it does not have to pass through the radio network or the public domain.
- The same CN transparency as MIP is maintained by CPAE, due to the FE functions. However, FEs have a reduced functionality compared to any MIP function, as the binding for the MNs are maintained only when there is downlink data traffic.

#### 7.3.2.3 3GPP Active Mode Mobility Support

For active mode mobility support, 3GPP standards include a set of solutions based on tunnels, such as GTP and PMIP. These solutions are logically similar from the perspective of the mobility support. They presume that a centralized anchor (i.e. PDN GW, GGSN) is selected for the MN and that redirection tunnels are created and maintained continuously between the current location of the MN and this centralized entity.

Compared with 3GPP Active Mode, CPAE has the following advantages:

- CPAE is able to dynamically select one of the multiple anchor nodes, adjusting the data path according to the momentary UE and network conditions.
- CPAE signaling is highly reduced, as the redirection tunnels are replaced by single forwarding binding messages. Therefore, no mobility association has to be established and especially deleted between the FEs and the anchor nodes.
- For the same data connection, multiple data paths are possible. First, multiple anchor nodes may be selected for the same data connection and data traffic can be dynamically forwarded to one of these entities. Second, multiple FEs can be used in parallel for the same MN. This was previously possible only with using multiple data connections, unable to transparently handover the data traffic.
- Through its anchor node reselection procedure, CPAE is very flexible, enabling the easy maintenance and energy efficient operation of the core network. The MNs are not bound to a single anchor node. CPAE can free or balance subscribers between different anchor nodes, if required. This was previously possible only between the entities with the same location.

#### 7.3.2.4 Host Identity Protocol(HIP) and Mobile IP with Route Optimization

In the MIP Route Optimization (MIP RO) solution, the MN signals to the CN its location change. The CN is the anchor of the data traffic which enables the usage of data paths, unmodified by the mobility management.

HIP is a a more evolved solution which separates completely from the data path entities and which takes into account multi-interface MNs. MIP with Route Optimization and HIP introduce a new layer between the network and the transport, which enable the CNs to match the MNs and to adapt the data path accordingly.

Compared to these solutions, CPAE has the following characteristics:

- CPAE maintains optimal data path stretch between the CNs and the MNs without CNs signaling. This was possible by the introduction of FEs at the border of the operator.
- The data path adaptation is transparent to the CN and to the operator of the CN. The FEs maintain the functionality of data path redirection on the behalf of the CNs.

- The adaptation signaling is contained to the operator where the MN resides. The delay of the operation is reduced to the signaling inside the single operator network.
- CPAE is transparent to the MN. The same functionality as in the current core network architecture is used i.e. based on GTP or PMIP. As the new layer in MIP RO and in HIP is the one identifying the MN to the CNs, it is not straight forward, if the functionality can be passed to the network.
- CPAE executes the data path adaptation as part of the attachment procedure similar to the 3GPP active mode solutions. Through this the data path adaptation delay is reduced compared to HIP or MIP RO, where CNs can be signaled only after the MN has attached to the network.

#### 7.3.2.5 Locator Identity Split Mobility Support (LISP)

The LISP Mobility Solution relies on the design of LISP protocol, which presumes that specific entities are located at the border of the operator domains, forwarding the data traffic received from other operators to the current locations of the MN. The LISP architecture was designed to solve the IP address fragmentation problem and not data path modifications.

The LISP mobility solution can be deployed in two modes: with functional entities which maintain the mobility control on behalf of the CN in the CN operator domain or in the MN operator domain. The solution in which the entities are located in the MN domain is similar to MIP and the same considerations apply. For the other solution, the considerations from the MIP RO and HIP apply.

#### 7.3.2.6 Transport Layer Identifier Solutions

The transport layer identifier solutions are using the transport layer of the CN to match the identity of the MN with its new location. These solutions must involve the MNs and the CNs into the procedure, requiring the modification of the network protocol stack. Because of this reason, they are not further compared with CPAE.

#### 7.3.2.7 Application Layer Identifier Solutions

These solutions use for identifying the MN a unique identifier, specific for a set of applications, such as a SIP URI in IMS and SIP services or an Apple or a Google Account. When the location of the MN in the network changes, the new location is registered, replacing the old one. For service continuity, a mechanism to announce the CNs to redirect the data traffic to the new location is required. Such solutions are already considered as part of the application level protocol, such as the SIP re-INVITE transactions. As these solutions are an application specific overlay over the network layer, they have to be signaled end-to-end.

Compared to these solutions, the SelfFit adaptation has the following advantages:

- SelfFit is addressing all types of data traffic, independent of application level protocols.
- SelfFit maintains the data path stretch between the CNs and the MNs without redirection due to the mobility support. This was possible by the introduction of FEs on the border of the operator domain.
- SelfFit is transparent to MNs and CNs. The transparency to MNs is maintained through the state of the art core network functionality, while to CNs is realized through FEs.
- The signaling is contained to the operator where the MN resides. The delay of the operation is reduced to the signaling inside the single operator network. In application level solutions the signaling has to be executed end-to-end.
- The signaling is done only with a limited number of FEs which had or have downlink data traffic for the MN. This is independent of the number of CNs.
- Application level procedures have to be executed after the attachment procedures. First an IP address has to be allocated and then the data path adaptation can be signaled. With CPAE, these operations are executed in a single step in the SelfFit solution.

#### 7.3.2.8 Conclusions

The SelfFit CPAE includes characteristics from both the network contained solutions such as MIP or 3GPP and from the end-to-end solutions such as HIP, SIP or mSCTP or LISP Mobility.

CPAE maintains from the network level solutions the transparency and the signaling reduction to an operator domain, which implies a reduced data path adaptation delay.

Also, it does not require redirections due to mobility control, the same as in the end-to-end solutions. CPAE signaling is contained to a single operator domain, bu introducing FEs which take the role of CN anchoring.

As an addition to the current solutions, CPAE can transmit mobility related events to the service level platforms, reducing the number of mobility support layers.

This is done through a re-organization of the functionality in the core network of the operator, through usage of a single mobility managed layer formed from CPAE and the anchor nodes and through dispersing the minimal information towards the FE located on the border of the operator domain.

### 7.3.3 Core Path Adaptation comparison with Distributed Mobility Management Solutions

In parallel with the development of the SelfFit core path adaptation solution, a set of solutions were considered by the IETF work group on Distributed Mobility Management (DMM) [226]. The DMM solutions also target the usage of multiple distributed anchor nodes and the adaptation of the data path through these anchors. In this subsection, the SelfFit CPAE is evaluated against the current DMM solutions. The comparison is based on the evaluation criteria from Chapter 3 and the metrics defined in Appendix A.2.1. The qualitative comparison is graphically illustrated in Figure 7.10 based on the data from Appendix A.2.1.

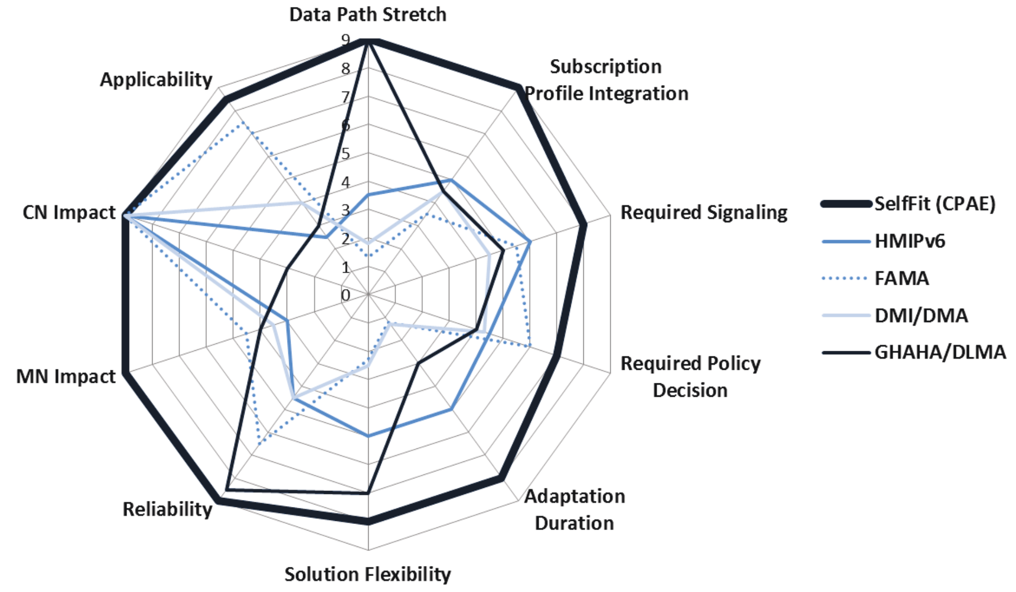


Figure 7.10: Core Network Path Adaptation vs. Distributed Mobility Management - Qualitative Comparison Representation

#### 7.3.3.1 Hierarchical Mobile IP(HMIP)

HMIP is a variant of the Mobile IP, especially addressing IPv6, having as goal to make local mobility transparent to CNs [227]. HMIP introduces a Mobility Anchor Point (MAP) which handles the micro-mobility of the MNs in a local domain. The adaptation procedures are reduced when the MN roams in a single domain by signaling the adaptation only with the MAP. For the mobility between domains, HMIP defaults to MIP functionality.

Compared to this solution, CPAE has the following advantages:

- CPAE does not depend on a static home anchor point. This enables the data traffic to be forwarded on a shorter path stretch compared to HMIP and its underlying MIP solution.
- The same transparency from the CNs is maintained in HMIP as in MIP and in SelfFit CPAE.
- The HMIP solution is more flexible compared to the MIP solution, as the MAP can be reselected whenever required. However, full flexibility, as in the case of the SelfFit CPAE, can not be attained due to maintenance of a unique home agent.

#### 7.3.3.2 Flat Access and Mobility Architecture (FAMA)

FAMA considers the placement of mobility anchor nodes close to the MNs, named here Distributed Anchor Routers (DARs) [228]. The DARs are able to give topologically correct IP addresses to the MNs during the attachment procedures. Additionally, any DAR acts as MIP Home Agents for IP addresses allocated, while the device was attached in other network areas.

Compared to FAMA, SelfFit CPAE has the following advantages:

- SelfFit CPAE maintains the minimal data path stretch for all the data traffic of the MN, while in FAMA, for the data traffic which uses the old IP address has to be redirected to DAR. FAMA has a longer data path than Mobile IP as the data traffic is transmitted to the initial access network of the MN.
- SelfFit CPAE maintains the signaling to the local domain of the MN, thus reducing the duration of the data path adaptation procedures.
- FAMA is highly inflexible, as it requires at least two DAR entities to maintain bindings for the same MN - the source DAR for the first IP address and the target DAR for the second one. The number of DAR entities maintaining bindings for the MN increases with its mobility, being directly proportionate

to the number of handovers. In case of SelfFit, this issue is mitigated by keeping a single binding in the CPAE which is transmitted on demand to FEs. No binding in the past anchor nodes is required.

#### 7.3.3.3 Dynamic Mobile IP (DMI)

DMI is a variant of the MIP which considers the allocation of a home agent and of a visited home address in the visited domains which enables the establishment of new communication sessions which do not require transparent service continuity [229]. This solution relies on MIP for all the other sessions requiring transparent communication with the CNs.

Compared to DMI, CPAE has the following major advantages:

- The data path stretch in CPAE is adapted for all the data traffic, especially addressing the data traffic requiring service continuity.
- The signaling in CPAE is reduced to the local domain. For DMI, a double signaling to the home and to the visited Home Agents is required each time the MN changes its network location.
- The DMI adaptation duration is equivalent to the one for Mobile IP, which, due to the signaling to the home domain of a previously allocated visited domain IP address, is higher than the one considered for CPAE.

A combination between FAMA and DMI based on PMIP, named Distributed Mobility Anchor (DMA) was proposed. DMA considers that the selection of the data path is done by the MN which has to support two distinct prefixes [230]. The DMI and the DMA have similar properties in report with the SelfFit proposal with the exception of the lower integration required in the MN.

#### 7.3.3.4 Global HA to HA (GHAHA)

GHAHA solution considers a distributed set of Home Agents throughout the IP domain each having a synchronized information on all the MNs. In order to reach the appropriate MN, a CN transmits an anycast message which will reach the closest HA which in its turn being aware of the current binding information of the MN redirects the data packets appropriately [231].

Compared to GHAHA, SelfFit CPAE has the following advantages:

- The SelfFit CPAE does not synchronize the information between multiple Home Agents. This is the largest disadvantage of the GHAHA solution as it requires a high signaling over the IP domain.

- GHAHA solution is not transparent to CNs. It requires the usage of anycast messages, which, apart from the security issues, requires modifications in the correspondent entities. SelfFit CPAE is transparent to CNs. In order to ensure that the FEs are reached, they are placed by deployment design at the entry in the operator domain.

- Due to the requirement to place home agents at a global scale and to support anycast messages the applicability of the GHAHA on top of the current deployments is highly reduced. Comparatively, SelfFit keeps its functionality limited to a single operator domain.

In order to reduce the signaling used by GHAHA, two additional solutions were considered. First, a new protocol named SAIL proposes the usage of the same GHAHA architecture in which the MN location information is stored in Distributed Hash Tables (DHT) [232]. This reduces the number of nodes which have to be synchronized, however still at a higher level than the SelfFit CPAE. The rest of the comparison remains the same.

Another approach named Dynamic Local Mobility Anchors (DLMA)is considering a split of the Home Agent functionality in GHAHA in functionality addressing the MN and the functionality addressing the forwarding from CNs [233], similar to the FEs separation from the the anchor nodes in the SelfFit CPAE. As SelfFit, DLMA can be deployed in a single operator domain. However, the signaling considered is similar to the Mobile IP signaling resulting in a high number of messages to be exchanged between the data path redirection entities and the anchor nodes.

Additionally, DLMA does not rely on the underlying IP forwarding, requiring IPv6 anycast messages to be sent by the CNs, thus requiring modifications in the CNs which highly deters its deployment applicability.

#### 7.3.3.5 Conclusions

In conclusion, SelfFit CPAE provides a more appropriate solution for carrier grade operator deployments in realizing a distributed mobility management solution. This is mainly due to the employment of FEs, at the border of the operator domain, enabling redirection while not interfering with the CN domains. As the forwarding through the Internet backbone is based on criteria, such as link availability and momentary peering costs, and considers fixed the IP address location, a redirection from the CN's domain does not optimize further the end-to-end data path.

Additionally, by keeping the communication transparent to the CNs and by reducing the data path redirection due to mobility to minimum, the SelfFit CPAE can be easier accepted as a suitable solution for evolving the current operator deployments.

### 7.3.4 Application Adaptation Solutions Comparison

In this section, the SelfFit AAE is compared with the other solutions from the literature, having similar goals. The comparison is based on the evaluation from Section 5.1, the evaluation criteria from Chapter 3 and the metrics defined in Appendix A.2.1. The qualitative comparison is graphically illustrated in Figure 7.11 based on the data from Appendix A.2.1.

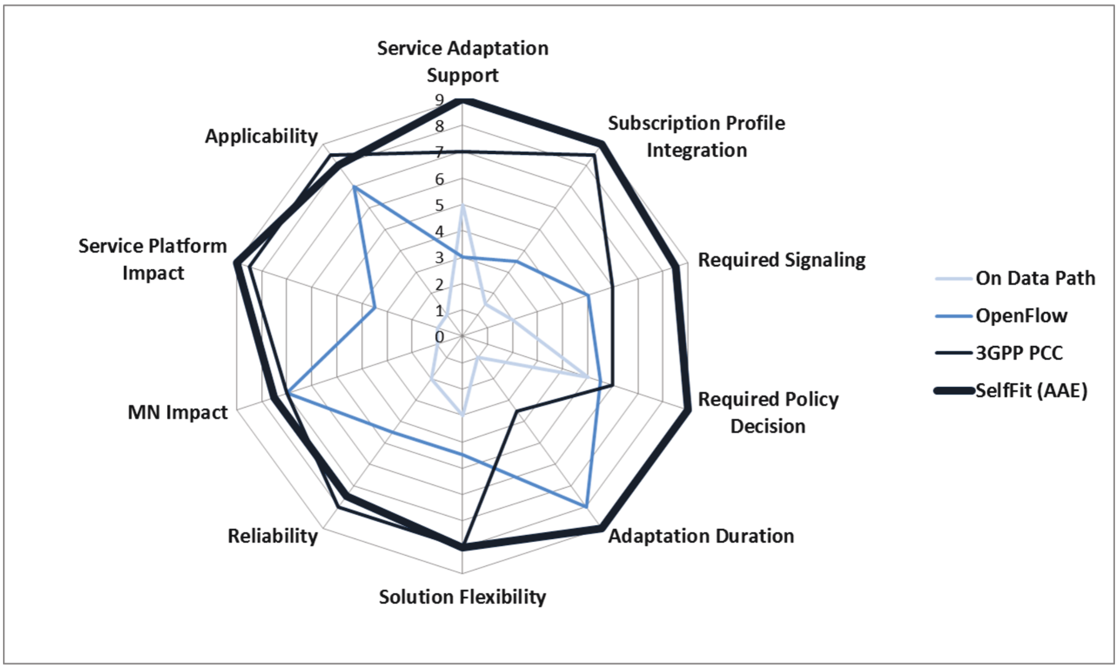


Figure 7.11: Application Adaptation Qualitative Comparison Representation

#### 7.3.4.1 Best Effort Delivery

The Best Effort application delivery is the main paradigm of the current Internet communications. However, as it does not offer any support in application delivery including application adaptation, the specific paradigm is not offering the expected features. Thus, the best effort delivery is not further analyzed here.

