Contextual Connectivity in Multi-Access Architectures

Siwar Ben Hadj Said

THÈSE / Télécom Bretagne
sous le sceau de l'Université européenne de Bretagne

pour obtenir le grade de Docteur de Télécom Bretagne
En accréditation conjointe avec l'Ecole doctorale Matisse
Mention : Informatique

présentée par
Siwar Ben Hadj Said
préparée dans le département Réseaux, sécurité et multimédia
Laboratoire Irisa

Contextual Connectivity in Multi-Access Architectures

Thèse soutenue le 15 décembre 2014
Devant le jury composé de :

César Viho
Professeur, Université de Rennes 1 / président

Abderrahim Benslimane
Professeur, Université d'Avignon / rapporteur

Sami Tabbane
Professeur, SUP'COM – Ecole supérieure des communications de Tunis / rapporteur

Thierry Turletti
Directeur de recherche, Inria - Sophia-Antipolis / examinateur

Farouk Kamoun
Professeur, Ecole Nationale des Sciences de l'Informatique - Tunisie / examinateur

Karine Guillouard
Ingénieur R&D, FT/Orange Labs – Cesson Sévigné / examinatrice

Jean-Marie Bonnin
Professeur, Télécom Bretagne / Directeur de thèse

N° d'ordre : 2014telb0341

Under the seal of European University of Brittany

Télécom Bretagne

In joint accreditation with University of Rennes 1

Ecole Doctorale – MATISSE

Contextual Connectivity in Multi-Access Architectures

PhD Thesis

In Computer Sciences

Presented by **Siwar Ben Hadj Said**

Department : Networks, Security and Multimedia (RSM)

Laboratory : Orange Labs

Thesis Advisor : Jean-Marie Bonnin

Defended in 15 December 2014

Jury :

Prof. César Viho, Full Professor, Université de Rennes 1, Rennes, France (President)
Prof. Abderrahim Benslimane, Full Professor, Université d'Avignon, Avignon, France (Reviewer)
Prof. Sami Tabbane, Full Professor, Sup'Com, Tunis, Tunisie (Reviewer)
Dr. Thierry Turletti, Research Scientist, INRIA Sophia-Antipolis, Nice, France (Examiner)
Prof. Farouk Kamoun, Professor Emeritus, ENSI, Tunis, Tunisie (Examiner)
Prof. Jean-Marie Bonnin, Full Professor, Telecom Bretagne, Rennes, France (Thesis Advisor)
Dr. Karine Guillouard, Senior Research Engineer, Orange Labs, Rennes, France (Supervisor)

N° d'ordre : 2014telb0341

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En accréditation conjointe avec l'Ecole Doctorale Matisse

Ecole Doctorale – MATISSE

Contextual Connectivity in Multi-Access Architectures

Thèse de Doctorat

Mention : Informatique

Présentée par **Siwar Ben Hadj Said**

Département : Réseaux, Sécurité et Multimédia (RSM)

Laboratoire : Orange Labs

Directeur de thèse : Jean-Marie Bonnin

Soutenue le 15 décembre 2014

Jury :

M. César Viho, Professeur, Université de Rennes 1, Rennes, France (Président)
M. Abderrahim Benslimane, Professeur, Université d'Avignon, Avignon, France (Rapporteur)
M. Sami Tabbane, Professeur, Sup'Com, Tunis, Tunisie (Rapporteur)
M. Thierry Turletti, Chercheur, INRIA Sophia-Antipolis, Nice, France (Examinateur)
M. Farouk Kamoun, Professeur Emérite, ENSI, Tunis, Tunisie (Examinateur)
M. Jean-Marie Bonnin, Professeur, Telecom Bretagne, Rennes, France (Directeur de thèse)
Mme. Karine Guillouard, Ingénieur, FT/Orange Labs, Rennes, France (Encadrant)

Acknowledgements

This thesis work would not have been possible without the inspiration and encouragement of many people.

I would like first of all to thank the jury members, Prof. Abderrahim Benslimane, Prof. Sami Tabbane, Dr. Thierry Turletti, Prof. Farouk Kamoun and Prof. Cesar Viho for their time spent in reviewing my dissertation.

I am truly grateful to my PhD advisor, Prof. Jean-Marie Bonnin for his help and guidance during my PhD studies. I would like to express my sincere gratitude to my PhD supervisor, Dr. Karine Guillouard, for her remarkable supervision and invaluable feedback. Her sincere devotion to her research and supporting attitude have turned my PhD experience with her fruitful and worthwhile.

I would like to thank all my colleagues at Orange Labs. I am particularly appreciative of my colleagues Meryem Ouzzif and Hanane Alabdellaouy, for their support, warm company and encouragement during the tough time of this thesis.

A special gratitude and love goes to my family and parents in-law for their support. I thank also my parents for their abiding love. Finally I want to express my deepest love and thanks to my loving husband, Monem, for his constant support, understanding, and encouragement during my research journey.

Abstract

Managing the network connectivity in multi-access architectures becomes a critical issue as these architectures should be able to interwork between heterogeneous technologies and to face the new ecosystem challenges. The 3GPP standards body proposed the multi-access 3GPP system that aims at providing ubiquitous network connectivity. This proposal has many benefits but it brings also a lot of challenges for network operators. In fact, within this system, network mechanisms such as mobility management and security mechanisms are designed to be activated in a systematic manner leading to rising network operating costs. For instance, the location update mechanism is always performed even for static devices. However, network operators face the challenge to host several categories of subscribers such as static subscribers, subscribers with high mobility, subscribers requiring high security level, subscribers satisfied with just low security level, subscribers requiring high/low bandwidth, etc.

This thesis includes three main contributions. In the first contribution, we analyze the current approaches in multi-access 3GPP system. The aim of such a study is to analyze whether the current network connectivity management approaches could face the challenges imposed by the new ecosystem. We started our study by decomposing the network connectivity into several network services. Particularly, our work addresses the mobility and security services. Through network usage scenarios, we showed that security and mobility mechanisms in the access network can be bypassed in certain cases. In addition, we specified a number of requirements for multi-access architectures to face the new ecosystem challenges.

In the second contribution, we concentrate on proposing a Context-Aware Connectivity Management (CACM) module for multi-access architectures. This module is able to manage and interwork between heterogeneous access technologies in an efficient manner. It selects and activates network mechanisms in accordance with the contextual information.

In the third contribution, we propose two concrete applications of the proposed model: (i) adaptive data traffic protection service and (ii) adaptive session continuity service. For the first, we setup a test bed that reproduces the untrusted non-3GPP

access in which we implemented a mechanism that activates/deactivates the encryption and integrity protection mechanisms within the IPsec tunnel according to the flow requirements. For the second application, we proposed an OpenFlow-based control plane for the LTE/EPC architecture to ensure adaptive session continuity service.

Keywords: network connectivity management, fixed and mobile convergence, context-awareness, security, mobility management, computer networking, adaptive networks

Résumé

La gestion de la connectivité réseau dans les architectures multi-accès devient une question primordiale parce que ces architectures devraient être en mesure de faire inter-fonctionner des technologies hétérogènes et de faire face aux défis imposés par le nouveau écosystème. L'organisme de normalisation 3GPP a proposé un système 3GPP multi-accès qui vise à fournir une connectivité réseau omniprésente. Cette proposition a indubitablement des avantages, mais apporte également plusieurs défis aux opérateurs de réseau. En effet, dans ce système, les mécanismes de réseau, telles que les mécanismes de gestion de la mobilité et de sécurité sont conçus pour être activés d'une manière systématique augmentant ainsi le coût d'exploitation du réseau. Par exemple, le mécanisme de la mise à jour de la localisation est toujours activé, même lorsque les terminaux ne sont pas mobiles. Cependant, les opérateurs de réseaux doivent relever le défi d'accueillir plusieurs catégories d'abonnés tels que des abonnés non-mobiles, des abonnés à forte mobilité, des abonnés qui demande un niveau de sécurité élevé, des abonnés qui sont satisfait avec un niveau de sécurité faible, des abonnés demandant une large/faible bande passante, etc.

Cette thèse contient principalement trois contributions. Dans la première contribution, nous analysons des approches actuelles dans le système 3GPP multi-accès. Le but d'une telle étude est d'analyser si les approches actuelles de gestion de la connectivité réseau pourraient faire face aux défis imposés par le nouveau écosystème. Nous avons commencé notre étude par la décomposition de la connectivité réseau en services. En particulier, notre travail porte sur les services de mobilité et de sécurité. Grâce à des scénarios d'usage du réseau, nous avons montré que les mécanismes de mobilité et de sécurité dans le réseau d'accès peuvent être contournés dans certains cas. En outre, nous avons spécifié un certain nombre d'exigences sur les architectures multi-accès pour faire face aux nouveaux défis imposés par l'écosystème.

Dans la deuxième contribution, nous proposons un module de gestion de la connectivité (CACM) pour les architectures multi-accès. Ce module est sensible au contexte des abonnés. Il est capable de gérer et de faire inter-fonctionner les accès hétérogènes d'une manière efficace. Il sélectionne et active les mécanismes du réseau

en conformité avec les informations contextuelles.

Dans la troisième contribution, nous proposons deux applications concrètes du modèle proposé : (i) l'adaptation du service de protection du trafic des donnés et (ii) l'adaptation du service de continuité de session. Pour la première, nous mettons en place un banc d'essai qui reproduit l'accès non-3GPP non sécurisé dans lequel nous avons implémenté un mécanisme qui active/désactive les mécanismes de chiffrement et de protection de l'intégrité dans le tunnel IPsec selon les exigences des flux applicatifs. Pour la deuxième application, nous avons proposé un plan de contrôle basé sur le protocole OpenFlow pour l'architecture LTE/EPC afin d'assurer un service de continuité de session adaptatif.

Mots clés: gestion de la connectivité réseau, convergence fixe et mobile, sensibilité aux contextes, securité, gestion de la mobilité, réseaux informatiques, réseaux adaptatifs

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General Context

Annual Cisco Visual Networking Index Forecast expects an exponential increase in data traffic until 2017 [Cis13]. This significant growth will be driven by new applications (e.g. smart meter, online gaming, multimedia services, etc.), variety of connected devices (e.g. Smart Phones, tablets, sensors, connected cars, smart cities, etc.), and also the ubiquitous of high bandwidth access networks.

A recent study predicts that network operators risk an "end of profit" sometime before mid 2015 [Tel11]. In fact, this study shows that despite the data traffic growth, the total network cost will exceed the total revenue very soon, if nothing is done before. This finding may be explained by the fact that existing access network architectures are ill-equipped to cope with the dual challenge of sustained data demand and falling revenues [Tel11]. Indeed, the cost to build, upgrade or operate the network is becoming too high while network operator' revenues are decreasing exponentially [PMJ13]. In the light of these predictions, network operators are invited to revisit the design and capabilities of their architectures with a twofold objective of reducing expenses and introducing new revenue generating services.

Moreover, network operators state that the coexistence of different architectures raises several problems such as the scalability issue, the functional redundancy, the complexity of network management, and the difficulty to inter-work between heterogeneous access points. For this reason, standard bodies such as BBF and 3GPP are seeking a new architecture that is able to interconnect heterogeneous access networks including fixed access networks (Wimax, WiFi, xDSL, etc.) and mobile access networks (3G, LTE, etc.)[3GP10a] [BBF12b]. As network operators expand their architectures to cover an ever greater proportion of subscribers, the number of connections grows and the OPEX and CAPEX required to maintain this growing network infrastructure increase.

With traditional connectivity management, network mechanisms such as mobility management, QoS control and security mechanisms are designed to be activated in a systematic manner leading to the rise of network operating costs. For instance, location update is always performed even for static devices. However, network operators face the challenge to host several categories of subscribers such as static subscribers, subscribers with high mobility, subscribers with small amount of data, etc.

Traditional connectivity is not aware of several contextual information such as subscriber's location, device type, running application, network status, time of the day, etc. For instance, it does not take into consideration that the mobile subscriber will probably not move during a given session (84% of subscribers are either static or nomadic during a session and 16% of subscribers are mobile users [TRKN09]). In addition, a recent study [PSBD11] shows that a large fraction of subscribers have limited mobility. Moreover, it was established in the same study that the subscriber mobility is predictable. Note that some statistics show that users spend more than 60% of their time at home or at work.

In this context, network operators should investigate new solutions to address the new ecosystem in a cost-efficient manner. In other words, network connectivity should be tailored to the subscriber context, application real needs and network status.

Traditional network connectivity in access networks are pre-configured to have systematic behavior. They are designed to activate the same network mechanisms in all situations and cannot be dynamically adapted, for instance, to emerging situational constraints. This static limitation is due to the absence of two main features, namely *(i) modularity*, and *(ii) context-awareness*.

- *modularity*: network connectivity needs to be decomposed into a given number of network services. Each network service requires the activation of specific network mechanisms. For example, the reachability service is one of the network services and is defined as the ability of the network operator to keep the user reachable for any incoming sessions. In case this service is activated, the network operator should update the user data path whenever the user change the access point. Therefore, the location update, the tunnel update and the IP address allocation mechanisms are all necessary for the reachability service.

- *context-awareness*: network connectivity needs to be aware of any contextual information and adapts its behavior accordingly. The contextual information includes any type of data that can be useful for improving and adjusting the network connectivity. Example of contextual information are user profile, mobility pattern (static/nomadic, low/high mobility), device type, session requirements, network status, etc.

Since network connectivity in access network is not modular and context-aware, network operators often end up activating several network mechanisms, within a given network connectivity, that the subscribers may not need. The following examples underline the need for modularity and context-awareness features:

- The end-user devices which are foreseen to be used in the future have highly augmented capabilities. The capabilities include not only different radio interfaces which enable different wireless connectivity, but also processing and storage capabilities. Indeed, the Annual Cisco VNI Forecast expects that, by 2017, mobile networks will host 8.6 billion handheld devices and 1.7 billion machine-to-machine (M2M) connections (e.g. sensors for medical applications, tracking systems in shipping, GPS systems in cars, etc.) [Cis13]. However, the same network mechanism does not fit all kinds of devices. Therefore, the network connectivity should include several network mechanisms and selects the adequate mechanism for each situation.

- With the variety of applications running over mobile as well as fixed access networks, the requirements in terms of network services (e.g. mobility management, QoS control, and security services) differ strongly. For instance, a static camera that surveys the street and uploads video periodically does not need the mobility management service. In addition, sessions that are already secured (e.g. SSL sessions established between medical sensors and their server, VPN sessions established between the traveling employees' laptops and the corporate intranet, etc.) need neither confidentiality nor integrity protection at the access network. Avoiding the session protection at the access level may bring down the packet process cost within network equipment; leading thereby to maximize resource utilization.

Hence, the modularity and context-awareness of network connectivity worth to be studied in greater depth.

1.2 Objectives and Contributions

This work was carried out within the "Multi Access Convergence Architecture (MCA)" team in "Convergent Multi-services Architecture networks (CMA)" laboratory of Orange Labs. The research topic of laboratory is centered around the access network architectures. The team works on fixed and mobile convergence, mobility management and network function virtualization aspects.

Our ultimate goal is to design an architecture that automatically adapts the network connectivity to the contextual information of each user. Depending on contextual information, this adjustment includes the activation/deactivation of network

mechanisms such as mobility management, QoS management, and security mechanisms. In other words, a given network mechanism is activated only when the user and/or the operator really requires its presence. This architecture should not offer functional redundancy, allow the flexible use of network resources and facilitate the interworking between heterogeneous access points.

The key challenge of this work is to make the network connectivity aware of required contextual information and to adapt its behavior accordingly. For instance, knowing that the launched data flow is already secured such as an IPsec flow, encryption and integrity protection mechanisms are considered redundant at the access network level and should be deactivated for such type of flow.

The contributions of this thesis fall into three primary fields: current approaches analysis, performance optimization and architecture design. Relevant to the first, our work is an attempt to analyze the current approaches in multi-access architectures. The aim of such a study is to analyze whether the current network connectivity management approaches could face the challenges imposed by the new ecosystem. We started our study by decomposing the network connectivity into three groups of network services: namely security, mobility management and QoS control services. Particularly, our work addresses the mobility and security services in multi-access 3GPP systems where the related mechanisms are systematically activated for each subscriber in all situations. Through network usage scenarios, we showed that security and mobility mechanisms in the access network can be bypassed in certain cases. In addition, we specified a number of requirements for the multi-access architectures to face the ecosystem challenges. Through analytical studies, we evaluated the security costs in LTE/EPC architecture, and showed how much network operator can save in terms of signaling, processing and transmission costs when the security mechanisms are deactivated for a given number of subscribers.

Second, relevant to performance optimization, this thesis offers a new connectivity management model that addresses the challenge of providing the user with the adequate network services. In our model, the network connectivity is decomposed into several network services. This effort attempts to fill an important research gap as the majority of the efforts in the literature focus on implementing and activating more and more network mechanisms (e.g. implementing mobility protocols in fixed accesses) and ignore that the systematic activation of these mechanisms for all subscribers increases network operator costs.

Third, relevant to architecture design, we proposed a Context-Aware Connectivity Management (CACM) module that should be introduced in multi-access architectures. This module is able to manage and inter-work between heterogeneous access technologies in an efficient manner. It selects and activates network mechanisms in accordance with the contextual information. The incorporation of context-aware network control functionality within our connectivity management module provides

value-added capability to decide the adequate network mechanisms to be activated in each situation; thereby decreasing the costs for network operators and, at the same time, provides subscriber with an adequate Quality of Experience (QoE). We proposed examples of implementation for two types of network services: (i) Data Traffic Protection and (ii) Session Continuity. For the first, we setup a test bed that reproduces the untrusted non-3GPP access in which we implemented a mechanism that activates/deactivates the encryption and integrity protection mechanisms within the IPsec tunnel according to the flow requirements. For the second, we proposed an OpenFlow-based control plane for the LTE/EPC architecture to manage the mobility services in a flexible manner.