#### 7.3.4.2 On Data Path Signaling

The resource reservations based on signaling executed over the data path is specific to circuit switched technologies and it was widely studied in the IP domain in the form of integrated services resulting in the RSVP and GIST systems.

Compared to the on-data-path signaling solutions, the SelfFit AAE has the following advantages:

- As AAE operations are under the control of a carrier grade operator which is aware of the available resources in the network, the adaptation is executed in a single step without requiring exception procedures, such as failure of reservations.
- As the AAE is communicating with the service platform, it makes appropriate resource reservations without requiring additional signaling between the MN and service platforms.
- The AAE is able to make decisions and to enforce them in the operator which deploys the specific enabler, thus resulting in a limited signaling and in a reduced adaptation duration.

#### 7.3.4.3 OpenFlow

The OpenFlow solution is offering a policy control mechanism for the network entities from a centralized controller enabling specific resource reservation rules to be installed on the data path for the specific MNs. However, at the moment of writing of this dissertation, there was no interface available towards the service platforms or integration with subscription profiles, limiting the solution to the lower data path level.

Compared to OpenFlow solution, AAE has the following advantages:

- AAE is integrated with the network subscription information and with the mobility support in the core network. It enables the adaptation to be executed in a single step, which is not currently possible without a new negotiation of the required resources between the MN and the service platforms.
- In OpenFlow, the adaptation can be executed only based on pre-allocated rules. These do not consider the dynamic context of the MN or the flexible application delivery. Thus, the resource adaptation is missing the required flexibility, as attained through the deployment of AAE.
- Due to the requirement of the communication between the MN and the service platform for the data delivery, the procedure has an extended delay.

#### 7.3.4.4 3GPP Policy and Charging Control (PCC)

The 3GPP PCC architecture is an extension of the basic policy control system towards applications and networks enabling the resource reservations on the data path, depending on the subscription profile of MNs and on the service platforms requirements.

Compared to the 3GPP PCC, SelfFit AAE has the following advantages:

- AAE is able to execute the resource reservation adaptation and the resources handover during an access network reselection in a single step while the 3GPP PCC has to consider two policy decisions and the afferent signaling.
- In AAE, the resources for the applications are reserved without any signaling to the MNs or the service platforms. This reduces the required communication during the adaptation, compared to the 3GPP EPC and, thus, reducing the delay of the overall procedure.
- As AAE relies on the 3GPP PCC solution, the same integration with the subscription profile, high reliability and flexibility are considered.
- As 3GPP PCC and SelfFit AAE are reserving resources for various applications, they impact in the same manner the service platforms.
- The 3GPP PCC is easier to be integrated into the currently deployed systems, as it requires only the resource reservation procedure and not the provisioning procedure. In order for AAE to function, the provisioning procedure has to be executed. In case this is not possible in specific deployments, default delivery levels have to be introduced in AAE which reduce the flexibility.

#### 7.3.4.5 Conclusions

The SelfFit AAE bases itself of the 3GPP PCC mechanisms for resource reservation and builds a specific functionality which enables the adaptation of the resources reserved for the different applications during the access network reselection procedures.

Due to this reduced reservation procedure, AAE requires less policy decisions, reduced signaling and is able to adapt the resources faster, compared to the available solutions.

Due to its integration with current standard solutions, SelfFit CPAE can be easily integrated in the future deployments.

## 7.4 Conclusions

In this chapter, the SelfFit framework was validated through proof of concept implementation and through expert comparison with the other available solutions from the literature. The testbed implementation led to a minimal evaluation in laboratory conditions of the proposed mechanisms using a software realization of an operator core network.

Through the evaluation of the implementation, it is considered that the proposed concepts and their practical mapping to the 3GPP architecture and afferent testbed realization are feasible to be deployed separately or as a complete framework bringing visible optimization to the currently working mobile data communication system.

The expert evaluation provided a critical assessment of the solution against the different available technologies. The evaluation was done separately for each of the three SelfFit enablers. The SelfFit enablers are considered better in a great majority of the cases compared to all the other assessed technologies.

It is assumed that the advantages offered by SelfFit on each of the three separate perspectives are further enhanced by the deployment of the complete SelfFit solution in the same system.

## Chapter 8

# Conclusions and Further Work

This dissertation has extensively studied the communication support in carrier grade operator networks, discussed the challenges of connectivity adaptation on a large scale wireless system, evaluated different approaches, and proposed a novel subscriber oriented adaptation framework from the perspective of access network selection, the efficient usage of core network resources, and data path resource adaptation. The adaptation framework has as main objective the enhancement of the scalability of the overall system and, through this, the improvement of the communication of the mobile users.

In the course of the research for this dissertation, as an outcome result, the following related publications have been disseminated:

1. Nine patent submissions
2. Six industry oriented conferences presentations without publication
3. Six tutorials in the area of packet switched core network architectures
4. Twenty-three conference papers and book chapters from which seventeen as first author.

This research work contributed to several industry projects, such as Huawei-Fraunhofer ConVeHO and Huawei-Fraunhofer FlatMore projects.

A large amount of the research in the form of practical realization was included in the Fraunhofer FOKUS OpenEPC toolkit currently deployed in the laboratories of more than twenty research companies from academia, equipment manufacturers and carrier grade operators.

Additionally, the OpenEPC toolkit is used as the basis for large scale projects such as the FI-PPP FI-WARE, EU FP7 Univerself, and EU FP7 MobileCloud-Networking.

This chapter summarizes achievements of this work and describes on-going future extension directions in some of the research projects.

## 8.1 Summary and Impact

This dissertation proposed a self-adaptation framework, named SelfFit, which introduces a subscriber oriented management functionality integrated with the core network infrastructure, addressing carrier-grade mobile network operator networks. This work can be considered as the foundation for the scalability increase of the operator infrastructures through subscriber oriented functionality adaptation.

With the purpose of improving the operator core network scalability, it was important to understand the mobile telecommunication trends and the architecture foreseen to be deployed. The work here is motivated by evolution of the wireless ecosystem towards massive broadband communication, which presumes a high number of heterogeneous devices communicating anywhere and anytime by using heterogeneous wireless technologies. The increasing trend in subscriber devices and network capacity brings the opportunity of novel services based on new paradigms such as cloud services, M2M and smart cities.

However, through the analysis in Chapter 2 of the core network architectures foreseen to be deployed for supporting the massive future mobile communication such as the 3GPP EPC, it was observed that all the subscribers and their communication are handled in the same manner, resulting in high overhead of signaling and throughput.

Chapter 3 detailed further these limitations into a set of requirements from the perspective of the subscriber, service platform and core network. From the requirements' description, a set of evaluation metrics were developed in order to be able to qualitatively compare the deployment opportunity in carrier grade operator networks. Through this methodology, the various heterogeneous technologies were further assessed.

Following, the different solutions available in the literature at the moment of writing of this dissertation were analyzed. As the number of technologies considered was high, they were grouped in classes based on their specific functionality target. Based on the observed results, this work identified the following main topics that would required further investigation, namely:

- Subscriber oriented access network selection
- Transparent core network path adaptation
- Automatic application resource reservation adaptation to network conditions

To tackle these issues, three complimentary directions were pursued, resulting in three distinct enablers from which the SelfFit framework consists of, as considered in Chapter 5. The interconnections between the three enablers, as well as between the enablers and the core network, were further considered.

The SelfFit ASE was designed to support access network selection, based on the required resources by the subscribers along with the subscriber information integration. This feature enables a harmonized usage of the available resources over the heterogeneous access networks at specific locations. Additionally, the SelfFit ASE was enhanced with the shallow context concept, enabling the reduction of the time critical period for access network handovers, without requiring pre-reservation of resources.

The SelfFit CPAE proposed two mechanisms for core path adaptation: synchronous and reactive to downlink data traffic. These mechanisms enable the redirection of the data traffic to the subscriber's momentary network location, resulting in a shorter data path stretch. Additionally, CPAE was extended with a concept for flexible data path reselection, enabling the adaptation of the data path through the specific operator support entities.

SelfFit AAE was designed to include an automatic mechanism for modifying the resources reserved for specific applications, depending on the access network currently used. This concept enables the efficient adaptation of the applications to the current access network used, through a reduced number of policy decisions and of the signaling over the wireless network.

For proving the applicability of the proposed SelfFit framework, the three enablers were mapped to the 3GPP EPC functionality in Chapter 6,as complimentary functions to the existing architecture and illustrated in Figure 8.1. This step of the dissertation proved that the given concepts could easily fit in the existing functionality and thus, be integrated as stand-alone extensions of the standardized carrier-grade network core network and to be deployed in the real-life environments. As a novel subscriber oriented management layer was proposed, it can be easily integrated with other not yet available, future network architectures, surpassing the 3GPP EPC.

Based on the functionality proposed and on the mapping to the 3GPP EPC, several specific items which were considered of high commercial importance by the industry project partners were protected through patent submissions.

Following, the generic enablers were implemented in the form of multiple prototypes and testbeds on top of the Fraunhofer FOKUS OpenEPC toolkit proving the practical applicability of the proposed features. As presented in Chapter 7, through comparative measurements on the practical implementation, it was observed that for each of the specific developed features, the SelfFit framework performs better in most of the use cases.

Furthermore, through qualitative assessment against the currently available solutions from the literature, the SelfFit is considered better on all the evaluation criteria. In designing the SelfFit enablers, a high emphasis was put on the reduction of the number of decisions and afferent signaling that has to be inte-

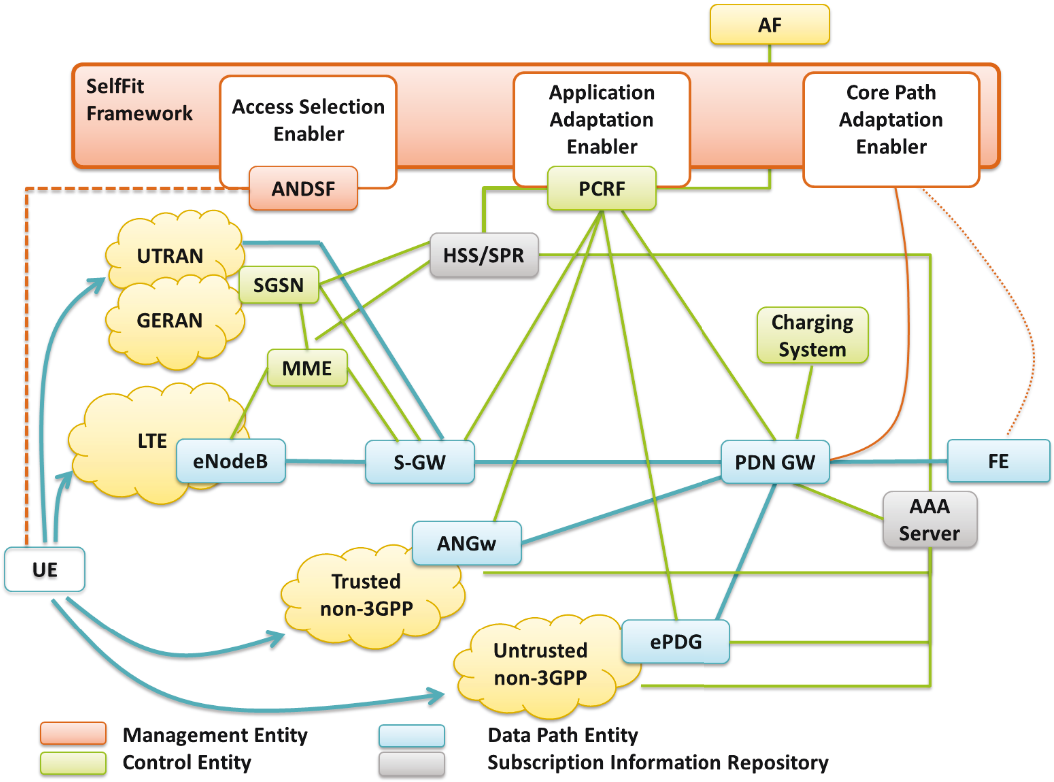


Figure 8.1: Mapping of SelfFit framework to 3GPP architecture

grated into the network system, especially on the communication with the mobile devices, as these characteristics were considered as the main deterring factors of scalability. A careful attention was given to the transparency of the operations to the correspondent nodes, which was achieved for all the functionality proposed. Additionally, except for the access network selection which has by its nature to involve the mobile device, the rest of the operations do not consider necessary any communication over the wireless links during the adaptation phases, characteristic which highly reduces the duration of the procedures.

The most outstanding results, including the novel concepts and their mapping on top of the 3GPP EPC, were published as part of academic conference proceedings, as presentations in industry conferences and as further development perspectives for operator functionality related tutorials. Through these dissemination means, the results of this dissertation were brought back to the scientific community for increasing awareness and for further validation. The list of publications appears throughout the references used in this dissertation and they are further listed in the Bibliography.

## 8.2 Outlook

Regarding further work, there are two main directions that the SelfFit framework could follow. The first consists on focusing on the existing functionality and its further dissemination. The second goes further into the automatic adaptation of the subscriber communication through the application of the SelfFit concept and their practical extension in prototype form addressing new research directions taken for improving the carrier grade operator core networks.

### 8.2.1 The SelfFit Dissemination

From the perspective of the features included, the SelfFit framework can be used as it is already implemented for further dissemination. Three distinct tracks are foreseen for the dissemination of non-modified SelfFit features, as illustrated in Figure 8.2:

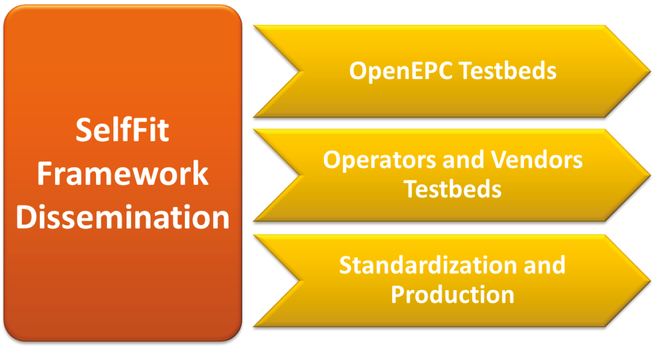


Figure 8.2: SelfFit Dissemination Directions

- OpenEPC testbeds - part of the SelfFit features were included in OpenEPC toolkit. Through these means, they are available to R&D departments for installing, using and extended as part of their own testbeds.
- Operator and Vendor testbeds - the features included in OpenEPC are available to operators and equipment manufacturers and can be customized according to their specific use cases, increasing the awareness of the importance of further deployment of the SelfFit framework.
- Standardization and Production - relying on the industry project partners, the most outstanding features of the SelfFit framework are to be further standardized and included as part of their products.

### 8.2.2 The SelfFit Extensions

The SelfFit concepts, design and implementation will serve in the next years as the foundation for several research projects:

- Design and reference implementation of the interface to networks for Future Internet as part of the FI-PPP FI-WARE project.
- Realizing a proof-of concept for SON orchestration as part of the EU FP7 Univerself project.
- Design, specification and implementation of a flexible core network using cloud computing concepts for the core network as part of the EU FP7 MobileCloud project.

The following subsections describe each project briefly.

#### 8.2.2.1 FI-PPP FI-WARE

The goal of the FI-WARE project is to introduce an innovative infrastructure enabling a cost-effective creation and delivery of services. For this, FI-WARE will provide a set of open networks and a reference implementation realizing the foundation for the Future Internet. It includes enablers in the area of networks, security, applications, Internet of Things, Cloud Computing and large data processing [74].

In this context, the SelfFit framework as part of OpenEPC toolkit was selected as the most representative prototype available for realizing an open interface to the networks, harmonizing the connectivity needs of the services build on top.