1.3 Outline of the Dissertation

The chapters of this report can be grouped into two parts:

- Part I: includes Chapter 2, 3, and 4. It reviews the current approaches of network connectivity management and analyses its readiness for the new ecosystem.
- Part II: includes Chapter 5, 6, and 7. It proposes a new model for the network connectivity management in multi-access architectures. In addition, it provides examples of implementation of two adaptive network services namely Data Traffic Protection and Session Continuity services.

The report is structured as follows:

Chapter 2 introduces mainly the basic background of our work. After presenting the considered network services and the related network mechanisms, it focuses on reviewing the connectivity management models in different category of access networks.

Chapter 3 presents the new ecosystem that access networks should address. Through several network usage scenarios, we analyze the network connectivity management in multi-access 3GPP systems and underline the main requirements for future access networks. This qualitative analysis shows that a new model of connectivity management is needed to suit future network usages. This chapter ends by outlining the related works.

Chapter 4 is dedicated to analyze the security services provided by the LTE access and evaluate the related signaling, processing and transmission costs. Through several scenarios, we showed that any added security service introduces additional signaling, processing or transmission loads.

Chapter 5 recalls the architecture requirements that were highlighted in previous chapters. Then, it introduces our vision of the connectivity management in multi-

access architectures. After that, we revisit the network usage scenarios of Chapter 3 to evaluate and compare the proposed model to the connectivity management model in multi-access 3GPP systems.

Chapter 6 proposes to demonstrate the feasibility of the proposed connectivity model via a test bed reproducing non-3GPP accesses. This chapter focus on the Data Traffic Protection (DTP) service. First, it proposes a mechanism that adapts DTP service to the application flow requirements in untrusted non-3GPP accesses. This mechanism ensures the DTP service for clear traffic only. Then, a test bed is set up to validate the proposed mechanism and highlight the related implementation challenges.

Chapter 7 proposes a theoretical implementation of adaptive mobility services in LTE/EPC architecture. It proposes a new control plane based on the OpenFlow (OF) protocol that adapts the Session Continuity service according to network status and application requirements. For example, when the current SGW is overloaded, the delay-tolerant data flows such as FTP flows are temporary transferred to another SGW. This reduces the overload on the SGW and ensures, therefore, the session continuity for the delay-sensitive as well as delay-tolerant data flows. The proposed control plane can be extended to ensures the Reachability service.

Finally, Chapter 8 concludes this report and provides an outlook on perspectives and future work.

Chapter 2

Background

2.1 Introduction

Fixed and mobile access networks are increasingly converging towards common IP-based core network. Providing effective network resource management in such complex heterogeneous environment requires unified, adaptive and scalable connectivity management solution to integrate and co-ordinate network mechanisms of different access technologies.

The scope of this chapter is to analyze the network connectivity management in the current access networks and identify the main network services. We define a network service as *"the set of mechanisms that a network operator implements to address a specific subscriber need"*. For instance, protecting the data traffic is considered as a service ensured for the subscriber. Several network mechanisms are implemented and activated to ensure these services. For example, the data traffic protection service is ensured via the encryption and integrity protection mechanisms.

We start by providing a list of the network services that we considered during this research work namely security and mobility services. Then, we describe different access networks while identifying how the security and mobility services are ensured in each access. By studying different types of access networks, we are able to shed light on the question of how similar or different they are in terms of the ensured network services.

2.2 Network services

The network connectivity analysis in several access networks shows that it is basically formed of three groups of network services namely Quality-of-Service (QoS), security and mobility services. Among the exhaustive list of the identified network services, we are mainly interested in security and mobility-related services. In this section,

we give an outline of the considered network services as well as the corresponding network mechanisms.

2.2.1 Security services

RFC 4949 [Shi07] defines a security service as "a processing or communication service that is provided by a system to give a specific kind of protection to system resources". Therefore, the security services may be defined as *"the set of network mechanisms that are activated by the network operator to avoid potential attacks such as user's identity theft, man-in-the-middle, and denial of service"*. Network security services may be implemented by several security mechanisms as shown in Table 2.1. In the following, we address an outline of the main security services.

Security Services	Network Mechanisms
Identification & Authentication	Login-password Identifier-Challenge/Response Identifier-Certificate
Access Control	Profil-based Access Control
Data traffic protection	Ciphering Integrity Protection
Privacy	Temporary Identifier Signaling exchange protection

Table 2.1: Security services and mechanisms.

2.2.1.1 Identification and Authentication

Identification service is crucial as it helps access providers to identify who is using their network. Several kinds of identity exists:

- Subscription identity: is allocated to the subscriber by the access network provider upon signing a subscription contract. It serves to identify subscribers when they request network resources. Based on the subscription identity, the access network provider determines the subscriber profile and identify the services to which he subscribed. The type of the subscription identity that shall be presented to an access network varies according to the authentication mechanism that was implemented in this access. For instance, the login represents the subscription identity in the login-password authentication mechanism.

- Device identity: is a piece of data that identify the device and is specific to vendors.
- Session identity: is a piece of data that is used in access network to identify a session (e.g. HTTP session) or a series of related message exchanges. It is different from the subscription and device identities. The sessions are typically short-lived as they expire after a preset time of inactivity which may be minutes or hours. Therefore, the session identity may become invalid after the session tear down.

The *Authentication* service consists in validating the identity presented by the user. The user must provide evidence to prove its identity to access the network. This proof can be a certificate or a response to a challenge. For security reasons, the authentication should be mutual in the access network, i.e., not only the user's identity is verified by the network but also the network credibility is checked by the user.
There are mainly three kinds of authentication mechanisms: pre-shared key authentication, mutual public key authentication, and tunneled authentication (i.e. first, the security authority is authenticated using its certificate, establishes encrypted channel with the subscriber and then authenticates the subscriber using the pre-shared secret).

2.2.1.2 Access Control

The *Access Control* service prevents either the unauthorized use of the network resources or the use of network resources in unauthorized manner. Therefore, it ensures two main properties:

- Authorized network resource usage: in general, the access network is intended to be used only by authorized subscribers.
- Appropriate network resource usage: once authorized, the subscriber is able to use network resources according to the restrictions provided in his profile and to network resource availability.

The resources can be bandwidth that supports a certain level of QoS (e.g.,latency and jitter), or an IP address that ensures the user reachability. The access control service should include two functions:

- Authorization: is intended to verify whether the subscriber has permission to use network resources.
- Admission control: is intended to control the manner with which the network is used. The purpose of this mechanism is to protect the network from overload

and congestion situations. Therefore, the network resources availability should be checked before accepting user resource request.

2.2.1.3 Data Traffic Protection (DTP)

The *Data Traffic Protection (DTP)* service prevents the disclosure and modification of data packets during their trip from the sender to the receiver. The data traffic protection is ensured with the ciphering and integrity protection mechanisms. The ciphering mechanism make the user communication unintelligible to a third-party. The integrity protection mechanism guarantees that any alteration of the original message by an unauthorized party will be notable at the receiver side.

2.2.1.4 Privacy

The access provider tends to maintain some basic information about subscribers such as subscription identities, devices identities, subscriber locations, etc. From privacy viewpoint, those information should be protected from other parties. In general, access providers should keep confidential the following information:

- Identity confidentiality: is defined as the act of keeping confidential the subscriber's identity. The identity privacy becomes crucial as it prevents identity theft, and therefore user impersonation attacks.
- Location confidentiality: is defined as the act of keeping confidential subscriber's locations. In fact, an intruder may observe and analyze location update messages sent by the subscriber each time he moves to a new access. This could lead to the disclosure of the subscriber location and therefore to privacy violations.

2.2.2 Mobility Services

The ITU-T (International Telecommunication Union - Telecommunication) defines the mobility as "the ability for the user, or other mobile entities, to communicate and access services irrespective of changes of the location or technical environment" [Uni06]. Therefore, we define the mobility service as *"the set of network mechanisms that are activated by the network operator to deliver the same networked services (e.g. voice call, file download, etc.), with their particular personalization, among alternative devices or access technologies"*. The mobility service also means the ability to receive services independently of the user location.

The concept of mobility includes the *seamless continuity of the session*, even when crossing network borders. This means that a process of handover or seamless handover is taking place among different access points. The mobility also includes

the *nomadism* aspect where users exit completely the previous session and logs on in the new access point, starting a session afresh. Moreover, the mobility includes the *reachability* aspect where the network operator maintains the user location up-to-date. The user location may be published in a global database such as DNS server, and it shall be available for serving incoming sessions such incoming VoIP call.

Mobility Services	Network Mechanisms
Session Continuity	Handover, Context Transfer, Data forwarding
Nomadism	Context Transfer
Reachability	Location update, Tunnel update, IP address allocation

Table 2.2: Mobility services and mechanisms.