Specifically for this purpose, a set of APIs including the Access Network Selection and the Application Adaptation from SelfFit along with the subscription information, resource and control charging interfaces will be extended and adapted as required by the Future Internet services and applications. As depicted in Figure 8.3, this enables the applications to indicate their delivery preferences as access network used and as reserved resources and thus triggering the SelfFit functionality.

#### 8.2.2.2 UniverSelf

EU FP7 UniverSelf project aims at overcoming the growing management complexity in the future network systems due to the introduction of more flexible functionality such as SON concepts. For this it uses the autonomic communication principles and technologies in order to realize a smooth and trustworthy environment in the carrier grade operator networks. The main functionality is

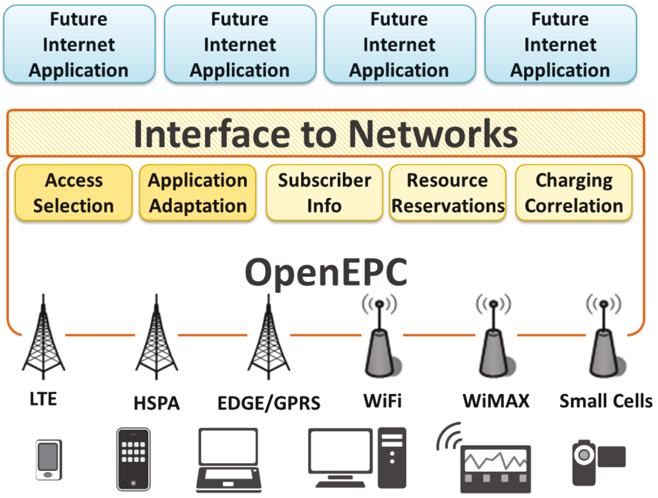


Figure 8.3: FI-WARE Interface to Networks

realized in the form of a Unified Management Framework (UMF) which enables the governance of multiple management systems [222].

The SelfFit Access Selection Enabler will be extended towards developing a first prototype of the collaboration means between different SON control loops, according to operator policies in which the different access selection policies are adapted depending on the momentary network context and on the utility of the various autonomic management features such as load balancing, energy efficiency and handover control in order to optimize various Key Parameter Indicators (KPIs) like the average drop rate.

#### 8.2.2.3 Mobile Cloud Networking

The Mobile Cloud Networking project aims at developing a novel mobile network architecture and the afferent technologies, leading the way towards fully cloud-based mobile communication systems and to extend the cloud computing paradigm in the direction of supporting elastic provisioning of mobile services [223].

From the operator network perspective, Mobile Cloud Networking has as main goal to offer cost-effective core network functionality, through usage of a common infrastructure and software implementation of the specific network components and through the sharing of the communication environment, enabling on-demand provisioning.

Specifically, it presumes the deployment of highly elastic, software based 3GPP EPC functionality on top of a heterogeneous programmable infrastructure offered by one or more infrastructure providers as illustrated in Figure 8.4. This requires further research in the area of flexibility of the operator core network functional-

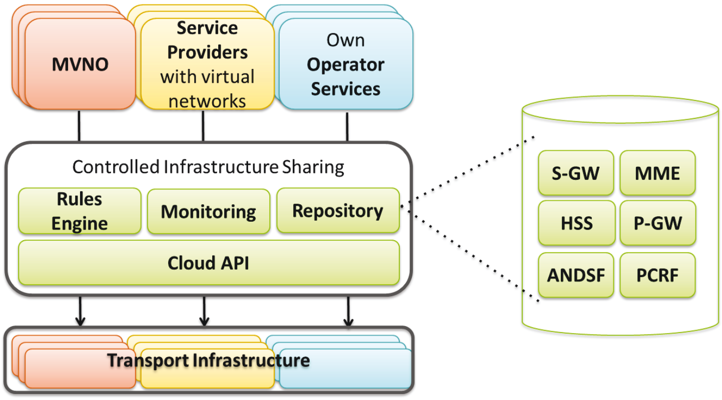


Figure 8.4: Mobile Cloud Networking Infrastructure Sharing

ity, as well as in the area of extending the cloud computing into supporting this flexibility.

The SelfFit framework concepts will be used in Mobile Cloud Networking, along with the OpenEPC implementation, especially in the area of realizing this flexibility. The adaptation paradigm will be translated for enabling the provisioning of virtual components according to the momentary infrastructure needs.

Additionally, SelfFit Core Path Adaptation enabler will be extended and adapted for supporting the transparent on-demand contextualization of the distinct data path entities. For this, adding to the functionality described in the SelfFit implementation, a novel functionality for context transfer, transparent to the operator core network functionality will be provided.

# List of Acronyms

| | |
|---|---|
| 3GPP | 3rd Generation Partnership Project |
| 3GPP2 | 3rd Generation Partnership Project 2 |
| AAA | Authenticated, Authorization and Accounting |
| AAE | Application Adaptation Enabler |
| AF | Application Function |
| AMBR | Aggregated Maximum Bit Rate |
| AN | Access Network |
| ANDSF | Access Network Discovery and Selection Function |
| ANGw | Access Network Gateway |
| AP | Access Point |
| APN | Access Point Name |
| ASE | Access Selection Enabler |
| BBERF | Bearer Binding and Event Rules Function |
| CDMA | Code Division Multiple Access |
| CN | Correspondent Node |
| CoA | Care of Address |
| COPS | Common Open Policy System |
| CPAE | Core Path Adaptation Enabler |
| CS | Circuit Switched |
| CSCF | Call Session Control Function |
| DAR | Distributed Anchor Router |
| DHT | Distributed Hash Table |
| Diff-Serv | Differentiated Services |
| DM | Device Management |
| DMI | Dynamic Mobile IP |
| DMM | Distributed Mobility Management |
| DNS | Domain Name System |
| DSL | Digital Subscriber Line |
| DSMIP | Dual Stack Mobile IP |
| EDGE | Enhanced Data rates for GSM Evolution |
| EID | Endpoint ID |
| EPC | Evolved Packet Core |
| ePDG | evolved Packet Data Gateway |
| ETR | Egress Tunnelling Router |

| | |
|---|---|
| FAMA | Flat Access and Mobility Architecture |
| GBR | Guaranteed Bit Rate |
| GHAHA | Global Home Agent to Home Agent |
| GIST | General Internet Signalling Transport |
| GPRS | General Packet Radio Service |
| GSM | Groupe Special Mobile |
| GTP | GPRS Tunnelling Protocol |
| HA | Home Agent |
| HDF | Handover Decision Function |
| HI | Host Identifier |
| HIP | Host Identity Protocol |
| HIT | Host Identifier Tag |
| HMIP | Hierarchical Mobile IP |
| HoA | Home Address |
| HSPA | High Speed Packet Access |
| HSS | Home Subscriber Server |
| I-CSCF | Interrogation Call Session Control Function |
| IETF | Internet Engineering Task Force |
| IMS | IP Multimedia Subsystem |
| IMSI | International Mobile Subscriber Identifier |
| Int-Serv | Integrated Services |
| IoT | Internet of Things |
| IP | Internet Protocol |
| IPv4 | Internet Protocol version 4 |
| IPv6 | Internet Protocol version 6 |
| ITR | Ingress Tunnelling Router |
| ITU | International Telecommunication Unit |
| ITU-T | ITU Telecommunication Standardization Sector |
| KPI | Key Parameter Indicator |
| LISP | Locator Identity Split Protocol |
| LMA | Local Mobility Anchor |
| LTE | Long Term Evolution |
| M2M | Machine to Machine (Communication) |
| MAG | Mobility Access Gateway |
| MAP | Mobility Anchor Point |
| MBR | Maximum Bit Rate |
| MICS | Media Independent Control Server |
| MIES | Media Independent Event Server |
| MIH | Media Independent Handover |
| MIHF | Media Independent Function |

| | |
|---|---|
| MIIS | Media Independent Information Server |
| MIP | Mobile IP |
| MIPv4 | Mobile IPv4 |
| MIPv6 | Mobile IPv6 |
| MLMF | Mobility Location Management Function |
| MN | Mobile Node |
| MTU | Maximum Transmission Unit |
| NGMN | Next Generation Mobile Network |
| NSIS | Next Steps in Signalling |
| MME | Mobility Management Entity |
| mSCTP | mobile SCTP |
| NGN | Next Generation Networks |
| OCS | On-line Charging System |
| OFCS | Of line Charging System |
| OMA | Open Mobile Alliance |
| OMA DM | OMA Device Management |
| ONF | Open Networking Foundation |
| PCC | Policy and Charging Control |
| PCEF | Policy and Charging Enforcement Function |
| PCRF | Policy and Charging Rules Function |
| P-CSCF | Proxy Call Session Control Function |
| PDN GW | Packet Data Network Gateway |
| PDP | Policy Decision Point |
| PEP | Policy Enforcement Point |
| PETR | Proxy ETR |
| PON | Passive Optic Fibre |
| PMIP | Proxy Mobile IP |
| PMR | Professional Mobile Radio |
| PS | Packet Switched |
| QoE | Quality of Experience |
| QoS | Quality of Service |
| RACF | Resource Admission and Control Function |
| RACS | Resource Admission and Control System |
| RLoc | Router Locator |
| RO | Routing Optimization |
| RSVP | Resource Reservation Protocol |
| RTT | Round Trip Time |
| RVS | Rendezvous Server |
| SCTP | Stream Control Transmission Protocol |
| S-CSCF | Serving Call Session Control Function |

| | |
|---|---|
| S-GW | Serving Gateway |
| SGSN | Serving GPRS Support Node |
| SIM | Subscriber Identity Module |
| SIP | Session Initiation Protocol |
| SLA | Service Level Agreement |
| SON | Self-Organizing Networks |
| SPR | Subscription Profile Repository |
| TCP | Transport Control Protocol |
| TFT | Traffic Flow Template |
| TTT | Time To Trigger |
| UDP | User Data Protocol |
| UE | User Endpoint |
| UMTS | Universal Mobile Telecommunications System |
| URI | Uniform Resource Identifier |
| VPN | Virtual Private Network |
| WiFi | Wireless Fidelity |
| WiMAX | Worldwide Interoperability for Microwave Access |

# Bibliography

[1] Rainie Lee Janna Anderson, "Internet evolution: Where hyperconnectivity and ambient intimacy take us", Pew Research Center: Internet & American Life Project, July 2010. (Cited on page 1.)

[2] Theodore Rappaport, *Wireless Communications : Principles and Practice*, Pearson Education, 2010. (Cited on page 1.)

[3] Cisco, "Cisco Visual Networking Index: Forecast and Methodology 2010-2015", Cisco Visual Networking Index, June 2011. (Cited on pages 1, 5 and 23.)

[4] GSA, "GSM/3G and LTE Market Update", Global Mobile Suppliers Association, March 2011. (Cited on pages 1, 25 and 38.)

[5] Gregor Maier, Fabian Schneider, and Anja Feldmann, "A First Look at Mobile Hand-Held Device Traffic", vol. 6032, pp. 161–170, 2010, 10.1007/978-3-642-12334-4_17. (Cited on page 1.)

[6] Mary et all Meeker, "The Mobile Internet Report", Morgan Stanley Research, December 2009. (Cited on pages 1, 5, 38 and 39.)

[7] Adam et all Holt, "Cloud Computing Takes Off: Market Set to Boom as Migration Accelerates", Morgan Stanley Research, May 2011. (Cited on page 1.)

[8] GSMA, "Connected Intelligence Database: 2020 connected devices overview", Machina Research Overview, October 2011. (Cited on pages 2, 5 and 40.)

[9] Rudolf Giffinger, Christian Fertner, Hans Kramar, Robert Kalasek, Natasa Pichler-Milanovic, and Evert Meijers, "Smart cities - Ranking of European medium-sized cities", EU Smart Cities Initative, 2009. (Cited on page 2.)

[10] European Comission, "EU Smart Cities and Communities Initiative", An initative of the European Comission, 2011. (Cited on pages 2 and 5.)

[11] ITU-T, "Y.2001: General overview of NGN", December 2004. (Cited on pages 2, 3, 10, 20 and 42.)

[12] Ray Horak, *Webster's New World Telecom Dictionary*, Number 978-0-471-77457-0. Wiley Publishing Inc., 2008. (Cited on pages 2, 6, 7 and 22.)

[13] Stueber Gordon, *Principles of Mobile Communication*, Number 978-1461403630. Springer, Berlin; Auflage: 3rd Edition., September 2011. (Cited on page 2.)

[14] Andre B. Bondi, “Characteristics of scalability and their impact on performance”, *Proceedings of the 2nd international workshop on Software and performance*, vol. 1, September 2000. (Cited on page 3.)

[15] ITU-T, “Y.2011: General Principles and General Reference Model for Next Generation Networks”, October 2004. (Cited on pages 3, 4, 11 and 90.)

[16] “Next Generation Mobile Networks Beyond HSPA & EVDO”, Online, December 2006. (Cited on pages 3, 5, 6, 26 and 41.)

[17] “Heterogeneous Networks - White Paper Summary”, Online, June 2011. (Cited on page 4.)

[18] 3rd Generation Partnership Project (3GPP), ”. (Cited on pages 4, 23 and 25.)

[19] Thierry Lucidarme Pierre Lescuyer, *Evolved Packet System (EPS): The LTE and SAE Evolution of 3G UMTS*, Number 978-0-470-05976-0. John Wiley and Sons, September 2008. (Cited on pages 4, 25 and 65.)

[20] Miguel-Angel Garcia-Martin Gonzalo Camarillo, *The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds*, Number 978-0-470-51662-1. John Wiley and Sons, 3rd Edition edition, 2008. (Cited on page 4.)

[21] 3GPP, “TS 23.402 - Architecture enhancements for non-3GPP accesses”. (Cited on pages 6, 9, 21, 26, 27, 28 and 58.)

[22] D. Liu C. Perkins, “DMM Comparison Matrix”, October 2011. (Cited on page 6.)

[23] T. Schierl K. Grueneberg L. Celetto C. Timmerer T. Zahariadis, C Lamy-Bergot, *Content Adaptation Issues in the Future Internet*, chapter Content Adaptation Issues in the Future Internet, pp. 283–292, Content Adaptation Issues in the Future Internet, 2009. (Cited on page 6.)

[24] 3GPP, “TS 23.401 - General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access”. (Cited on pages 9, 21, 26, 27, 28, 30, 57, 65 and 66.)

[25] J. Manner and M. Kojo, “Mobility Related Terminology”, RFC 3753 (Informational), June 2004. (Cited on pages 10 and 23.)

[26] ITU-T, "Next Generation Networks - Generalized mobility, Mobility management requirements for NGN", November 2006. (Cited on pages 10 and 22.)

[27] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description;", December 2011. (Cited on page 12.)

[28] 3GPP, "TR 36.902 - Self-configuring and self-optimizing network (SON) use cases and solutions", March 2011. (Cited on page 12.)

[29] Holger Ziemek, Marius Corici, and Radu Popescu-Zeletin, "COMMUNICATION DEVICE AND METHOD FOR FILTERING DATA ACCORDING TO A DATA POLICY", 2006. (Cited on page 16.)

[30] Liu Xiaoyu, Lee Su Won, Corici Marius, Murarasu Alin, and Arbanowski Stefan, "MOBILE TERMINAL AND NETWORK DEVICE FOR PROVIDING HANDOVER SERVICE", 2009. (Cited on page 16.)

[31] Su-Won Lee, Xiaoyu Liu, John Williams Floroiu, Eun Sun Jung, Hee Jin Jang, Marius Corici, and Byoung-Joon Lee, "MOBILE TERMINAL AND NETWORK DEVICE FOR PROVIDING HANDOVER SERVICE", 2008. (Cited on page 16.)

[32] Xiaoyu LIU, Su Won Lee, Byung Suk Kim, Byoung-Joon Lee, and Marius-Iulian Corici, "APPARATUS FOR DISTRIBUTING DATA TRAFFIC IN HETEROGENEOUS WIRELESS NETWORKS", 2009. (Cited on page 16.)

[33] M. Corici, A. Diez, T. Magedanz, C. Pampu, and Q. Zhou, "SYSTEM AND METHOD FOR MANAGING AN ACCESS NETWORK RESELECTION", 2010. (Cited on page 16.)

[34] C. Pampu, Q. Zhou, M. Corici, A. Diez, T. Magedanz, and D.VIngarzan, "METHOD FOR RE-SELECTING A COMMUNICATION NETWORK", 2010. (Cited on page 16.)

[35] M. Corici, V. Dragos, T. Magedanz, K. Pentikousis, Q. Zhou, and C. Pampu, "METHOD AND DEVICE FOR ADAPTING A DATA PATH FOR A MOBILE ENTITY IN A COMMUNICATION NETWORK", 2011. (Cited on page 16.)

[36] M. Corici, V. Dragos, T. Magedanz, K. Pentikousis, Q. Zhou, and C. Pampu, "METHOD FOR MANAGING A DATA PATH, MOBILITY CONTROLLER AND COMMUNICATION SYSTEM", 2011. (Cited on page 16.)

[37] M. Corici, V. Dragos, T. Magedanz, K. Pentikousis, Q. Zhou, and C. Pampu, “METHOD FOR ESTABLISHING A DATA PATH, DEVICE AND COMMUNICATION SYSTEM”, 2011. (Cited on page 16.)

[38] D. Sisalem, M. Corici, S. Ehlert, and R. Popescu-Zeletin, “VDSat: Nomadic Satellite-Based VoIP Infrastructure”, in *Wireless Communication Systems, 2005. 2nd International Symposium on*, sept. 2005, pp. 619 –623. (Cited on page 16.)

[39] J. Williams Floroiu, M. Corici, Byoung-Joon Lee, Suwon Lee, S. Arbanowski, and T. Magedanz, “A Vertical Handover Architecture for End-to-End Service Optimization”, in *Mobile and Wireless Communications Summit, 2007. 16th IST*, july 2007, pp. 1 –5. (Cited on page 16.)

[40] J.W. Floroiu, M. Corici, and S. Arbanowski, “An Architecture for Resource Management in IMS Wireless Access Networks”, in *Communications Workshops, 2008. ICC Workshops ’08. IEEE International Conference on*, may 2008, pp. 446 –449. (Cited on page 16.)

[41] M.I. Corici, F.C. de Gouveia, and T. Magedanz, “A Network Controlled QoS Model over the 3GPP System Architecture Evolution”, in *Wireless Broadband and Ultra Wideband Communications, 2007. AusWireless 2007. The 2nd International Conference on*, aug. 2007, p. 39. (Cited on page 16.)

[42] M.I. Corici, A.F. Murarasu, S. Arbanowski, T. Magedanz, Suwon Lee, and Xiaoyu Liu, “Multimedia Mobility Service Solution”, in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, sept. 2008, pp. 1 –5. (Cited on page 16.)

[43] T. Magedanz M. Corici, F. Gouveia, *Advances in Broadband Communication and Networks - A Network Controlled QoS Model over the 3GPP Evolved Packet Core*, Number Chapter 12. River Publishers Series in Communication, September 2008. (Cited on page 16.)

[44] A.D. Albaladejo, F.C. de Gouveia, M.I. Corici, and T. Magedanz, “The PCC Rule in the 3GPP IMS Policy and Charging Control Architecture”, in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 30 2008-dec. 4 2008, pp. 1 –5. (Cited on page 16.)

[45] M. Corici, A. Diez, D. Vingarzan, T. Magedanz, C. Pampu, and Zhou Qing, “Enhanced access network discovery and selection in 3GPP Evolved Packet Core”, in *Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on*, oct. 2009, pp. 677 –682. (Cited on page 16.)

[46] Gouveia F. Corici M. Magedanz T. Diez, A., *Quality of Service Architectures for Wireless Networks: Performance Metrics and Management - Chapter on Evolution of QoS control in Next Generation Mobile Networks*, Number Chapter 26. Information Science Reference, January 2010. (Cited on page 16.)

[47] Marius Corici, Fabricio Gouveia, Thomas Magedanz, and Dragos Vingarzan, "OpenEPC: A Technical Infrastructure for Early Prototyping of NGMN Testbeds", in *Testbeds and Research Infrastructures. Development of Networks and Communities*, Thomas Magedanz, Anastasius Gavras, Nguyen Huu Thanh, Jeffry S. Chase, Ozgur Akan, Paolo Bellavista, Jiannong Cao, Falko Dressler, Domenico Ferrari, Mario Gerla, Hisashi Kobayashi, Sergio Palazzo, Sartaj Sahni, Xuemin (Sherman) Shen, Mircea Stan, Jia Xiaohua, Albert Zomaya, and Geoffrey Coulson, Eds., vol. 46 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 166–175. Springer Berlin Heidelberg, 2011, 10.1007/978-3-642-17851-1_13. (Cited on page 16.)

[48] Marius Corici, Jens Fiedler, Thomas Magedanz, and Dragos Vingarzan, "Access Network Discovery and Selection in the Future Wireless Communication", *Mobile Networks and Applications*, vol. 16, pp. 337–349, 2011, 10.1007/s11036-011-0309-3. (Cited on page 16.)

[49] M. Corici, Magedanz T., and Vingarzan D., "3GPP Evolved Packet Core - the Next Generation Mobile Networks all-IP architecture", in *Proceedings of World Telecommunications Congress 2010: The Infrastructure for the 21st Century,*. September 2010, OVE GIT, ITG (Hrsg.). (Cited on page 16.)

[50] Marius Corici, Dragos Vingarzan, Thomas Magedanz, Cornel Pampu, and Qing Zhou, "Proactive Vertical Handover Optimizations in the 3GPP Evolved Packet Core", in *Mobile Networks and Management*, Kostas Pentikousis, Ramon Aguero, Marta Garcia-Arranz, Symeon Papavassiliou, Ozgur Akan, Paolo Bellavista, Jiannong Cao, Falko Dressler, Domenico Ferrari, Mario Gerla, Hisashi Kobayashi, Sergio Palazzo, Sartaj Sahni, Xuemin (Sherman) Shen, Mircea Stan, Jia Xiaohua, Albert Zomaya, and Geoffrey Coulson, Eds. 2011, vol. 68 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 109–122, Springer Berlin Heidelberg, 10.1007/978-3-642-21444-8_11. (Cited on pages 16 and 169.)

[51] M. Corici, D. Vingarzan, T. Magedanz, C. Pampu, and Qing Zhou, "Access Network Reselection Based on Momentary Resources in a Converged Wire-

less Environment", in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, dec. 2010, pp. 1 –6. (Cited on pages 16 and 167.)

[52] M. Corici, T. Magedanz, D. Vingarzan, and P. Weik, "Enabling Ambient Aware Service Delivery in Heterogeneous Wireless Environments", in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, dec. 2010, pp. 1 –6. (Cited on pages 16 and 174.)

[53] M. Corici, J. Fiedler, A. Onofrei, and D. Vingarzan, "Enabling dynamic service delivery in the 3GPP Evolved Packet Core", in *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, dec. 2010, pp. 2012 –2016. (Cited on page 16.)

[54] Marius Corici, Jens Fiedler, Thomas Magedanz, and Dragos Vingarzan, "Access Network Discovery and Selection in the Future Broadband Wireless Environment", in *Mobile Wireless Middleware, Operating Systems, and Applications*, Ying Cai, Thomas Magedanz, Minglu Li, Jinchun Xia, Carlo Giannelli, Ozgur Akan, Paolo Bellavista, Jiannong Cao, Falko Dressler, Domenico Ferrari, Mario Gerla, Hisashi Kobayashi, Sergio Palazzo, Sartaj Sahni, Xuemin (Sherman) Shen, Mircea Stan, Jia Xiaohua, Albert Zomaya, and Geoffrey Coulson, Eds. 2010, vol. 48 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 70–83, Springer Berlin Heidelberg, 10.1007/978-3-642-17758-3_6. (Cited on page 16.)

[55] M. Corici, A. Onofrei, D. Vingarzan, J. Mueller, and T. Magedanz, "Massive deployment of small coverage area cells in the all-IP wireless system - Opportunities, issues and solutions", in *Telecom World (ITU WT), 2011 Technical Symposium at ITU*, oct. 2011, pp. 147 –152. (Cited on page 16.)

[56] J. Mueller, T. Magedanz, M. Corici, and D. Vingarzan, "UE amp; network initiated QoS reservation in NGN and beyond", in *Network of the Future (NOF), 2011 International Conference on the*, pp. 62 –67. nov. 2011. (Cited on page 16.)

[57] M. Corici, J. Fiedler, T. Magedanz, and D. Vingarzan, "Evolution of the resource reservation mechanisms for machine type communication over mobile broadband Evolved Packet Core architecture", in *GLOBECOM Workshops (GC Wkshps), 2011 IEEE*, dec. 2011, pp. 718 –722. (Cited on page 16.)

[58] M. Corici, J. Fiedler, D. Vingarzan, and T. Magedanz, "Optimized low mobility support in massive mobile broadband Evolved Packet Core architecture", in *Networks (ICON), 2011 17th IEEE International Conference on*, dec. 2011, pp. 322 –327. (Cited on page 16.)

[59] M. Corici, J. Fiedler, D. Vingarzan, and T. Magedanz, "Device connectivity management for mobile All-IP broadband network architecture", in *Signal Processing and Communication Systems (ICSPCS), 2011 5th International Conference on*, dec. 2011, pp. 1 –7. (Cited on page 16.)

[60] Marius Corici, Thomas Magedanz, and Dragos Vingarzan, "Developing Optimized Future Mobile Broadband Platforms and Services", Informa 3G, HSPA+ & LTE Optimization, March 2010. (Cited on page 16.)

[61] Marius Corici, Thomas Magedanz, and Dragos Vingarzan, "Deploying a Next Generation Core Network for 3G - What are the Challenges with the Current 3G Architecture?", Informa Next Generation Core Networks (NGCN) Summit, May 2010. (Cited on page 16.)

[62] Marius Corici, Thomas Magedanz, and Dragos Vingarzan, "How can Operators Escape "Black Box" Standards? Unleashing Innovative Interoperability by the Right Toolkits and Open Testbeds", Informa Next Generation Core Networks (NGCN) Summit, March 2011. (Cited on page 16.)

[63] Thomas Magedanz, Marius Corici, and Florian Schreiner, "Evolved Packet Core as a Smart Bitpipe to Mobile Clouds", Informa 3G, HSPA+ & LTE Optimization, March 2011. (Cited on page 16.)

[64] Marius Corici and Thomas Magedanz, "The Strategic Importance of the 3GPP Evolved Packet Core (EPC) for Seamless Mobile Cloud Access", Informa Transport Networks for Mobile Operators (TMNO), March 2011. (Cited on page 16.)

[65] Marius Corici and Thomas Magedanz, "NGN2FI - Interconnection Challenges in the Emerging Future Internet", Marcus Evans - 5th Annual Interconnection Strategies for Advanced Telecoms Analysing Next Generation Networks and their Effect on Interconnections, September 2011. (Cited on page 16.)

[66] Marius Corici, Thomas Magedanz, and Dragos Vingarzan, "Towards the Wireless Future Internet - Understanding the Next Generation of Mobile Networks and the Evolved Packet Core", The 6th International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities, May 2010. (Cited on page 16.)

[67] Marius Corici, Thomas Magedanz, and Dragos Vingarzan, "Getting Started with the 3GPP Evolved Packet Core (EPC) - Basics, Standards, Architecture and Capabilities", Informa 3G, HSPA+ & LTE Optimization, March 2010. (Cited on page 16.)

[68] Marius Corici and Thomas Magedanz, "Evolved Packet Core Workshop: The OpenEPC Toolkit for Rapid Prototyping of Mobile Next Generation Networks and Emerging Future Internet Infrastructures and Applications", Informa Transport Networks for Mobile Operators (TMNO), March 2010. (Cited on page 16.)

[69] Thomas Haustein, Thomas Magedanz, and Marius Corici, "Understanding the Next Generation of Mobile Broadband Communications: LTE and EPC Concepts, Architectures, Protocols and Applications", Fraunhofer FOKUS Future Seamless Communication (FUSECO) Forum, October 2010. (Cited on page 16.)

[70] Marius Corici and Dragos Vingarzan, "Getting started with OpenEPC and Open IMS Core to set-up your own FUSECO Testbed", Fraunhofer FOKUS Future Seamless Communication (FUSECO) Forum, November 2011. (Cited on page 16.)

[71] Marius Corici, Hakan Coskun, and Thomas Magedanz, "Telecom Signaling Evolution for IMS, OTT and M2M Services - Introducing Evolved Packet Core (EPC) and Machine Type Communication (MTC", Marcus Evans 13th Annual NGN Signalling Evolution Forum 2012, January 2012. (Cited on page 16.)

[72] Marius Corici and all, *OpenEPC Toolkit Website*, Fraunhofer FOKUS Institute. (Cited on pages 16, 128 and 129.)

[73] Marius Corici and all, *OpenEPC Reference Customers*, Fraunhofer FOKUS Institute, 2012. (Cited on page 16.)

[74] European Comission FI-PPP Framework, *The Future of the Internet Public Private Partnership (FI-PPP) - Future Internet Core Platform (FI-WARE) Project*, 2012. (Cited on pages 16, 128 and 199.)

[75] ITU-T, "Functional requirements and architecture of next generation networks", April 2010. (Cited on page 21.)

[76] 3GPP, "TS 22.038 - (U)SIM Application Toolkit (USAT); Service description; Stage 1", February 2011. (Cited on page 21.)

[77] S. Barkley D. Nelson B. Patil M. Stevens B. Wolff D. Mitton, M. St.Johns, "RFC 3127 Authentication, Authorization, and Accounting: Protocol Evaluation", June 2001. (Cited on page 21.)

[78] 3GPP, "TS 23.060 - General Packet Radio Service (GPRS); Service description", December 2011. (Cited on pages 21, 28, 57 and 65.)

[79] 3GPP, “TS 23.203 - Policy and Charging Control Architecture”, December 2011. (Cited on pages 22, 28, 30, 32, 80 and 81.)

[80] C. Perkins, “IP Mobility Support”, RFC 2002 (Proposed Standard), Oct. 1996, Obsoleted by RFC 3220, updated by RFC 2290. (Cited on pages 23 and 32.)

[81] 3GPP, “TR 23.893 - Feasibility Study on Multimedia Session Continuity”, June 2008. (Cited on pages 23 and 72.)

[82] Cisco, “Cisco Visual Networking Index: Forecast and Methodology 2011-2016”, Cisco Visual Networking Index, February 2012. (Cited on pages 23, 25, 38, 39, 41 and 42.)

[83] 3GPP, “TR 36.912 - Feasibility study for Further Advancements for E-UTRA (LTE-Advanced)”, March 2011. (Cited on page 23.)

[84] 3GPP, “TR 36.913 - Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced)”, April 2011. (Cited on page 23.)

[85] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, “LTE-advanced: next-generation wireless broadband technology [Invited Paper]”, *Wireless Communications, IEEE*, vol. 17, no. 3, pp. 10 –22, june 2010. (Cited on page 23.)

[86] Martin Sauter, *From GSM to LTE: An Introduction to Mobile Networks and Mobile Broadband*, Wiley Publishing Inc., 2011. (Cited on pages 24 and 25.)

[87] ITU-T, “REPORT ITU-R M.2134: Requirements related to technical performance for IMT-Advanced radio interface(s)”, January 2008. (Cited on page 25.)

[88] Alan Gatherer Vikram Chandrasekhar, Jeffrey G. Andrews, “Femtocell Networks: A Survey”, *Communications Magazine, IEEE*, vol. 46 Issue 9, no. ISSN: 0163-6804, pp. 59–67, September 2008. (Cited on page 25.)

[89] Nazmul Haq Siddique Syed Faraz Hasan and Shyam Chakraborty, “Femtocell versus WiFi - A Survey and Comparison of Architecture and Performance”, *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on*, vol. 1, no. ISBN 978-1-4244-4066-5, pp. 916 – 920, May 2009. (Cited on page 25.)

[90] GSA The Global mobile Suppliers Association, “LTE Developments Worldwide”, Global Mobile Suppliers Association, November 2011. (Cited on page 25.)

[91] Ranveer Chandra Alec Wolman Brian Zill Rohan Murty, Jitendra Padhye, “Designing high performance enterprise Wi-Fi networks”, *NSDI’08 Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, vol. 1, no. ISBN: 111-999-5555-22-1, pp. 73–88, 2008. (Cited on page 25.)

[92] 3GPP2 Vision AdHoc, “The 3GPP2 Vision”, 3rd Generation Partnership Project 2 (3GPP2), October 2009. (Cited on page 25.)

[93] Broadband Forum, “Next Generation Broadband Access White Paper”, Broadband Forum, August 2009. (Cited on page 26.)

[94] W. Joseph and L. Martens, “Performance evaluation of broadband fixed wireless system based on IEEE 802.16”, in *Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE*, april 2006, vol. 2, pp. 978 –983. (Cited on page 26.)

[95] 3GPP, “TR 23.882 - 3GPP system architecture evolution (SAE): Report on technical options and conclusions”, September 2008. (Cited on page 26.)