2.2.2.1 Session Continuity

The *Session Continuity* service is defined as the ability of the network operator to maintain ongoing sessions for a moving object. The *Handover* mechanism is needed to keep the session continuity when the user move from one access point to another. This mechanism updates the user data path inside the access network by updating data tunnels/bearers and temporarily forwarding data traffic between the previous and current access points/gateways. There are two types of Handover: Seamless (i.e. the session continuity is preserved without any impact on the perceived communication quality) and Hard (i.e. preserve the session continuity with some impact of the perceived communication quality such as extra delay). The handover mechanisms may trigger the *Context Transfer* and *Data Forwarding* mechanisms. Forwarding can be for example accomplished by tunnels between network equipment (e.g. Proxy Mobile IP (PMIP) protocol [GLD+08]) or controlled traffic redirection (e.g. OpenFlow [Ope]).

2.2.2.2 Nomadism

The *Nomadism* service is defined as the ability to provide network services irrespective of environment changes of a moving object. When changing the network access point, the user's session is completely stopped and then started again, i.e., the session continuity is not preserved. The network operator may allocate different user identifiers/addresses when the user moves from one access point to another. The *Nomadism* service requires the presence of the user profile at the new access

point. Therefore, a *Context Transfer* mechanism is required to retrieve the user profile from the previous access network or from the home database.

2.2.2.3 Reachability

The *Reachability* service is defined as the ability of the network operator to keep the user reachable for any incoming sessions. The network operator may update the user data path whenever the user change the access point. The *IP address allocation* mechanism is required to enable applications or other devices to reach the subscriber device. The *Tunnel update* mechanism is required to keep the user data path updated. Maintaining the data path for each user even when there is no active data sessions is consuming in terms of device energy and network resources. Therefore, a power saving mode was introduced in Mobile access networks where the network resources are released for inactive users. Generally, when the device enters the power saving mode, it sends less signaling messages such as location update messages. Therefore, the *Location update* mechanism is required to keep the network operator up-to-date with the user location. It includes user registration, location update and paging procedures. First, the user registers his device to the location registry. The network assigns a permanent IP address through which any application can reach the subscriber regardless of its location. Then, the location update procedure is required to be able to find the user location when needed.

2.3 Access Network Architectures

The Access Network (AN) is a set of functions that ensure the data traffic delivery between the user and the corresponding application or network service provider. In the literature, the ANs are classified into Fixed access network (F-AN) and Mobile access network (M-AN) categories. The M-AN differs from the F-AN in that the mobility service is natively ensured in M-AN. Lately, the multi-access network category appeared in order to ensure the convergence between heterogeneous access networks. In this section, we examine the differences and similarities in the current AN architectures. Generally, ANs offering the same network services adhere to the same architectural model. First, a generic model for the AN architecture is given. Then, examples of the different AN categories are provided. Each time, we map the AN architecture to our generic model to identify where and how network services are performed.

2.3.1 Generic Access Network Model

A typical access network is divided into *Control Plane (AN-CP)* and *Transfer Plane (AN-TP)* as shown in Figure 2.1. The AN-CP consists of network entities that are

involved in the network connectivity establishment, control, charging and termination procedures. The AN-TP, by contrast, consists of network entities that process and forward user data traffic. Generally, this separation makes it possible to centralize the control entities while distributing the network entities involved in the data traffic forwarding.

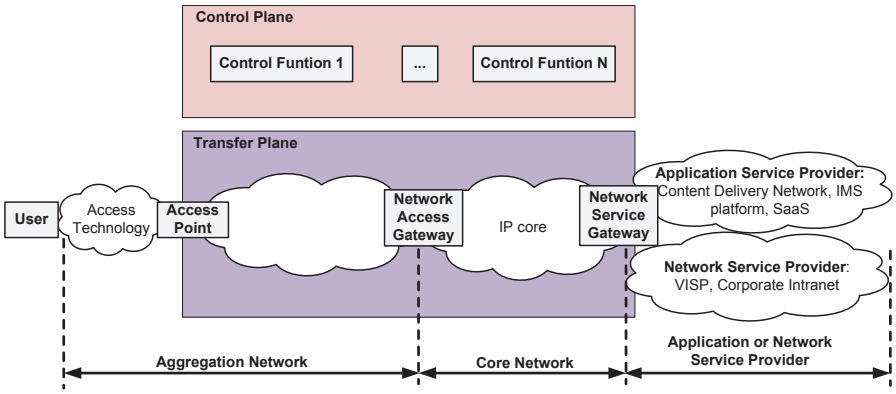


Figure 2.1: Generic Access Network Model

Basically, the AN-TP is divided into aggregation network and core network domains. The *aggregation network* encompasses the elements of the access network from the network interface of the user's device to the Network Access Gateway. This network typically aggregates traffic coming from multiple Access Points and depends on the access technology in use. The *core network* represent the center core of the access network. The switching, routing or tunneling functions are the main functions contained herein to enable the proper distribution of the data traffic.

The Packet Data Networks (PDNs) includes Application Service Providers (ASP) and Network Service Providers (NSPs). The ASP and NSP definitions were given in [BBF03]. The NSP provides access to the Internet. It includes Virtual Internet Service Providers (VISPs) and Corporate Intranets. The NSP authenticates and allocates IP address to their subscribers. The access network provider may act as a wholesale Internet provider especially in case of the VISP case. The ASP provides services to the subscriber application (e.g. gaming, IP telephony, Video on Demand, etc.). It includes the Content Delivery Network (CDN), IMS platform, Software-as-a-Service, etc. The ASP does not handle the IP address allocation.

The AN-TP includes the following logical entities:

- **Access Point (AP)**: represents the first contact of the user's terminal with the access network. It aggregates traffic from the users that are in the some AP coverage.

- **Network Access Gateway (NAGw)**: aggregates the data traffic coming

from multiple APs. This gateway ensures the basic network services such as authentication, access control, IP address allocation and charging.

- **Network Service Gateway (NSGw)**: terminates the AN-TP and gives access to the other Packet Data Networks (PDNs). This gateway may be enabled with the added-value network services such as anchoring user session for mobility purpose, enforcing policies for better QoS and filtering user traffic.

2.3.2 Fixed access networks

2.3.2.1 xDSL access

The basic DSL architecture was specified in TR-059 [BBF03] and later updated by TR-101 [BBF06], TR-134 [BBF12a], and TR-203 [BBF12b]. The xDSL architectural model is depicted in Figure 2.2. The aggregation network in xDSL accesses is based on the Digital Subscriber Line (DSL) access technology. The DSL technology enables high-speed data rate over the existing copper telephone wires that connect subscriber's homes or offices to their access network provider. The core network is based on the IP protocol and consists of routers.

The Residential Gateway (RG) is the equipment that connects the user devices

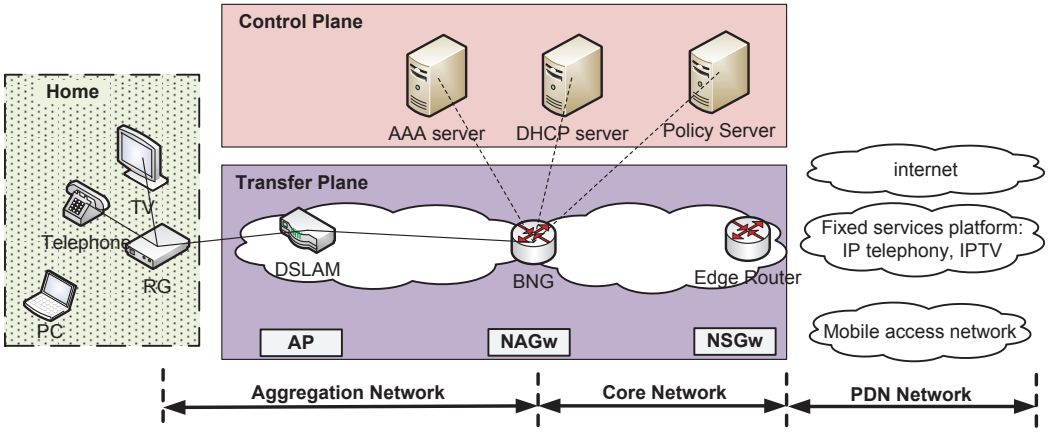


Figure 2.2: xDSL architecture

to the xDSL access. The xDSL access just authenticates the RG and provides an "Always-ON" connection. The subscriber devices in home such as Computers and Smartphones share the same xDSL connection. A local authentication may take place between the subscriber device and the RG.

The xDSL Transfer Plane is made up of two main entities:

- Digital Subscriber Line Access Multiplexer (DSLAM): aggregates the data traffic coming from several subscribers onto a single high-capacity uplink (ATM

or Gigabit Ethernet backhaul) to the xDSL core network. The DSLAM acts as the AP in our generic access model.

- Broadband Network Gateway (BNG): aggregates data traffic issuing from several DSLAMs. The BNG acts as the NAGw entity as it ensures the basic network services such as relaying the authentication exchanges and DHCP requests/replies between the device and the control plane. According to [BBF03], the BNG also performs the QoS mechanisms such as packet classification and scheduling.

- Edge Router: corresponds to our NSGw entity as it gives access to other PDNs. However, the Edge Router provides no added-value services.

The xDSL control plane (xDSL-CP) is responsible for user authentication, access control and IP address allocation. It includes the following logical entities:

- Authentication, Authorization and Accounting (AAA) server: authenticates the subscriber's credentials and validates the users access policies. The exchange between the AAA server and the BNG is based on RADIUS [RWRS00] or Diameter [FALZ12] protocol.

- DHCP server: is responsible for IP address allocation [Dro97].

- Policy Server: was introduced by TR-134 [BBF12a] to provide broadband network services based on policies. This policies may be activated in the BNG dynamically and/or statically.

The subscriber authentication in the xDSL access can be explicit through the use of login/password method. In that case, the credential are stored in the RG. The BNG uses the received Login parameter as the username in the RADIUS or Diameter authentication request. The subscriber authentication can be implicit based on the Line ID wherein the DSLAM inserts this identifier in the user DHCP request message (i.e. the DHCP Option 82 information). Therefore, the BNG uses the Line ID as the username in the authentication request.

2.3.2.2 WLAN access

Nowadays, the Wi-Fi access is becoming more and more ubiquitous. In fact, fixed broadband operators encourage their subscribers to share their private Wi-Fi (i.e. community Wi-Fi schemes). In addition, the number of users demanding Wi-Fi access in shops, cafes, and hotels is increasing. Moreover, Wireless Internet Service Providers (WISPs) are new types of operator that grants Internet access and location-based services using public Hotspots. Therefore, the Wi-Fi coverage is overlapping and the Wi-Fi access points are owned by various operators.

Where the IEEE standardization bodies focus on specifying the set of medium access control (MAC) and physical layer (PHY), they did not extend the standardization work to the WLAN organic architecture. This gives WiFi service providers the freedom to design the best architecture for their access network. We analyzed several WLAN architecture proposals [Cis10], [AL12] [Iri13] and come up with the most recurrent architectural model. The main difference between these proposals consists in the authentication and access control mechanisms. First, we describe the WLAN architecture with an open access (see Figure 2.3). After that, we describe the most implemented authentication and access control mechanisms in WLAN accesses.

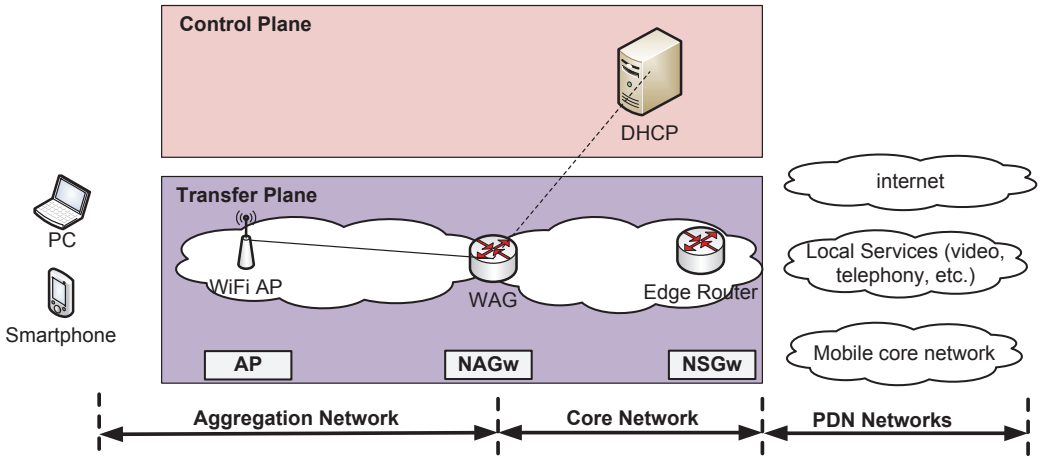


Figure 2.3: WLAN architecture

The WLAN Transport Plane (WLAN-TP) includes:

- Wi-Fi Access Point (WAP): represents the first point of contact of the user with the WLAN network. It represents the boundary between the wireless coverage and the wired backbone. Depending on security configuration, WAP may behave differently. For instance, an open WAP with no security configuration serves any user within its coverage area as long as the maximum number of users that it can serve simultaneously is not reached. In protected Wi-Fi access networks, the WAP acts as a gatekeeper. In this case, it opens the door only for data traffic related to authenticated and authorized users. Otherwise, the WAP rejects user association request as long as the authentication procedure fails. Two possible security configurations for a protected Wi-Fi access will be described in the following.

- WLAN Access Gateway (WAG): represents the first hop router in WLAN networks. It aggregates traffic issuing from multiple WAPs. The WAG cor-

responds to our NAGw as it provides the basic connectivity functions such as packet routing, IP address allocation and access control. It also relays the authentication exchanges between the user device and the authentication entities.

- Edge Router: corresponds to our NSGw entity as it gives access to other PDNs. However, the Edge Router provides no added-value services.

Generally, the user association in WLAN access includes the following main phases:

Discovery: represents the first phase in the association procedure in WLAN. Upon exchanging the security capabilities, the user device and the WAP agree on security parameters such as the authentication method to be used, the confidentiality and integrity protection algorithms, and the cryptographic key management approach. At the end of this phase, the user device is just associated to the WAP and needs to be authenticated and authorized before getting the complete access to WLAN network.
Authentication: validates the user identity before authorizing access. The web portal-based access and the IEEE 802.1x are the most familiar authentication mechanisms.
Key Management: consists in activating the confidentiality and integrity protection services at the WLAN-TP and more precisely at the Radio level. During this phase, a 4-way handshake between the user device and the WAP takes place. The aim of this exchange is to establish fresh security keys to protect link-layer frames.
Data Transfer: represents the last phase of the association procedure. During this phase, the user device is allowed to send data traffic through the corresponding WAP. In case the confidentiality and integrity protection services are activated, the user device and the WAP run the confidentiality and integrity protection algorithms during data traffic delivery.

To enable seamless user authentication and strong access control in WLAN access, the IEEE 802.11i amendment [IEE04] proposed the IEEE 802.1x protocol. On the other side, the Web portal-based access continues to be demanded by WISPs because it requires no special application setup nor security credentials configuration in the user device. Unlike the web portal-based access, the IEEE 802.1x ensures confidentiality and integrity protection services at the transfer plane. In the following we gives more details about these two authentication mechanisms.

IEEE 802.1x Authentication

Security implementation in WLAN access is undoubtedly continuously improving. The Wired Equivalent Privacy (WEP) was the first security algorithm specified in IEEE 802.11 standards. The aim of this algorithm was to provide a level of security

comparable to that of wired network [IEE97]. WEP ensures the confidentiality protection for subscriber's messages. However, the secret key (WEP key) of the WAP should be manually configured in each device connecting to this WAP. Security analysis showed that security implementation in the original IEEE 802.11 is vulnerable to several attacks capable of WEP key cracking [BBO08]. To overcome this security weakness, the WiFi alliance introduced the WiFi Protected Access (WPA) as a replacement of WEP. WPA was just an intermediary version of the security implementation in WLAN. Finally, IEEE 802.11i [IEE04] was proposed as an amendment of the original IEEE 802.11 and is referred to as Robust Security Network (RSN). It adopted the IEEE 802.1X protocol to ensure port-based access control.
The 802.1x introduces three main functional entities in WLAN architecture:

- *Supplicant*: represents the party that needs to be authenticated.
- *Authenticator*: is defined as the gatekeeper of the access network. It delegates the authentication service to the authentication server and waits for its approval to open the access for the subscriber.
- *Authentication Server (AS)*: is defined as the entity that provides authentication service to the authenticator. Therefore, the AS task is to check the supplicant credentials and to send the supplicant rights to the authenticator.

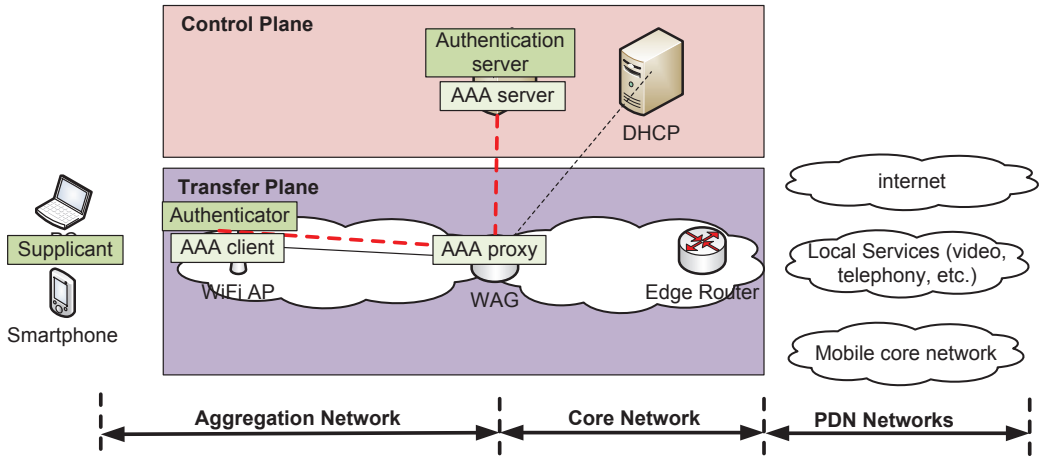


Figure 2.4: WLAN architecture with IEEE 802.1x access control

To adopt the 802.1x access control, the WLAN control plane should include a AAA server (Figure 2.4). In addition, the supplicant, the authenticator and the AS functions should be integrated within user device, WAP and AAA server entities, respectively. The communication between the AS and the authenticator is based

on the AAA protocol such as RADIUS and Diameter. Therefore, the WAP should include AAA client. Moreover, a AAA proxy can be implemented in the router in order to relay the WAP-AS exchanges.