[96] Analysys Mason Limited, “The momentum behind LTE worldwide”, NGMA, 2011. (Cited on page 26.)

[97] ITU-T, “Y.2018 - Mobility management and control framework and architecture within the NGN transport stratum”, September 2009. (Cited on pages 27 and 55.)

[98] 3GPP, “Overview of 3GPP Release 11”, 3GPP, January 2012. (Cited on page 27.)

[99] 3GPP, “Overview of 3GPP Release 12”, 3GPP, January 2012. (Cited on page 27.)

[100] 3GPP, “TS 23.228 - IP Multimedia Subsystem (IMS)”, December 2011. (Cited on pages 28, 71 and 75.)

[101] 3GPP, “TS 24.301 - Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS)”, December 2011. (Cited on pages 28, 30, 57, 65 and 66.)

[102] 3GPP, “TS 24.302 - Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks”, December 2011. (Cited on pages 28, 31, 58, 59 and 60.)

[103] 3GPP, "TS 29.272 - Evolved Packet System (EPS); Mobility Management Entity (MME) and Serving GPRS Support Node (SGSN) related interfaces based on Diameter protocol", December 2011. (Cited on pages 30, 32 and 57.)

[104] 3GPP, "TS 29.273 - Evolved Packet System (EPS); 3GPP EPS AAA interfaces", December 2011. (Cited on pages 30 and 32.)

[105] 3GPP, "TS 29.274 - 3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C)", December 2011. (Cited on pages 30, 32, 58 and 65.)

[106] 3GPP, "TS 23.234 - 3GPP system to Wireless Local Area Network (WLAN) interworking; System description", March 2011. (Cited on page 30.)

[107] 3GPP, "TS 29.212 - Policy and Charging Control (PCC) over Gx/Sd reference point", December 2011. (Cited on pages 30, 32 and 80.)

[108] 3GPP, "TS 29.213 - Policy and charging control signalling flows and Quality of Service (QoS) parameter mapping", December 2011. (Cited on pages 30, 32 and 81.)

[109] 3GPP, "TS 29.214 - Policy and charging control over Rx reference point", December 2011. (Cited on pages 30, 32, 35 and 81.)

[110] 3GPP, "TS 29.215 - Policy and Charging Control (PCC) over S9 reference point", December 2011. (Cited on pages 30 and 32.)

[111] 3GPP, "TS 24.312 - Access Network Discovery and Selection Function (ANDSF) Management Object (MO)", December 2011. (Cited on pages 31, 32 and 133.)

[112] P. Calhoun, J. Loughney, E. Guttman, G. Zorn, and J. Arkko, "Diameter Base Protocol", RFC 3588 (Proposed Standard), September 2003, Updated by RFCs 5729, 5719, 6408. (Cited on page 32.)

[113] George Tsirtsis Hesham Soliman, "Dual Stack Mobile IPv6", July 2005. (Cited on pages 32 and 65.)

[114] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy Mobile IPv6", RFC 5213 (Proposed Standard), Aug. 2008. (Cited on pages 32 and 65.)

[115] 3GPP, "TS 29.275 - Proxy Mobile IPv6 (PMIPv6) based Mobility and Tunnelling protocols", December 2011. (Cited on page 32.)

[116] "OMA Device Management Protocol", December 2009. (Cited on page 32.)

[117] 3GPP, "TS 33.401 - 3GPP System Architecture Evolution (SAE); Security architecture", December 2011. (Cited on pages 33 and 140.)

[118] 3GPP, "TS 33.402 - 3GPP System Architecture Evolution (SAE); Security aspects of non-3GPP accesses", December 2011. (Cited on page 33.)

[119] Nielsen, "Android Leads in U.S. Smartphone Market Share and Data Usage", online, April 2011. (Cited on page 38.)

[120] R. Wakikawa C. Ng K. Kuladinithi T. Ernst, N. Montavont, "Motivations and Scenarios for Using Multiple Interfaces and Global Addresses", May 2008. (Cited on page 38.)

[121] Google, *Android SDK: The Developer's Guide*, Google. (Cited on page 40.)

[122] Apple Inc., *iOS Developer Library: Reference*, Apple Inc. (Cited on page 40.)

[123] 3GPP, "TS 22.368 - Service requirements for Machine-Type Communications (MTC)", October 2011. (Cited on page 40.)

[124] Gottfried Punz, *Evolution of 3G Networks - The Concept, Architecture and Realizationof Mobile Networks Beyond UMTS*, Number ISBN 978-3-211-09439-6. Springer-Verlag/Wien, 2010. (Cited on page 41.)

[125] IEEE Standards Society, "IEEE Standard and Local and Metropolitan Area Networks - Part 21: Media Independent Handover Services", 2009. (Cited on page 52.)

[126] L. Eastwood, S. Migaldi, Qiaobing Xie, and V. Gupta, "Mobility using IEEE 802.21 in a heterogeneous IEEE 802.16/802.11-based, IMT-advanced (4G) network", *Wireless Communications, IEEE*, vol. 15, no. 2, pp. 26 –34, april 2008. (Cited on page 54.)

[127] Ashutosh Dutta, Subir Das, David Famolari, Yoshihiro Ohba, Kenichi Taniuchi, Victor Fajardo, and Toshikazu Kodama, "Secured Seamless Convergence across Heterogeneous Access Networks", 2006. (Cited on page 54.)

[128] G. Lampropoulos, A.K. Salkintzis, and N. Passas, "Media-independent handover for seamless service provision in heterogeneous networks", *Communications Magazine, IEEE*, vol. 46, no. 1, pp. 64 –71, january 2008. (Cited on page 54.)

[129] Shun-Fang Yang, Jung-Shyr Wu, and Hsu-Hung Huang, "A vertical Media-Independent Handover decision algorithm across Wi-Fi and WiMAX networks", in *Wireless and Optical Communications Networks, 2008. WOCN*

*'08. 5th IFIP International Conference on*, may 2008, pp. 1 –5. (Cited on page 54.)

[130] M.A. Abdelatif, G.K. Kalebaila, and H.A. Chan, "A Cross-Layer Mobility Management Framework based on IEEE802.21", vol. 1, pp. 1 –6, sept. 2007. (Cited on page 54.)

[131] Su Kyoung Lee, K. Sriram, Kyungsoo Kim, Jong Hyup Lee, Yoon Hyuk Kim, and N. Golmie, "Vertical Handoff Decision Algorithm Providing Optimized Performance in Heterogeneous Wireless Networks", pp. 5164 –5169, nov. 2007. (Cited on page 54.)

[132] Georgios Aristomenopoulos, Timotheos Kastrinogiannis, and Symeon Papavassiliou, "A unified approach for efficient network selection in multi-service integrated CDMA/WLAN systems", in *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly*, New York, NY, USA, 2009, IWCMC '09, pp. 1263–1268, ACM. (Cited on page 54.)

[133] N. Niebert, M. Prytz, A. Schieder, N. Papadoglou, L. Eggert, F. Pittmann, and C. Prehofer, "Ambient networks: a framework for future wireless internetworking", in *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, may-1 june 2005, vol. 5, pp. 2969 – 2973 Vol. 5. (Cited on page 54.)

[134] N. Niebert, A. Schieder, H. Abramowicz, G. Malmgren, J. Sachs, U. Horn, C. Prehofer, and H. Karl, "Ambient networks: an architecture for communication networks beyond 3G", *Wireless Communications, IEEE*, vol. 11, no. 2, pp. 14 – 22, apr 2004. (Cited on page 54.)

[135] B. Ahlgren, L. Eggert, B. Ohlman, and A. Schieder, "Ambient networks: bridging heterogeneous network domains", in *Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on*, sept. 2005, vol. 2, pp. 937 –941 Vol. 2. (Cited on page 54.)

[136] T. Melia, G. Bajko, S. Das, N. Golmie, and JC. Zuniga, "IEEE 802.21 Mobility Services Framework Design (MSFD)", RFC 5677 (Proposed Standard), Dec. 2009. (Cited on pages 54 and 55.)

[137] T. Melia, "Mobility Services Transport: Problem Statement", RFC 5164 (Informational), Mar. 2008. (Cited on page 54.)

[138] G. Bajko and S. Das, "Dynamic Host Configuration Protocol (DHCPv4 and DHCPv6) Options for IEEE 802.21 Mobility Services (MoS) Discovery", RFC 5678 (Proposed Standard), Dec. 2009. (Cited on page 54.)

[139] G. Bajko, “Locating IEEE 802.21 Mobility Services Using DNS”, RFC 5679 (Proposed Standard), Dec. 2009. (Cited on page 55.)

[140] JC. Zuniga M. Watfa H.W. Kim A. Rahman, U. Olvera-Hernandez, “Transport of Media Independent Handover Messages Over IP”, July 2007. (Cited on page 55.)

[141] Corici Marius Murarasu Alin Arbanowski Stefan Liu Xiaoyu, Lee Su Won, “MOBILE TERMINAL AND NETWORK DEVICE FOR PROVIDING HANDOVER SERVICE”, December 2008. (Cited on page 55.)

[142] IEEE Standards Society, “IEEE Std 1900.4-2009, IEEE Standard for Architectural Building Blocks Enabling Network-Device Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Access Networks”, January 2009. (Cited on page 56.)

[143] S. Taranu and J. Tiemann, “General method for testing context aware applications”, June 2009. (Cited on page 56.)

[144] J. Perez-Romero, O. Sallent, R. Agusti, J. Nasreddine, and M. Muck, “Radio Access Technology Selection enabled by IEEE P1900.4”, in *Mobile and Wireless Communications Summit, 2007. 16th IST*, july 2007, pp. 1 –5. (Cited on page 57.)

[145] Yuan Xue, Yuewei Lin, Zhiyong Feng, Huying Cai, and Cheng Chi, “Autonomic Joint Session Scheduling Strategies for Heterogeneous Wireless Networks”, in *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, 31 2008-april 3 2008, pp. 2045 –2050. (Cited on page 57.)

[146] Guo Xiaoli, Li Yuhong, and Hu Weiqi, “An autonomic flow based path selection method for multi-homed nodes”, in *Broadband Network Multimedia Technology, 2009. IC-BNMT '09. 2nd IEEE International Conference on*, oct. 2009, pp. 85 –89. (Cited on page 61.)

[147] I.F. Akyildiz, Jiang Xie, and S. Mohanty, “A survey of mobility management in next-generation all-IP-based wireless systems”, *Wireless Communications, IEEE*, vol. 11, no. 4, pp. 16 – 28, aug. 2004. (Cited on page 63.)

[148] C. Perkins, “IP Mobility Support for IPv4”, RFC 3344 (Proposed Standard), Aug. 2002, Obsoleted by RFC 5944, updated by RFC 4721. (Cited on page 63.)

[149] D. Johnson, C. Perkins, and J. Arkko, “Mobility Support in IPv6”, RFC 3775 (Proposed Standard), June 2004, Obsoleted by RFC 6275. (Cited on page 63.)

[150] E. Nordmark and M. Bagnulo, “Shim6: Level 3 Multihoming Shim Protocol for IPv6”, RFC 5533 (Proposed Standard), June 2009. (Cited on page 64.)

[151] John Ronan and John McLaughlin, “An Empirical Evaluation of a Shim6 Implementation”, in *Mobile Networks and Management*, Kostas Pentikousis, Ramon Aguero, Marta Garcia-Arranz, Symeon Papavassiliou, Ozgur Akan, Paolo Bellavista, Jiannong Cao, Falko Dressler, Domenico Ferrari, Mario Gerla, Hisashi Kobayashi, Sergio Palazzo, Sartaj Sahni, Xuemin (Sherman) Shen, Mircea Stan, Jia Xiaohua, Albert Zomaya, and Geoffrey Coulson, Eds., vol. 68 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 62–70. Springer Berlin Heidelberg, 2011, 10.1007/978-3-642-21444-8_6. (Cited on page 64.)

[152] A. Dhraief and N. Montavont, “Rehoming Decision Algorithm: Design and Empirical Evaluation”, in *Computational Science and Engineering, 2009. CSE ’09. International Conference on*, aug. 2009, vol. 2, pp. 464 –469. (Cited on page 64.)

[153] G. Aristomenopoulos, T. Kastrinogiannis, Z. Li, M. Wilson, M.G. Juan, A.L. Jose, Y. Li, V. Kaldanis, and S. Papavassiliou, “Autonomic mobility and resource management over an integrated wireless environment; A GANA oriented architecture”, in *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, dec. 2010, pp. 545 –550. (Cited on page 64.)

[154] R. Wakikawa, V. Devarapalli, G. Tsirtsis, T. Ernst, and K. Nagami, “Multiple Care-of Addresses Registration”, RFC 5648 (Proposed Standard), Oct. 2009, Updated by RFC 6089. (Cited on page 64.)

[155] G. Tsirtsis, H. Soliman, N. Montavont, G. Giaretta, and K. Kuladinithi, “Flow Bindings in Mobile IPv6 and Network Mobility (NEMO) Basic Support”, RFC 6089 (Proposed Standard), Jan. 2011. (Cited on page 64.)

[156] 3GPP, “TS 29.281 - General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)”, December 2011. (Cited on page 65.)

[157] Consortium of Daidalos, “Final Daidalos II Global Architecture (Deliverable DII-124)”, Daidalos II Project, Decenver 2008. (Cited on pages 66 and 72.)

[158] P. Nikander, A. Gurtov, and T.R. Henderson, “Host Identity Protocol (HIP): Connectivity, Mobility, Multi-Homing, Security, and Privacy over IPv4 and IPv6 Networks”, *Communications Surveys Tutorials, IEEE*, vol. 12, no. 2, pp. 186 –204, quarter 2010. (Cited on page 67.)

[159] J. Ylitalo P. Nikander and J. Wall, "Integrating Security, Mobility, and Multi-Homing in a HIP Way", *Proc. of Network and Distributed Systems Security Symposium (NDSS 03), IEEE*, feb. 2003. (Cited on page 67.)

[160] R. Moskowitz, P. Nikander, P. Jokela, and T. Henderson, "Host Identity Protocol", RFC 5201 (Experimental), Apr. 2008, Updated by RFC 6253. (Cited on page 67.)

[161] P. Nikander, T. Henderson, C. Vogt, and J. Arkko, "End-Host Mobility and Multihoming with the Host Identity Protocol", RFC 5206 (Experimental), Apr. 2008. (Cited on page 67.)

[162] Keisuke Uehara Hiroshi Esaki Fumio Teraoka Masahiro Ishiyama, Mitsunobu Kunishi, "LINA: A new Approach to mobility support in wide area networks", *IEICE trans. Commun.*, vol. E84-B, no. 8, pp. 2076–2086, August 2001. (Cited on page 67.)

[163] Keisuke Uehara Hiroshi Esaki Fumio Teraoka Mitsunobu Kunishi, Masahiro Ishiyama, "LIN6: A New Approach to Mobility Support in IPv6", 2000. (Cited on page 67.)

[164] R. Moskowitz and P. Nikander, "Host Identity Protocol (HIP) Architecture", RFC 4423 (Informational), May 2006. (Cited on page 67.)

[165] J. Laganier, T. Koponen, and L. Eggert, "Host Identity Protocol (HIP) Registration Extension", RFC 5203 (Experimental), Apr. 2008. (Cited on page 67.)

[166] J. Laganier and L. Eggert, "Host Identity Protocol (HIP) Rendezvous Extension", RFC 5204 (Experimental), Apr. 2008. (Cited on page 67.)

[167] P. Jokela, R. Moskowitz, and P. Nikander, "Using the Encapsulating Security Payload (ESP) Transport Format with the Host Identity Protocol (HIP)", RFC 5202 (Experimental), Apr. 2008. (Cited on page 67.)

[168] P. Nikander and J. Laganier, "Host Identity Protocol (HIP) Domain Name System (DNS) Extensions", RFC 5205 (Experimental), Apr. 2008. (Cited on page 68.)

[169] Cedric de Launois Olivier Bonaventure Bruno Quoitin, Luigi Iannone, "Evaluating the Benefits of the Locator/Identifier Separation", *MobiArch 07 Proceedings of 2nd ACM/IEEE international workshop on Mobility in the evolving internet architecture*, vol. 1, no. ISBN 978-1-59593-784-1, pp. 1–6, August 2007. (Cited on page 68.)

[170] D. Meyer D. Lewis D. Farinacci, V. Fuller, "Locator/ID Separation Protocol (LISP)", February 2012. (Cited on page 68.)

[171] D. Meyer C. White D. Farinacci, D. Lewis, "LISP Mobile Node", February 2012. (Cited on pages 68 and 69.)

[172] D. Farinacci V. Fuller D. Lewis D. Meyer S. Brim, N. Chiappa, "LISP-CONS: A Content distribution Overlay Network Service for LISP", April 2008. (Cited on page 69.)

[173] D. Meyer D. Lewis V. Fuller, D. Farinacci, "LISP Alternative Topology (LISP+ALT)", December 2011. (Cited on page 69.)

[174] E. Lear, "NERD: A Not-so-novel EID to RLOC Database", march 2010. (Cited on page 69.)

[175] D. Farinacci V. Fuller D. Lewis, D. Meyer, "Interworking LISP with IPv4 and IPv6", march 2010. (Cited on page 69.)

[176] Matthew Caesar, Tyson Condie, Jayanthkumar Kannan, Karthik Lakshminarayanan, and Ion Stoica, "ROFL: routing on flat labels", in *Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, New York, NY, USA, 2006, SIGCOMM '06, pp. 363–374, ACM. (Cited on page 69.)

[177] Luis M. Correia, Henrik Abramowicz, Martin Johnsson, and Klaus Wuenstel, *Architecture and Design for the Future Internet: 4WARD Project: 4WARD EU Project (Signals and Communication Technology)*, Springer Science+Business Media, 2011. (Cited on page 70.)

[178] P. Bertin, Rui L Aguiar, M. Folke, P Schefczik, and X Zhang, "Paths to Mobility Support in the Future Internet", in *ICT-MobileSummit 2009: ICT Mobile and Wireless Communications Summit*, European Union, Ed., 2009. (Cited on page 70.)

[179] A. Eriksson and B. Ohlman, "Dynamic Internetworking Based on Late Locator Construction", in *IEEE Global Internet Symposium, 2007*, may 2007, pp. 67 –72. (Cited on page 70.)

[180] Jianli Pan, S. Paul, R. Jain, and M. Bowman, "MILSA: A Mobility and Multihoming Supporting Identifier Locator Split Architecture for Naming in the Next Generation Internet", in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 30 2008-dec. 4 2008, pp. 1 –6. (Cited on page 70.)

[181] Jianli Pan, R. Jain, S. Paul, M. Bowman, Xiaohu Xu, and Shanzhi Chen, "Enhanced MILSA Architecture for Naming, Addressing, Routing and Security Issues in the Next Generation Internet", in *Communications, 2009. ICC '09. IEEE International Conference on*, june 2009, pp. 1 –6. (Cited on page 70.)

[182] Sebastien Barre, Christoph Paasch, and Olivier Bonaventure, "MultiPath TCP: From Theory to Practice", in *NETWORKING 2011*, Jordi Domingo-Pascual, Pietro Manzoni, Sergio Palazzo, Ana Pont, and Caterina Scoglio, Eds., vol. 6640 of *Lecture Notes in Computer Science*, pp. 444–457. Springer Berlin / Heidelberg, 2011, 10.1007/978-3-642-20757-0_35. (Cited on page 70.)

[183] A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar, "Architectural Guidelines for Multipath TCP Development", RFC 6182 (Informational), Mar. 2011. (Cited on page 70.)

[184] M. Handley O. Bonaventure A. Ford, C. Raiciu, "TCP Extensions for Multipath Operation with Multiple Addresses", January 2012. (Cited on page 70.)

[185] J.R. Iyengar, P.D. Amer, and R. Stewart, "Concurrent Multipath Transfer Using SCTP Multihoming Over Independent End-to-End Paths", *Networking, IEEE/ACM Transactions on*, vol. 14, no. 5, pp. 951 –964, oct. 2006. (Cited on page 70.)

[186] Seok Joo Koh, Moon Jeong Chang, and Meejeong Lee, "mSCTP for soft handover in transport layer", *Communications Letters, IEEE*, vol. 8, no. 3, pp. 189 – 191, march 2004. (Cited on page 70.)

[187] Ed. M. Riegel, M. Tuexen, "Mobile SCTP", November 2007. (Cited on page 70.)

[188] Z. Zhu, R. Wakikawa, and L. Zhang, "A Survey of Mobility Support in the Internet", RFC 6301 (Informational), July 2011. (Cited on page 71.)

[189] P. Vixie, S. Thomson, Y. Rekhter, and J. Bound, "Dynamic Updates in the Domain Name System (DNS UPDATE)", RFC 2136 (Proposed Standard), Apr. 1997, Updated by RFCs 3007, 4035, 4033, 4034. (Cited on page 71.)

[190] B. Wellington, "Secure Domain Name System (DNS) Dynamic Update", RFC 3007 (Proposed Standard), Nov. 2000, Updated by RFCs 4033, 4034, 4035. (Cited on page 71.)

[191] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261 (Proposed Standard), June 2002, Updated by RFCs 3265,

3853, 4320, 4916, 5393, 5621, 5626, 5630, 5922, 5954, 6026, 6141. (Cited on page 71.)

[192] H. Schulzrinne and E. Wedlund, "Application-layer mobility using SIP", pp. 29 –36, 2000. (Cited on page 71.)

[193] Wu Q. Lin Y.-B. Huang, S.-M. and C.-H. Yeh, "SIP mobility and IPv4/IPv6 dual-stack supports in 3G IP multimedia subsystem", *Wireless Communications and Mobile Computing*, vol. 6, pp. 585–599, July 2006. (Cited on page 71.)

[194] Zahid Ghadialy, "Over-the-top (OTT) Applications and Services". (Cited on page 74.)

[195] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633 (Informational), June 1994. (Cited on page 75.)

[196] J. Wroclawski, "The Use of RSVP with IETF Integrated Services", RFC 2210 (Proposed Standard), Sept. 1997. (Cited on page 75.)

[197] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification", RFC 2205 (Proposed Standard), Sept. 1997, Updated by RFCs 2750, 3936, 4495, 5946, 6437. (Cited on page 75.)

[198] S. Shenker, C. Partridge, and R. Guerin, "Specification of Guaranteed Quality of Service", RFC 2212 (Proposed Standard), Sept. 1997. (Cited on page 75.)

[199] J. Wroclawski, "Specification of the Controlled-Load Network Element Service", RFC 2211 (Proposed Standard), Sept. 1997. (Cited on page 75.)

[200] Geng-Sheng Kuo and Po-Chang Ko, "Dynamic RSVP for mobile IPv6 in wireless networks", in *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, 2000, vol. 1, pp. 455 –459 vol.1. (Cited on page 76.)

[201] A. Acharya A. K. Talukdar, B. R. Badrinath, "MRSVP: A Resource Reservation Protocol for an Integrated Services Network with Mobile Hosts", *ANUP KUMAR TALUKDAR*, vol. 7, pp. 5–19, 2001. (Cited on page 76.)

[202] M. Brunner, "Requirements for Signaling Protocols", RFC 3726 (Informational), Apr. 2004. (Cited on page 76.)

[203] H. Chaskar, “Requirements of a Quality of Service (QoS) Solution for Mobile IP”, RFC 3583 (Informational), Sept. 2003. (Cited on page 76.)

[204] R. Hancock, G. Karagiannis, J. Loughney, and S. Van den Bosch, “Next Steps in Signaling (NSIS): Framework”, RFC 4080 (Informational), June 2005. (Cited on page 76.)

[205] K. Nichols, S. Blake, F. Baker, and D. Black, “Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers”, RFC 2474 (Proposed Standard), Dec. 1998, Updated by RFCs 3168, 3260. (Cited on page 76.)

[206] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differentiated Service”, RFC 2475 (Informational), Dec. 1998, Updated by RFC 3260. (Cited on page 76.)

[207] H. Schulzrinne and R. Hancock, “GIST: General Internet Signalling Transport”, RFC 5971 (Experimental), Oct. 2010. (Cited on page 76.)

[208] T. Tsenov, H. Tschofenig, X. Fu, C. Aoun, and E. Davies, “General Internet Signaling Transport (GIST) State Machine”, RFC 5972 (Informational), Oct. 2010. (Cited on page 76.)

[209] J. Manner, G. Karagiannis, and A. McDonald, “NSIS Signaling Layer Protocol (NSLP) for Quality-of-Service Signaling”, RFC 5974 (Experimental), Oct. 2010. (Cited on page 76.)

[210] A. Westerinen, J. Schnizlein, J. Strassner, M. Scherling, B. Quinn, S. Herzog, A. Huynh, M. Carlson, J. Perry, and S. Waldbusser, “Terminology for Policy-Based Management”, RFC 3198 (Informational), Nov. 2001. (Cited on page 77.)

[211] B. Moore, E. Ellesson, J. Strassner, and A. Westerinen, “Policy Core Information Model – Version 1 Specification”, RFC 3060 (Proposed Standard), Feb. 2001, Updated by RFC 3460. (Cited on page 77.)

[212] D. Durham, J. Boyle, R. Cohen, S. Herzog, R. Rajan, and A. Sastry, “The COPS (Common Open Policy Service) Protocol”, RFC 2748 (Proposed Standard), Jan. 2000, Updated by RFC 4261. (Cited on page 77.)

[213] Tom Anderson Larry Peterson Hari Balakrishnan Jennifer Rexford Scott Shenker Jonathan Turner Nick McKeown, Guru Parulkar, “OpenFlow: Enabling Innovation in Campus Networks”, *ACM SIGCOMM Computer Communication Review*, vol. 38(2), pp. 69–74, April 2008. (Cited on page 79.)

[214] Open Networking Foundation (ONF), ". (Cited on page 79.)

[215] Brandon Heller Casey Barker Dan Cohn Dan Talayco David Erickson Edward Crabbe Glen Gibb Guido Appenzeller Jean Tourrilhes Justin Pettit KK Yap Leon Poutievski Martin Casado Masahiko Takahashi Masayoshi Kobayashi Nick McKeown Peter Balland Rajiv Ramanathan Reid Price Rob Sherwood Saurav Das Tatsuya Yabe Yiannis Yiakoumis Zoltan Lajos Kis. Ben Pfaff, Bob Lantz, "OpenFlow Switch Specification", February 2011. (Cited on page 79.)

[216] 3GPP, "TS 32.240 - Telecommunication management; Charging management; Charging architecture and principles", December 2011. (Cited on page 80.)

[217] 3GPP, "TS 32.260 - Telecommunication management; Charging management; IP Multimedia Subsystem (IMS) charging", December 2011. (Cited on page 80.)

[218] 3GPP, "TS 32.295 - Telecommunication management; Charging management; Charging Data Record (CDR) transfer", April 2011. (Cited on page 80.)

[219] 3GPP, "TS 32.297 - Charging management; Charging Data Record (CDR) file format and transfer", December 2011. (Cited on page 80.)

[220] 3GPP, "TS 32.298 - Charging management; Charging Data Record (CDR) parameter description", December 2011. (Cited on page 80.)

[221] 3GPP, "TS 32.299 - Charging management; Diameter charging applications", December 2011. (Cited on page 80.)

[222] European Comission, *EU Framework Programme 7 (FP7) - UniverSelf Project*, 2012. (Cited on pages 128 and 200.)

[223] European Comission, *EU Framework Programme 7 (FP7) - MobileCloud Project*, 2012. (Cited on pages 128 and 200.)

[224] Andrei Pelinescu-Onciul and all, *SIP Express Router*, Open Source Project. (Cited on page 129.)

[225] Vingarzan D. and all, "ETSI TS 186 008-2 IMS/NGN Performance Benchmark, Part 2: Subsystem Configurations and Benchmarks1", October 2010. (Cited on page 168.)

[226] R Wakikawa L. Zhang R. Kuntz, D. Sudhakar, "A Summary of Distributed Mobility Management", August 2011. (Cited on pages 181 and 186.)

[227] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier, "Hierarchical Mobile IPv6 (HMIPv6) Mobility Management", RFC 5380, Oct. 2008. (Cited on page 187.)

[228] F. Giust CJ. Bernardos, A. de la Oliva, "A IPv6 Distributed Client Mobility Management approach using existing mechanisms", March 2011. (Cited on page 187.)

[229] L. Beloeil X. Brouckaert M. Kassi-Lahlou, C. Jacquenet, "Dynamic Mobile IP (DMI)", January 2003. (Cited on page 188.)

[230] P. Seite and P. Bertin, "Dynamic Mobility Anchoring (DMA)", May 2010. (Cited on page 188.)

[231] R. Wakikawa, Z. Zhu, and L. Zhang, "Global Home Agent to Home Agent (GHAHA) Protocol Specification", July 2009. (Cited on page 188.)

[232] Zhenkai Zhu, R. Wakikawa, and Lixia Zhang, "SAIL: A scalable approach for wide-area IP mobility", *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, pp. 367 – 372, april 2011. (Cited on page 189.)

[233] H. Chan, F. Xia, J. Xiang, and H. Ahmed, "Distributed Local Mobility Anchors (DLMA)", March 2010. (Cited on page 189.)

# Implementation Aspects

## A.1 Access Selection Enabler

### A.1.1 PCRF-ASE Interface

The following procedures are supported:

1. Report Request - ASE queries the PCRF a report on the active services of a current subscriber or the availability of resources in an access network in the following instances:
    (a) At the beginning of communications between the ASE and the UE via the S14 reference point
    (b) After receiving a request from the UE via the S14 reference point for access discovery and selection
    (c) At the receipt from a notification of change in the parameters from the UE via the S14 reference point (e.g. availability of access networks, locations etc.)
    (d) When for an access network selection decision based on internal triggers (i.e. policies), a report on the active sessions of the subscriber or the availability of resources in the access networks is needed
2. Report Subscription - The ASE shall request the PCRF, subscription to bearer level reports in the following instances:
    (a) At the beginning of communications between the ASE and the UE via the S14 reference point
    (b) At the receipt from a communication including change in the parameters from the UE via the S14 reference point (e.g. availability of access networks, location etc.)
    (c) Upon internal triggers (i.e. policies)
3. Report Notification - The PCRF shall notify with reports the ASE upon modifications in the subscriber sessions or the available resources of access networks if the ASE has previously requested subscription to such modifications.

4. Access Re-selection Request - The PCRF sends an access re-selection request to the ASE after a session authorization request was received from other PCC functionality (e.g. the AF, PCEF, BBERF) and the access network to which the UE is connected does not have enough resources for the session to be accepted but other access networks are available and have resources.

   NOTE: The PCRF may send access re-selection request after release of a session and internal triggers require so (e.g. subscriber profile, policies etc.)

   If an error is return as result code, the PCRF shall continue the session establishment rejection as specified in the PCC procedures for it.

   If the access re-selection is possible the ASE may include the Target-Access-Network AVP in the response and the Handover-Type AVP to indicate if it is a final handover or a volatile handover. In case the handover type is volatile, the Expiration-Time AVP has to be included to the interval of time for the PCRF to wait for handover completion event. The ASE may also indicate in the Report-Triggers AVP that it wants to be notified upon handover completion.

   The PCRF shall behave in this case as it had received a Handover Indication request.

   If the ASE foresees that it can not provide the handover indication instantly (e.g. because communication with the UE via S14 is involved) it shall reply to the PCRF with a response including the experimental result code DIAMETER_SUCCESS_AR_PROCESSING (2XXX).