Upon a successful authentication, the AS may generate and forward cryptographic keys associated with the authenticated user device to the WAP. Therefore, the 802.1x Authentication mechanism can be used only in a trusted WiFi network (i.e. tusted WAP). The user device is not allowed to get an IP address unless it is successfully authenticated by the AS.
The authentication exchange between the AS and the supplicant relies on the Extensible Authentication Protocol (EAP) [ABV+04]. As a generic authentication framework, EAP supports multiple authentication methods, called EAP methods, such as EAP-MD5-challenge (based on pre-shared secret), EAP-TLS (based on certificates) and, EAP-AKA (based on the smartcards). In addition, EAP messages may be carried over either data link layer protocols like PPP and 802.1X or network layer protocols such as PANA, IKEv2 and DHCP. The number of messages exchanged during an authentication procedure depends on the selected authentication method. The IEEE 802.1X has adopted five EAP methods as the official authentication methods namely EAP-TLS, EAP-SIM, EAP-AKA, LEAP and EAP-TTLS. Among these methods, we give a brief overview of EAP-TLS, EAP-AKA and EAP-TTLS.

- *EAP-Transport Layer Security (TLS)*: is a mutual authentication method where the supplicant and the server use certificates as credentials. A Public Key Infrastructure (PKI) is required for the smooth running of the method. However, the disadvantages of PKI infrastructures reside in their expensive cost and implementation complexity. Generally, access network providers use the vendor certificate associated with the MAC address at the user side and provide the AS with a certificate. Upon a successful authentication, session key is generated at both sides (supplicant and AS).

- *EAP-Authentication and Key Agreement (AKA)*: is a well-known mutual authentication method. Likewise the authentication mechanism used in the 3GPP cellular access, it is based on a pre-shared key authentication method. The pre-shared key is delivered to the subscriber through the USIM card. The same key is stored in the access provider database.

- *EAP-Tunneled Transport Layer Security (TTLS)*: is a hybrid authentication mechanism based on the use of pre-shared secret and certificates. The AS authenticates itself to the user using its certificate. Then, a tunnel is established between the user device and the AS to secure the authentication exchanges. The user uses a pre-shared secret authentication method to authenticate itself

to the AS.

The IEEE 802.1x authentication provides a greatly enhanced user experience because it allows the seamless authentication of users (i.e. without the need for user intervention).

Web authentication

The Web portal-based access is another authentication method that continues to be demanded by WISPs because it needs no specific application to be installed at the user device.

The portal-based access control requires two basic entities namely Portal server and AAA server (Figure 2.5). The Portal server provides free portal services and web-based authentication. This server also retrieves the user credentials and send them to the AAA server. In fact, after attaching to the WAP, the user opens the web browser (i.e. an HTTP request). The WAG redirects the HTTP request to the Portal server which sends to the user an authentication page to enter his credentials (login and password). Then, the Portal server retrieves the user credential and transfers them to the AAA server through the WAG.

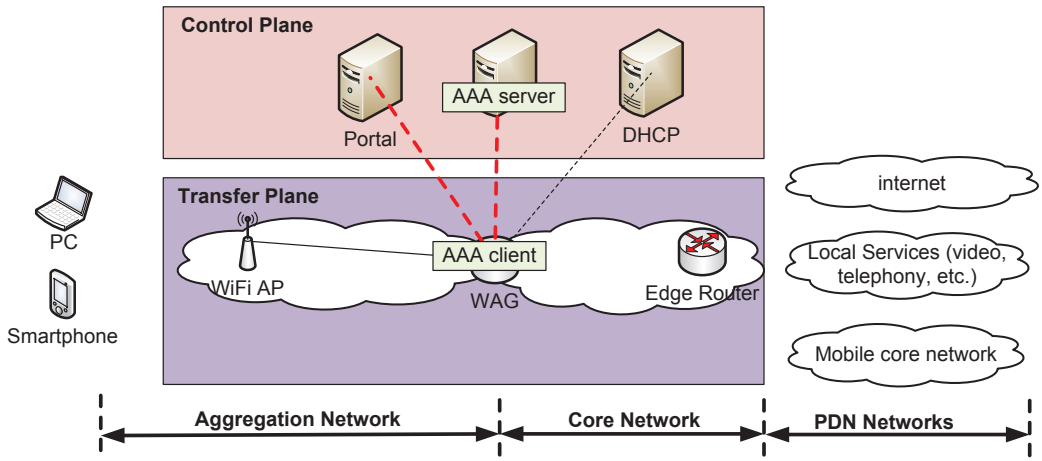


Figure 2.5: WLAN architecture with web access control

With this authentication method, a public IP address should be manually configured at the user side or obtained from the DHCP server before the authentication procedure. With this IP address, the unauthenticated user can only connect to the Portal server and use free local services. The WAG should implement the HTTP redirect mechanism in order to redirect the unauthenticated HTTP request to the portal server. After authentication, the user is allowed to connect to the internet. Although the web authentication is the most familiar mechanism found in untrusted WiFi network such public hotspots, it presents some limitations. For instance, the

captive portals require the presence of a web browser at the user device side. Moreover, users need to open the web browser and authenticate themselves before running any other applications. Finally, terminals that have Wi-Fi and a TCP/IP stack but do not have a web browser such as game consoles cannot use this mechanism.

2.3.3 Mobile access networks

Over the last few decades, we have witnessed the emergence of several wireless accesses which cover various requirements. The cellular access represents the first category of mobile access networks. The Global System for Mobile communication (GSM) [ETS97] was the first standard developed for the cellular access. Initially, the GSM was dedicated for voice traffic. The General Packet Radio Service (GPRS) extends the GSM capabilities to support the IP packet transfer. One decade later, the Universal Mobile Telecommunication System (UMTS) [3GP 9] was introduced as an evolution of the GSM access combined with the GPRS access. It provides a better data rate on the radio interface. One more decade later, the Long Term Evolution/Evolved Packet Core (LTE/EPC) architecture [3GP11b] appeared as the solution that will increase the capacity and the speed of cellular accesses. In the following, we present the UMTS and LTE architectures.

2.3.3.1 UMTS access

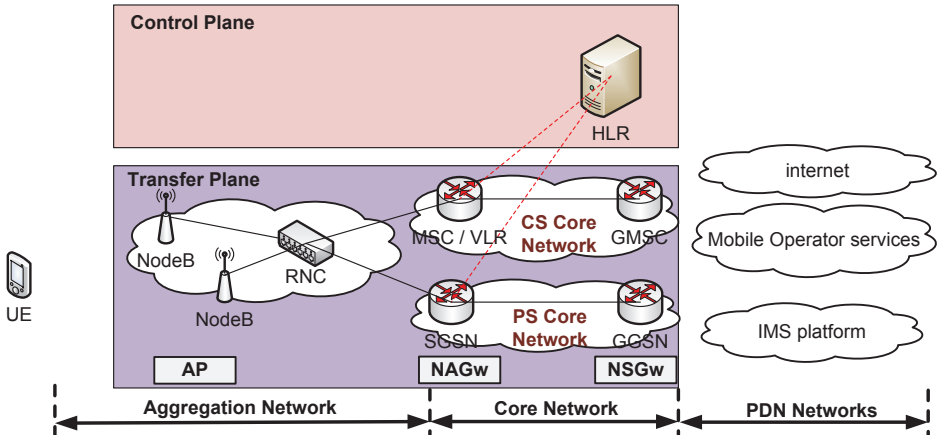


Figure 2.6: UMTS architecture

The UMTS architecture as was provided in Release 9 [3GP 9] is depicted in Figure 2.6. The aggregation network includes two main entities: Node B (NB) and Radio Network Controller (RNC). The NB corresponds to the AP in our generic model as it connects the user device to the UMTS network. The RNC controls a

pool of NBs by managing the radio resources and inter-NB mobility. Moreover, the data encryption is terminated at the RNC.

The UMTS core network is divided into two parts: the Circuit Switched Core Network (CS CN) for voice calls and SMS and the Packet Switched Core Network (PS CN) for data traffic. The CS CN is composed of Mobile service Switching Center (MSC) and Gateway Mobile Switching Center (G-MSC) entities. The MSC and G-MSC are responsible for setting up and releasing end-to-end connection, routing voice calls and SMS, handling mobility and call traffic charging. A Visitor Location Register (VLR) may be co-located with the MSC. Similarly, the Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN) are responsible for data path setup and mobility management in the PS CN.

The UMTS control plane includes the Home Location Register (HLR) where the subscribers profiles are stored. The subscriber location is always updated in the HLR entity.

The GGSN and the G-MSC corresponds to the NSGw in our generic model. The GGSN is the first IP router. It is responsible for IP address allocation and QoS enforcing. In addition, the GGSN acts as a data anchor point for inter-SGSN mobility.