### A.1.2 Procedures and Data Flows

## A.2 Core Path Adaptation Enabler

### A.2.1 Procedures and Data Flows

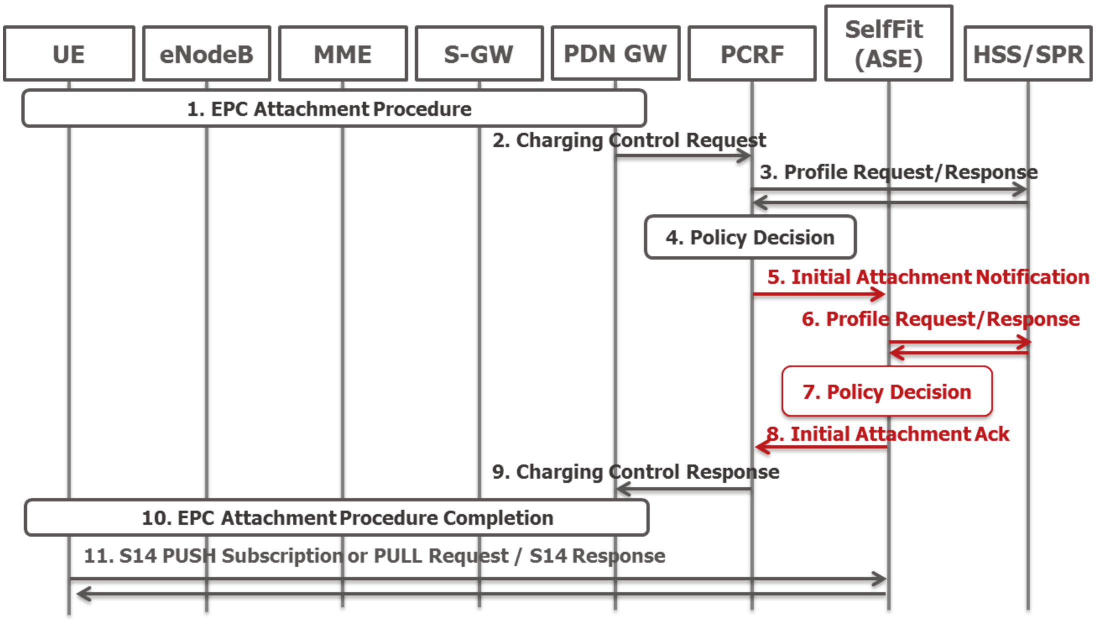


Figure 5: Access Selection Enabler - Initial Attachment with PCRF notification

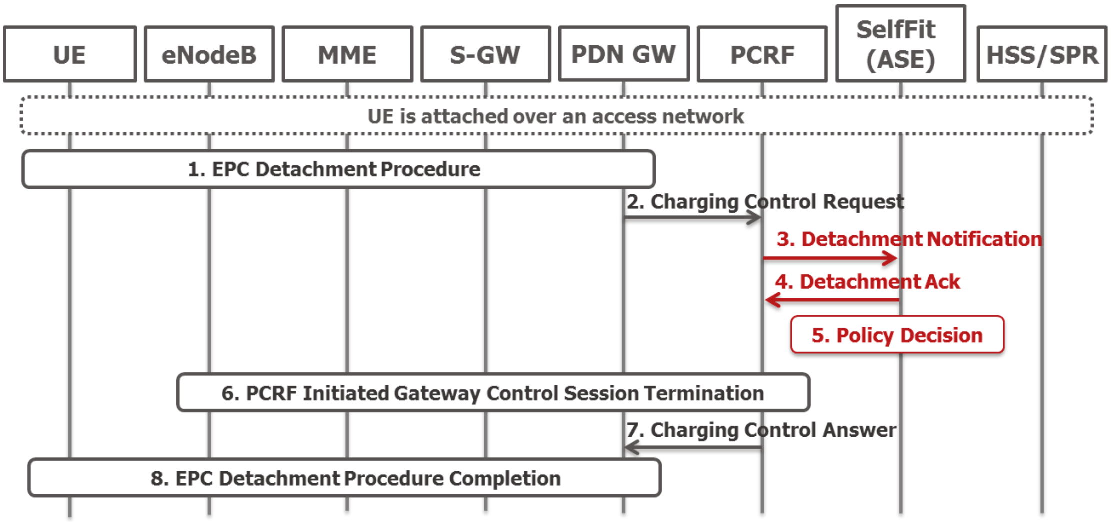


Figure 6: Access Selection Enabler - Detachment with PCRF notification

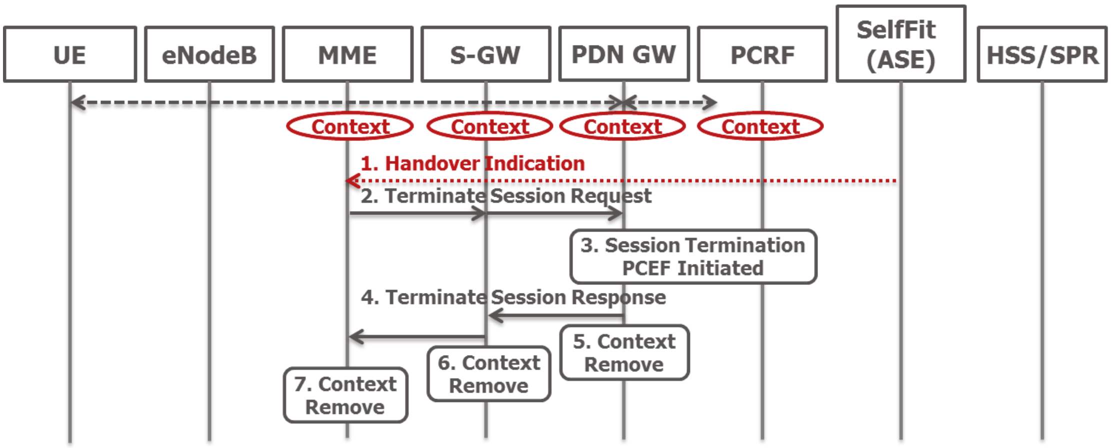


Figure 7: Access Selection Enabler - Volatile Shallow Context Termination

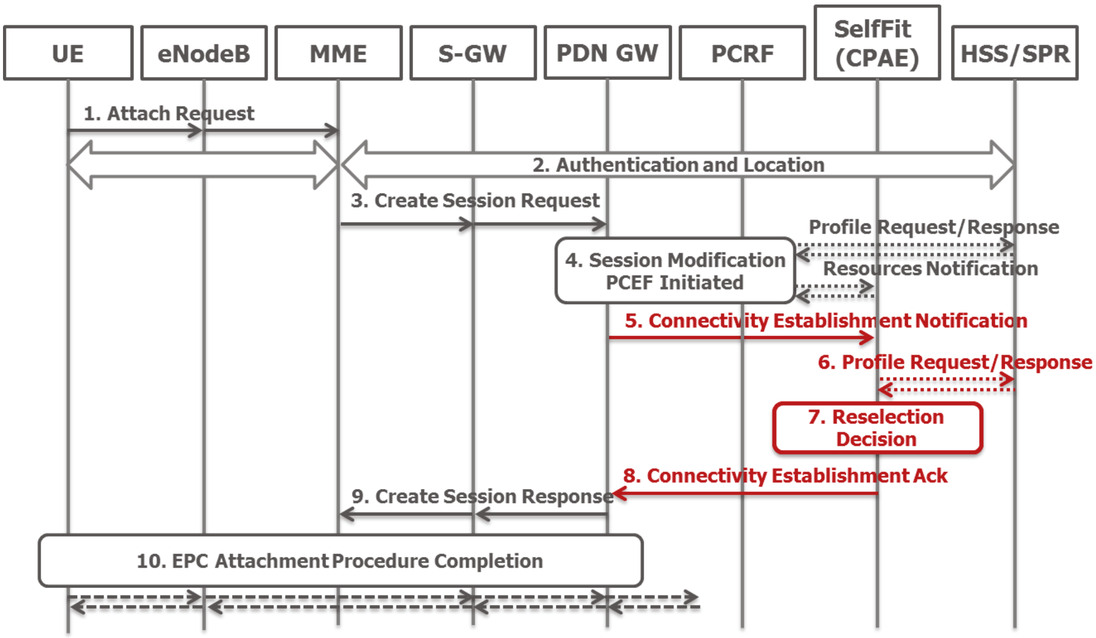


Figure 8: Core Path Adaptation Enabler - Initial Attachment

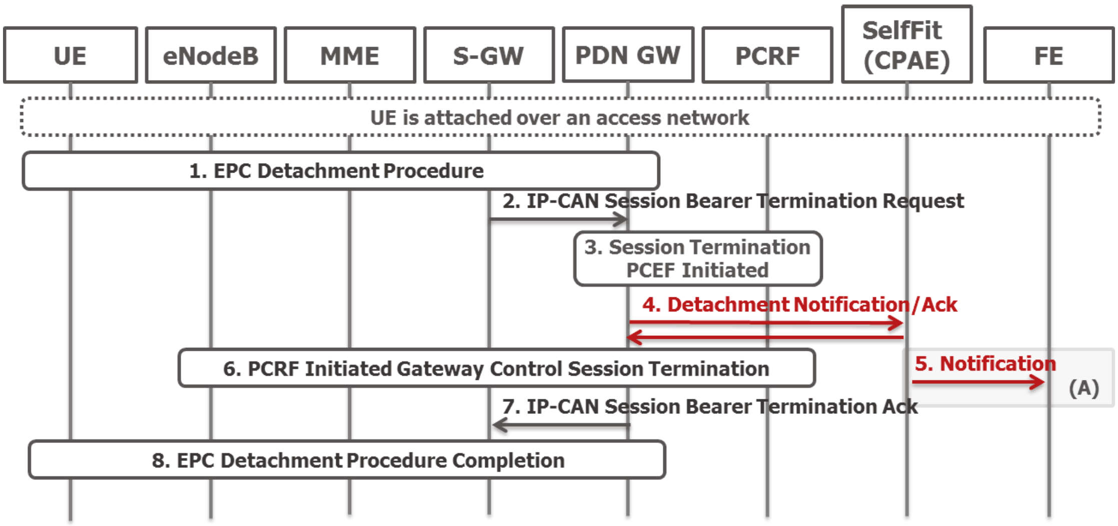


Figure 9: Core Path Adaptation Enabler - Network Detachment

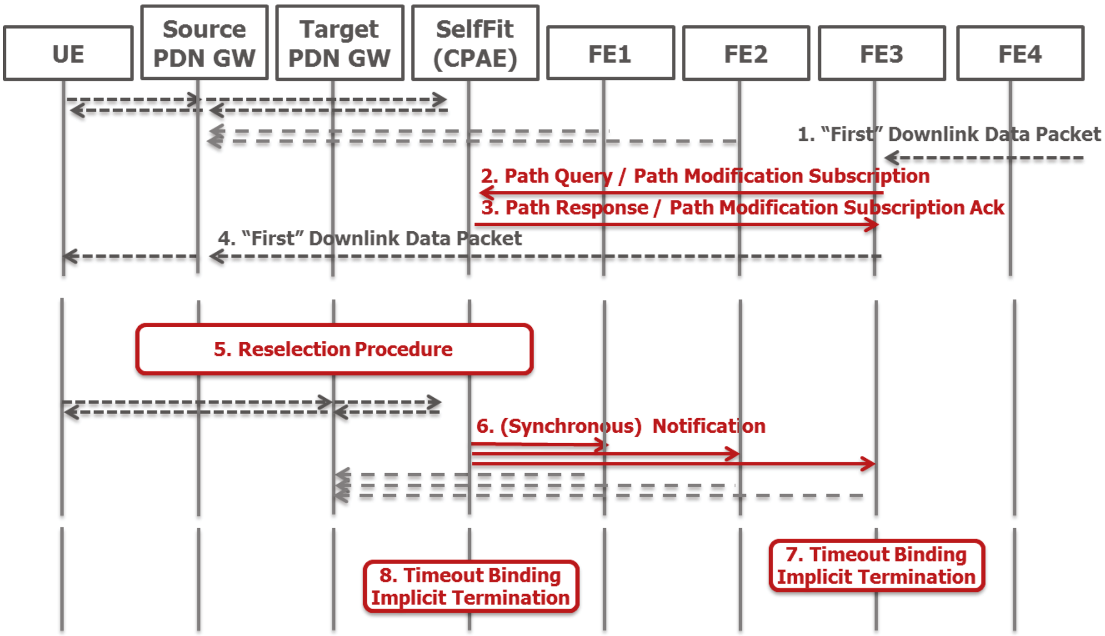


Figure 10: Core Path Adaptation Enabler - Synchronous Adaptation

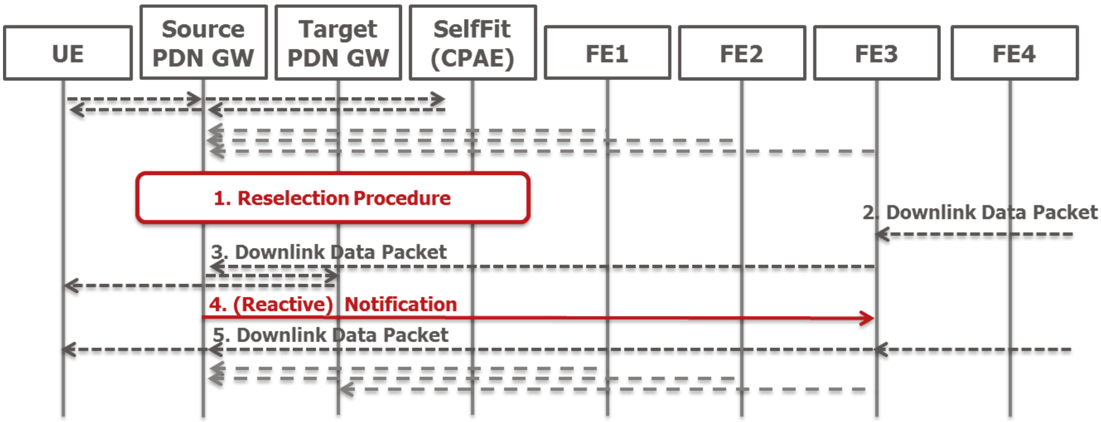


Figure 11: Core Path Adaptation Enabler - Reactive Adaptation

## B.3 Evaluation Metrics

- **Resource Based Access Network Selection** - A solution supports better resource based access network selection when its decisions and enforcement are made as to handover to an access network to which the connectivity service has the available resources to run seamlessly. The metric spans from:
  - **1** - the solution does not consider any resource information. Additionally, it could not be extended unless the solution is completely redesigned
  - **2** - the solution does not consider any resource information. Additionally, it could not be extended without large modifications to the overall architecture
  - **3** - the solution does not consider any resource information. Additionally, it could not be easily extended using the existing functionality
  - **4** - the solution does not consider any resource information. Additionally, it could not be extended using only the existing functionality.
  - **5** - the solution does not consider resource information, but it can be easily extended.
  - **6** - the solution considers resource information only for part of the decision (not all cases)
  - **7** - the solution considers resource information for all the cases as part of the decision
  - **8** - the solution considers resource information as part of the decision and is able to execute a preparation phase of a handover when needed
  - **9** - the solution considers the resource information as part of the decision, is able to make decisions triggered by resource changes and is able to execute a preparation of a handover when needed
- **Data Path Stretch** - A solution supports a better data path stretch when it does require less redirection of data traffic due to mobility support in the core network. The metric does not consider the limitations in forwarding due to the underlying network layer i.e. IP forwarding limitations and sub-optimal IP routing protocols. It spans between:

  - **1** - the solution redirects all the data traffic to a centralized network anchor, regardless of the location of the mobile device. It can not be modified unless the solution is completely redesigned.
  - **2** - the solution redirects all the data traffic to a centralized network anchor regardless of the location of the mobile device. It can be extended with large modifications to the overall architecture.
  - **3** - the solution redirects the data traffic to a centralized anchor selected based on the initial location of the device. The anchor can not be changed without high modifications to the system.
  - **4** - the solution redirects the data traffic to a centralized anchor selected based on the initial location of the device. The anchor can be changed through replicating state between a large number of anchors.
  - **5** - the solution redirects the data traffic to more centralized anchors selected based on the initial location depending on the specific data traffic. The anchors can be changed through replicating state between a large number of anchors.
  - **6** - the solution requires redirection of some of the data traffic, with no selection mechanism
  - **7** - the solution requires redirection of some data traffic, which may be selected
  - **8** - the solution requires redirection of the data traffic in the correspondent node domain. The selection of the anchor in the correspondent node domain is independent of the anchor in the home domain.
  - **9** - the solution requires no redirection. The data path stretch is not affected by the mobility support.

- **Service Adaptation Support** - A solution is better supporting service adaptation when it is able to adapt the resources required per specific data flow, triggered and according to modified network conditions and according to subscriber and application requirements. The metric spans between:
  - **1** - the solution does not allow for any automatic adaptation of the resources. A new resource reservation has to be executed through the termination of the initial service and the establishment of a new one.
  - **3** - the solution does not allow for any automatic adaptation of resources. Adaptation of resources has to be triggered externally. The solution only executing the newly specified rules independent of subscriber or data flow information.

  - **5** - the solution does not allow for any automatic adaptation of resources. Adaptation has to be triggered externally. The solution maintains data flow information in the network nodes, however no subscriber or application specific decisions can be taken.
  - **7** - the solution does not allow for any automatic adaptation of resources. Adaptation has to be triggered externally. The solution maintains subscriber and data flow information as well as network location information.
  - **9** - the solution allows for automatic adaptation of the resources for the specific data flows of the subscriber triggered by and based on the network conditions modifications.