The SGSN and the MSC/VLR represents the NAGw entity in our generic model. The SGSN or the VLR performs the authentication and the access control functions interacts for this objective with the HLR database. Also, the SGSN/VLR carries out the privacy function and stores the user security keys. The SGSN acts as a data anchor for mobility between RNCs.

The User Equipment (UE) represents the subscriber in UMTS access. It consists of a smart card called Universal Integrated Circuit Card (UICC) and a terminal called Mobile Equipment (ME). UICC houses the UMTS Subscriber Identity Module (USIM application). In general, UICC is known as the USIM card. The subscriber identity and the corresponding cryptographic key are contained in the USIM card.

2.3.3.2 LTE/EPC access

The LTE/EPC architecture (see Figure 2.7) is an IP-based access network that was introduced by 3GPP in Release 8. Similarly to UMTS, the network operator delivers to each subscriber a USIM card that includes the subscriber identity and the permanent security key.

The LTE/EPC aggregation network consists of several eNodeBs (eNB) and is called the Radio access network (RAN). The Core Network (CN) is called Evolved Packet Core (EPC) and consists of three main entities namely Mobility Management Entity (MME), Serving Gateway (S-GW) and PDN Gateway (P-GW)[3GP11b].

The LTE/EPC Transfer Plane includes three entities namely eNodeB (eNB),

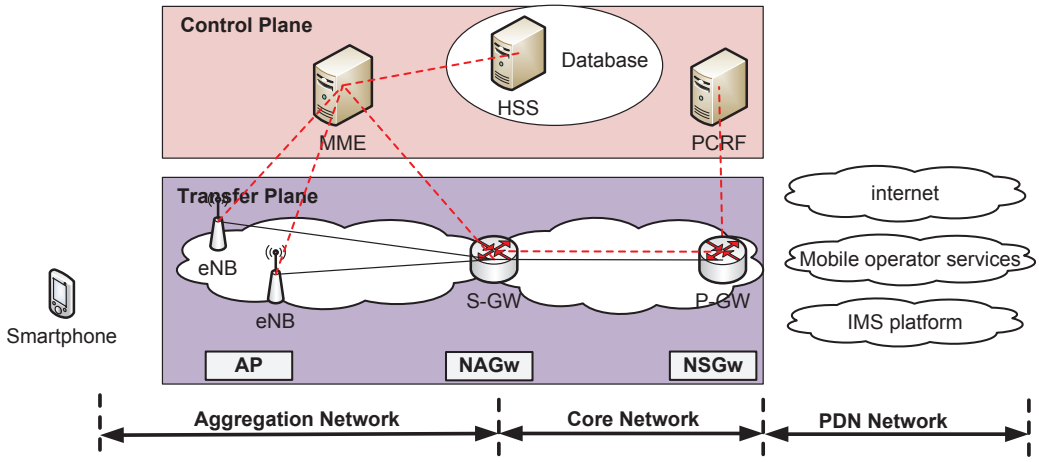


Figure 2.7: EPS architecture

Serving Gateway (S-GW), and Packet Data Network (PDN) Gateway (P-GW).

The eNB represents the first contact point in the LTE/EPC architecture. Typically, the eNB represents the AP in our generic model. It acts as a gatekeeper opening the door only for data traffic related to authorized users. The eNB relays the signaling traffic exchange between the UE and the core network.

The SGW acts as a demarcation point between the eNB and the core network. The SGW represents the NAGw in our generic model. It anchors UE data traffic for local mobility (i.e. packets are still routed through this point while UE moves between eNBs connected to the same S-GW). For each UE, it maintains an always-on connection with the P-GW to ensure its reachability. It triggers the UE paging operation in the MME and buffers packets when downlink data arrive for an idle UE.

The PGW is the termination point of the core network towards one or more PDN networks. It acts as the NSGw in our generic model. The P-GW is responsible for IP address allocation from either its own pool or the target PDN networks. The P-GW is also responsible for the QoS enforcement of the incoming traffic. The P-GW anchors UE data traffic for intra-LTE mobility (i.e. packets are still routed through this point while UE moves between several S-GWs).

The LTE/EPC Control Plane is made up of three entities: Mobility Management Entity (MME), Home Subscriber Server (HSS), and Policy Control and Charging Rules Function (PCRF).

The MME acts as the connectivity manager in LTE/EPC architecture. It is responsible for the subscriber authentication, authorization and mobility. Upon successful authentication, it generates the required keying material to secure its signaling exchange with the user and the subscriber data traffic over the wireless

link. It selects the adequate gateways (i.e. S-GW and P-GW) and triggers the admission control procedure within these gateways. It handles session continuity and user reachability when the user move from an eNB to another.

The HSS acts as a central database of all subscriber-specific information such as access restrictions for roaming, the subscribed QoS profiles and the subscriber preferences. For instance, the HSS contains information about the Packet Data Networks (PDNs) with which the subscriber is allowed to communicate. This may be in the form of Access Point Name (APN) which is a label in accordance with the DNS naming conventions or a PDN IP addresses. In addition, the HSS holds dynamic information such as the identity of the MME that currently hosts the subscriber. The HSS may also integrate the authentication center (AUC) where the authentication vectors are generated for authentication purpose.

The PCRF is responsible for policy decision-making, as well as for enforcing the decided policies in the Policy Control Enforcement Function (PCEF). Generally, the PCEF resides in the P-GW. The PCRF decides how a certain data traffic will be treated in the LTE/EPC access by providing QoS parameters such as QoS class identifier (QCI) and bit rates. Also, the PCRF checks that the decided QoS is in accordance with the user's subscription profile.

The 3GPP specification describes several procedures that are required to establish and maintain the data path inside the access network. The basic procedures are:

- Network attachment: is called as the registration procedure. The UE needs to perform this procedure to receive services from the network. Upon receiving an initial attach request, the MME connects the UE to a default P-GW. A tunnel between the S-GW and P-GW is established and maintained as long as the UE is registered to the network.

- Service Request: is performed when the UE is already registered and the data path is not already established inside the access network. This procedure may be initiated by the UE when a new application is running or upon receiving a paging request.

- Tracking Area Update (TAU): helps the network to be up-to-date of the latter UE location. It is called also UE idle mobility. In fact, the cellular network is divided in different areas called tracking areas (TA). An area consists of several cells. The TAU procedure is initiated by the UE when a change of TA is detected.

- Handover: enables session continuity when the UE moves from an eNB to another. The UE mobility may include changing the serving S-GW or MME.

Security in LTE/EPC

Each time the UE initiates one of this procedures with the network, security mechanisms are invoked. For instance, the MME should verify the user identity (i.e. authentication). Then, it downloads the corresponding user profile and decide whether the UE is authorized to use network resources (i.e. access control). In addition, an agreement on security parameters such as cryptographic keys and algorithms is performed between the UE and the MME to protect their exchanges (i.e. privacy). A second agreement is performed between the UE and the eNB to protect the data traffic at the radio level (i.e. Data Traffic Protection).

Mobility in LTE/EPC

The mobility services are completely ensured in the LTE/EPC architecture for each subscriber. The MME is responsible for managing the UE mobility by keeping track of its location, updating its data path and paging it for incoming sessions.

The user reachability service is ensured by establishing an "always-on" data bearer between the SGW and PGW related to this user (i.e. S5 default bearer) and by keeping an up-to-date track of the user location in the MME. In fact, the S5 bearer is established during the network attachment and updated during the SGW relocation. When there is no active sessions, the user device turns to IDLE state to save battery. In this mode, the Radio Access Bearer (RAB) (i.e. the radio bearer between the UE device and the eNB and the S1 bearer between the eNB and the SGW) are released. In this case, the location management functions such as location update and paging functions are required to keep the user reachable. Upon receiving an incoming packet for an IDLE UE, the SGW triggers the paging mechanism in the MME. The network coverage in LTE/EPC is divided into Tracking Areas (TAs). Each TA includes a number of eNBs. Whenever the UE crosses the TA boundary, the UE update its location in the MME via the TAU procedure.

The handover service takes place when the user with active session change the access point. A temporary data forwarding may be performed between previous and new eNBs until the data path is updated. The UE sends to the MME the handover indication to trigger the data path update procedure.

The presence of the UE profile at the serving MME enables the nomadism service. Using the UE profile, each MME can ensure for the UE the same subscribed services. During the UE mobility, the MME may download the UE profile from the HSS or request it from the previous MME.

2.3.4 Multi-Access 3GPP system

The emergence of multi-interface mobile terminals (3G, LTE, WLAN, WMAN, etc.) has completely changed access network architectures. In Release-8, the core network

EPC was specified to host multiple access network (3GPP (UMTS, LTE, etc.) or non-3GPP (WiFi, WiMAX, ADSL, etc.)) and to handle the handover between these accesses. The rational behind this was to bring convergence using a unique IP-based core network and to provide the same services over multiple access technologies.