- **Service Profile Integration** - A solution is better when it uses all the information available in the mobile system related for the specific user. It spans between:
  - **1** - no network subscription profile information is used. It can not be modified unless the solution is completely redesigned.
  - **2** - no network subscription profile information is used. It can be modified, however it requires high modifications to the system
  - **3** - an independent subscription profile is used for the specific adaptation. The connection to the other subscription information available in the network or to the information in the mobile node requires extensive communication mechanisms across technologies
  - **4** - an independent subscription profile is used for the specific adaptation. Connection with the other subscription information or to the mobile node requires extensions to the architecture which may extend the communication across technologies
  - **5** - an independent subscription profile is used for the specific adaptation. Connection with the other subscription information in the network or the mobile node requires extensions to the architecture
  - **6** - an independent subscription profile is used for the specific adaptation, including usage of the mobile node information. Connection with the other network subscription information requires extensions to the architecture
  - **7** - subscription profile in the network is used in a very limited form or only for part of the adaptation
  - **8** - subscription profile information available is used for the specific adaptation decisions. No specific extensions of the subscription information are considered in order to make adaptation decisions better.

  - **9** - the subscription profile information available in the network and in the mobile device is used for the specific adaptation decision. Extensions which enable a better decisions are considered.

- **Required Signaling** - A solution is considered better when it requires less messages to be exchanged over the network for the specific adaptation. Especially messages over the wireless link are considered detrimental. This metric spans between:
  - **1** - the solution requires signaling to all the network nodes
  - **2** - the solution requires extensive signaling to network nodes i.e. all base stations in an area, all network nodes part of the data path etc.
  - **3** - the solution requires extensive signaling. As the solution provides only a template of the adaptation, multiple times the adaptation has to be signaled. All correspondent nodes, if any, have to be signaled.
  - **4** - the solution requires extensive signaling a distributed set of nodes
  - **5** - the solution requires point-to-point signaling of all the entities involved in the adaptation, including signaling over the radio network
  - **6** - the solution requires the signaling only on a limited part of components, mainly only in the network side with less signaling over the wireless network
  - **7** - the solution requires the signaling only on a limited part of components, only in the network side or only to the mobile device
  - **8** - the solution requires the signaling only on a limited part of components, only in the network side or only to the mobile device or no re-execution of the adaptation is required
  - **9** - very limited signaling/no signaling is required for the specific adaptation

- **Required Policy Decisions** - A solution is considered better when it requires less policy decisions to be taken. If to obtain the require result, multiple policy decision rounds have to be executed, it is considered highly detrimental from this perspective. This metric spans between:
  - **1** - the solution requires a continuous number of policy decisions for a specific adaptation
  - **2** - the solution's success in adaptation is independent of the actual policy decisions. Thus, a high number of policy decisions are required, as the decisions and the adaptations are in orthogonal planes

  - **3** - the solution's success in adaptation is loosely coupled to the actual policy decisions. Thus, a high number of policy decisions are required
  - **4** - the solution's success in adaptation is better coupled to the actual policy decisions. A large number of policy decisions are required
  - **5** - policy decisions are addressing the complete communication of the mobile node, transparent to the actual data traffic and its characteristics or the subscription profile or the decisions are covering only part of the available network resources which can be adapted. Multiple decisions may be required
  - **6** - policy decisions are considering the complete mobile device communication. Multiple decisions may be required as only a limited number of parameters are considered
  - **7** - the policy decisions include the mobility related available parameters, resulting an unlikely repetition of the procedures
  - **8** - the policy decisions include the mobility and resource related available parameters. Repetition of the decision is an exception case
  - **9** - a single policy decision is considered at the beginning of the specific procedures, the adaptations are completely automatic i.e. the policy decisions are stored

- **Adaptation Duration** - The adaptation duration is measured between the moment an adaptation is required and the moment when the data path is already adapted. The metric spans between:

  - **1** - a large number of nodes have to be notified on the adaptation or the adaptation uses a slow notification mechanism
  - **2** - the adaptation requires that the specific patters in the data traffic of the subscriber changes i.e. the previous active communication sessions are terminated
  - **3** - a large number of nodes have to be notified. The notification has to be transmitted end-to-end
  - **4** - a large number of nodes have to be notified. The notification has to reach at least a specific node in the network such as intermediary service platform which has to take a distinct policy decision
  - **5** - the adaptation procedure is executed over a limited number of nodes related only to the core and forwarding network and mobility control
  - **6** - the adaptation procedure involves a limited number of nodes from either the core network, the forwarding network or the mobility control

  - **7** - the adaptation is constraint to a single operator domain
  - **8** - the adaptation is included as part of other existing procedures such as network attachment and thus no other delay is added
  - **9** - the adaptation duration is shorter than the other existing procedures

- **Solution Flexibility** - A solution is better when it can adapt easier to topology changes through reselection of entities serving a mobile node. It spans between:
  - **1** - the solution is not able to adapt to any change in network topology, thus affecting the communication of the subscribers
  - **2** - the solution is not able to change any network node which ensures the communication of the subscriber without affecting its communication
  - **3** - the solution does not include any means to flexibly select a specific node
  - **4** - the solution does not include any means to flexibly select a node outside a specific group e.g. between some nodes located at the same geographic location
  - **5** - the flexibility is ensured through the mobile node or from forwarding data plane triggers of adaptation procedures
  - **6** - the core network can ensure the flexibility of the data path through some architectural additions
  - **7** - data path can be flexibly adapted to the new network conditions through an extension mechanism to the specific technology
  - **8** - data path is automatically adapted
  - **9** - data path is automatically adapted in a very fast manner, through preparation procedures

- **Reliability** - A solution is better when interruption in communication due to node failure is faster mitigated. The metric spans between:
  - **1** - if any network node fails, the communication does not recover
  - **2** - if a communication node on the data or control plane of the subscriber fails, the communication does not recover
  - **3** - if a specific node fails on the data path the communication can not recover
  - **4** - if a node from a group of nodes from the data plane fails, than the communication can not recover

  - **5** - if a node from the data or control plane fails, the communication can not recover
  - **6** - if a node from the forwarding plane fails the communication can be recovered through architectural extensions notifying the control plane
  - **7** - if a forwarding plane node fails, the subscriber can automatically recover communication through the already available information
  - **8** - if a forwarding plane node fails, the data path can be automatically recovered
  - **9** - if a large number of network nodes fail, the data path can be recovered

- **MN Impact** - this metric evaluates the impact on the mobile devices i.e. the changes that have to be brought to a plain mobile node in order to function according to the solution requirements. It spans between:
  - **1** - large modifications have to be brought to the mobile devices or new architecture of mobile devices have to be considered
  - **2** - a large set of modifications have to be brought in the network stack which have to be included in the kernel or the firmware of the mobile node
  - **3** - a set of modifications have to be brought to the network stack which have to be included in the kernel or the firmware of the mobile node
  - **4** - a minimal set of modifications have to be included in the network stack which have to be included in the kernel. The current network can be used in parallel
  - **5** - a control mechanism of the network stack in the mobile devices has to be included orchestrating the network level operations e.g. changes of routes etc.
  - **6** - a set of functions should be brought completing the network stack i.e. for each application type requiring the adaptation. They can run in user space, as separate applications, adapting to the network stack changes
  - **7** - the mechanism comes with the existing network stack, in order to ensure the flexibility an external control suffices
  - **8** - the mechanism relies on existing functionality
  - **9** - the solution is transparent to the mobile node

- **CN Impact** - this metric evaluates the impact on the correspondent nodes, service platforms or correspondent node's operator domain i.e. the changes that have to be brought to the corresponding entities in order to sustain the solution. It spans between:

  - **1** - large modifications have to be brought to the correspondent nodes or new architecture of correspondent nodes have to be considered
  - **2** - a large set of modifications have to be brought in the network stack of the correspondent nodes, which have to be included in the kernel or the firmware of the correspondent nodes
  - **3** - a set of modifications have to be brought to the network stack which have to be included in the kernel or the firmware of the mobile node
  - **4** - a minimal set of modifications have to be included in the network stack which have to be included in the kernel of the correspondent nodes or a modification in the correspondent node's domain entities.
  - **5** - a control mechanism of the network stack in the correspondent nodes has to be included orchestrating the network level operations e.g. changes of routes etc.
  - **6** - a set of functions should be brought completing the network stack i.e. for each application type requiring the adaptation. They can run in user space, as separate applications, adapting to the network stack changes.
  - **7** - the mechanism comes with the existing network stack, in order to ensure the flexibility an external control suffices
  - **8** - the mechanism relies on existing functionality
  - **9** - the solution is transparent to the correspondent node

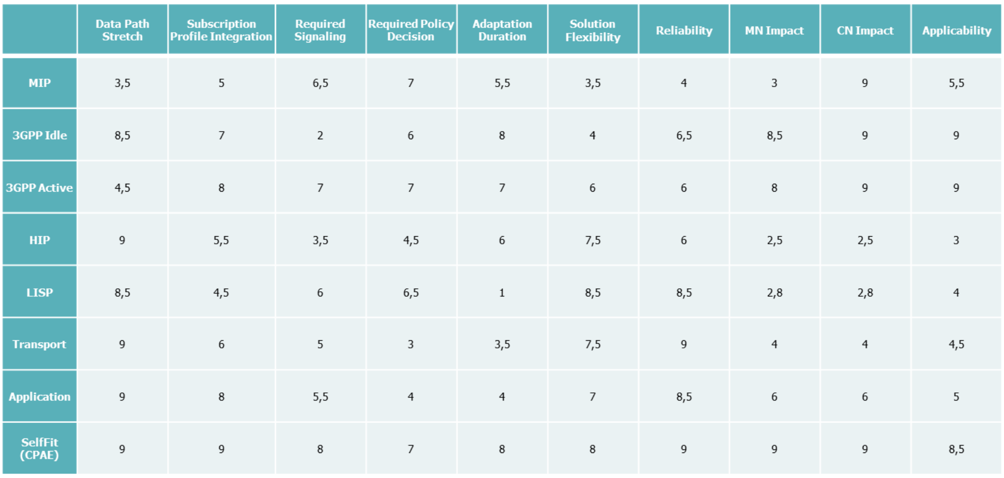
- **Applicability** - this metric evaluates the easiness to add the solution to the current network architecture. It spans between:
  - **1** - a new network architecture is required
  - **2** - a large set of modifications have to be brought to the network including redesign of some nodes
  - **3** - the solution requires replacement or complete modifications of already existing functions
  - **4** - the solution requires introduction of a large set of nodes in all network domains
  - **5** - the solution requires the modification of at least the mobile device and a correspondent node in a different network operator
  - **6** - the solution requires the modification of the mobile node and requires the replacement of a function in its operator domain
  - **7** - the solution requires modifications to the mobile node and additions to functions available in its operator domain

- **8** - the solution is integrated in the system, a reduced number of adaptations are required
- **9** - the solution is the current running system

## B.4 Technology Comparison

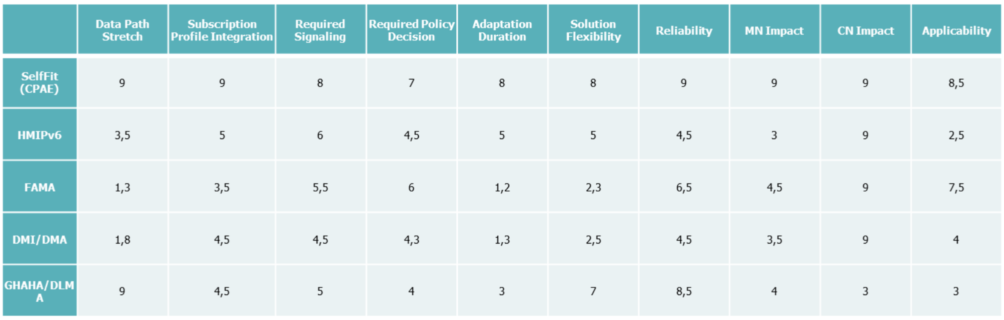
| | Resource Based Access Network Selection | Subscription Profile Integration | Required Signaling | Required Policy Decision | Adaptation Duration | Solution Flexibility | Reliability | MN Impact | CN Impact | Applicability |
|---|---|---|---|---|---|---|---|---|---|---|
| **MIH** | 2 | 2 | 6,5 | 5 | 5,5 | 3 | 5 | 3 | 9 | 4 |
| **P1900.4** | 2,5 | 4 | 6 | 5 | 5 | 4 | 5 | 3,5 | 9 | 4 |
| **Intra-3GPP** | 1 | 7 | 7,5 | 5 | 6 | 7 | 7 | 7 | 9 | 9 |
| **ANDSF** | 3 | 4 | 3 | 2 | 7 | 5 | 7 | 5 | 9 | 8 |
| **SelfFit(ASE)** | 9 | 9 | 8 | 8 | 9 | 9 | 7 | 6 | 9 | 7,5 |

Table 2: Access Selection Technologies Qualitative Comparison Matrix



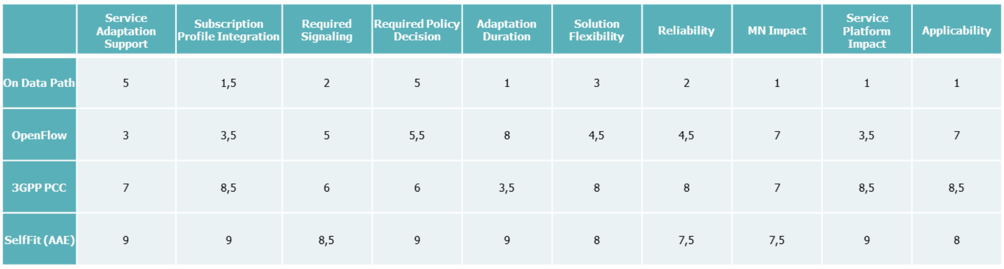
| | Data Path Stretch | Subscription Profile Integration | Required Signaling | Required Policy Decision | Adaptation Duration | Solution Flexibility | Reliability | MN Impact | CN Impact | Applicability |
|---|---|---|---|---|---|---|---|---|---|---|
| **MIP** | 3,5 | 5 | 6,5 | 7 | 5,5 | 3,5 | 4 | 3 | 9 | 5,5 |
| **3GPP Idle** | 8,5 | 7 | 2 | 6 | 8 | 4 | 6,5 | 8,5 | 9 | 9 |
| **3GPP Active** | 4,5 | 8 | 7 | 7 | 7 | 6 | 6 | 8 | 9 | 9 |
| **HIP** | 9 | 5,5 | 3,5 | 4,5 | 6 | 7,5 | 6 | 2,5 | 2,5 | 3 |
| **LISP** | 8,5 | 4,5 | 6 | 6,5 | 1 | 8,5 | 8,5 | 2,8 | 2,8 | 4 |
| **Transport** | 9 | 6 | 5 | 3 | 3,5 | 7,5 | 9 | 4 | 4 | 4,5 |
| **Application** | 9 | 8 | 5,5 | 4 | 4 | 7 | 8,5 | 6 | 6 | 5 |
| **SelfFit (CPAE)** | 9 | 9 | 8 | 7 | 8 | 8 | 9 | 9 | 9 | 8,5 |

Table 3: Core Network Path Adaptation Technologies Qualitative Comparison Matrix



| | Data Path Stretch | Subscription Profile Integration | Required Signaling | Required Policy Decision | Adaptation Duration | Solution Flexibility | Reliability | MN Impact | CN Impact | Applicability |
|---|---|---|---|---|---|---|---|---|---|---|
| SelfFit (CPAE) | 9 | 9 | 8 | 7 | 8 | 8 | 9 | 9 | 9 | 8,5 |
| HMIPv6 | 3,5 | 5 | 6 | 4,5 | 5 | 5 | 4,5 | 3 | 9 | 2,5 |
| FAMA | 1,3 | 3,5 | 5,5 | 6 | 1,2 | 2,3 | 6,5 | 4,5 | 9 | 7,5 |
| DMI/DMA | 1,8 | 4,5 | 4,5 | 4,3 | 1,3 | 2,5 | 4,5 | 3,5 | 9 | 4 |
| GHAHA/DLMA | 9 | 4,5 | 5 | 4 | 3 | 7 | 8,5 | 4 | 3 | 3 |

Table 4: Distributed Mobility Management (DMM) Technologies Qualitative Comparison Matrix



| | Service Adaptation Support | Subscription Profile Integration | Required Signaling | Required Policy Decision | Adaptation Duration | Solution Flexibility | Reliability | MN Impact | Service Platform Impact | Applicability |
|---|---|---|---|---|---|---|---|---|---|---|
| On Data Path | 5 | 1,5 | 2 | 5 | 1 | 3 | 2 | 1 | 1 | 1 |
| OpenFlow | 3 | 3,5 | 5 | 5,5 | 8 | 4,5 | 4,5 | 7 | 3,5 | 7 |
| 3GPP PCC | 7 | 8,5 | 6 | 6 | 3,5 | 8 | 8 | 7 | 8,5 | 8,5 |
| SelfFit (AAE) | 9 | 9 | 8,5 | 9 | 9 | 8 | 7,5 | 7,5 | 9 | 8 |

Table 5: Application Adaptation Qualitative Comparison Matrix