The multi-access 3GPP system is depicted in Figure 2.8. It is made up from: several access networks (ANs) and a common core network (EPC). The Access Networks (AN) are classified into three main categories: 3GPP accesses, trusted non-3GPP accesses and untrusted non-3GPP accesses. The 3GPP accesses include the cellular accesses such as GERAN of 2G, UTRAN of UMTS and E-UTRAN of LTE. The non-3GPP accesses include the fixed accesses such as WiFi, WiMAX and xDSL. Generally, the access trustworthiness is decided by the network operator.

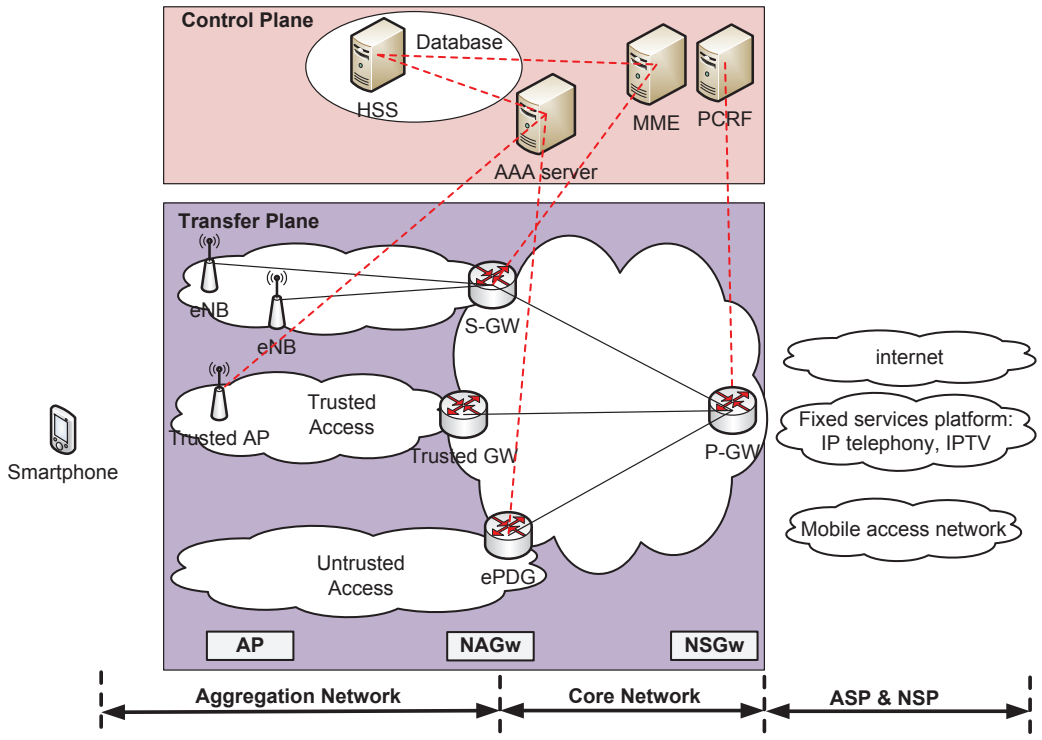


Figure 2.8: Multi-Access 3GPP System

In the control plane, a *3GPP AAA Server* is added to authenticate and authorize users that are connecting via non-3GPP access.

The untrusted non-3GPP access networks are connected to the EPC via the *evolved Packet Data Gateway (ePDG)*[3GP10a]. The ePDG represents the first point of contact of the UE with the EPC at the transfer plane when the untrusted access is in use. It corresponds to the NAGw entity in our generic model. The ePDG relays the authentication exchanges between the UE and the 3GPP AAA

server. The IPsec framework [KS05] is used to secure IP traffic between UE and ePDG. IPsec uses two protocols to ensure traffic security: Authentication Header (AH) [Ken05a] and Encapsulating Security Payload (ESP)[Ken05b]. In addition, IPsec provides two modes of operation: Transport mode, and Tunnel modes. The transport mode provides transport layer protection. In this mode, the communication goes to its destination directly. Whereas in Tunnel mode, the communication should pass through a security gateway. The ESP in tunnel mode was chosen to be used between the UE and the ePDG [3GP10a].

The trusted non-3GPP access network is right connected to the EPC via a trusted Gateway (i.e. acts as the NAGw entity). This gateway may be BNG in xDSL access or WAG in WLAN access.

Table 2.3 summarizes the security mechanisms that are implemented in three kind of accesses: LTE, Trusted WiFi and Untrusted WiFi accesses. The UE is identified with a unique permanent identity, called the International Mobile Subscriber Identity (IMSI) in all accesses. The EAP-AKA was specified as the authentication method that should be used in Trusted non-3GPP accesses (e.g. Trusted WiFi). In these accesses, the authentication is based on the same credentials as in 3GPP access (i.e. the USIM card).

security services	Access category		
	LTE	Trusted WiFi	Untrusted WiFi
Identification	IMSI	IMSI	IMSI
Authentication	EPS-AKA	802.1X	EAP-AKA transported over IKEv2
Access control	UE profile downloaded by the MME from the HSS	UE profile downloaded by the 3GPP AAA server from the HSS	UE profile downloaded by the 3GPP AAA server from the HSS
Data Traffic Protection	protection at radio link	protection at radio link	protection at IP level (i.e. IPsec tunnel)
Privacy	Temporary Identity (GUTI) Specific protection for the signaling exchanges	no temporary identity no specific protection for the signaling exchanges	no temporary identity no specific protection for the signaling exchanges

Table 2.3: Security services in multi-access 3GPP system.

The UE profile download procedure is performed in each access to ensure the access control service. The Data Traffic Protection service is ensured at radio link in LTE and Trusted WiFi. The same service is ensured at IP level in untrusted WiFi. The Privacy Service is completely ensured in LTE accesses. Actually, the

subscriber identity (IMSI) is protected by using a Temporary identity (GUTI). In addition, a specific protection is setup for the signaling exchanges between the UE and the MME to protect the mobility exchanges such as tracking area updates (i.e. the subscriber location is then protected).

In non-3GPP accesses, the subscriber mobility is managed at the network (i.e. there are no explicit mobility signaling exchanges between the UE and the network). Therefore, the privacy service is not ensured in these accesses. In case the subscriber decides to launch a host-based mobility protocol (e.g. MIP), the signaling exchanges (i.e. Binding Update and Binding Acknowledgment) with the Home Agent (HA) functionality in the P-GW should be protected. A specific IPsec tunnel should be setup between the UE and the PGW for this purpose [Per10] [3GP12a].

The multi-access 3GPP system is led by the mobility management constraints. The UE mobility within 3GPP accesses is managed by the MME and SGSN. The UE mobility between 3GPP and non-3GPP accesses is managed at IP level by the P-GW. This results in a hierarchical architecture with a centralized IP mobility anchor (i.e. P-GW) and different intermediate mobility anchors (i.e. SGW, Trusted GW or ePDG). The mobility services (i.e. session continuity, reachability as well as nomadism) are completely ensured in this model.

In Release 10, the multi-access 3GPP system has evolved to support the use cases where the same UE connect simultaneously to both 3GPP and non-3GPP accesses [OS12]. Simultaneous multi-access connectivity can be provided in several ways:

- *Multi-Access PDN Connectivity (MAPCON)*: The ability to have one PDN connection on a cellular access and another PDN on a non-3GPP access.
- *IP Flow Mobility (IFOM)*: The ability on a per IP flow basis to choose on which access each flow should be supported and to move them seamlessly between accesses.
- *Non-seamless Offload*: The ability on a per IP flow basis to choose on which access each flow should be supported, but assuming the flows over the non-3GPP access will not go through the core network (hence without support for session continuity).

2.3.5 Discussion

When we map the above fixed accesses to the generic model, we remark that the xDSL and WLAN accesses have similar architecture. In both accesses, the SAGw entity just gives the access to PDN networks. The IP address allocation and potentially the QoS enforcing are performed at the NAGw entity. Depending on the authentication mechanism, the AP entity may perform the access control by allowing only authenticated and authorized users (e.g. 802.1X in WLAN) or provide an

open access leaving the access control function to the NAGw entity (e.g. access control in xDSL or Web authentication in WLAN).

The xDSL and WLAN accesses present also similarity in that the user's session continuity is not natively granted when the subscriber moves from an AP to another. However, nothing prevents having the session continuity between two wifi AP. This depends on the additional mechanisms (e.g. Mobile IP) to be implemented in this particular access network. The nomadism is ensured in WLAN accesses as the user can use different APs to receive the same services. As the authentication in xDSL is related to the xDSL line (i.e. AP), the nomadic user cannot receive his networked service in a different xDSL line. Unlike the xDSL access, the WLAN access with the 802.1x access control is able to host multiple devices with heterogeneous authentication credentials (e.g. login/password, security key, certificate, etc.).

The LTE/EPC and the UMTS accesses have the same aim which is ensuring the mobility services for all subscribers. However, the LTE/EPC architecture converted the UMTS core network into an IP-based core network. Moreover, it reduced the hierarchy by removing the RNC entity and distributing the radio resource management and data encryption function between eNBs. In addition, the mobility management function has been moved from the data transfer plane (i.e. the RNC and SGSN entities manage the subscriber mobility and forward the data traffic at the same time) to the control plane (i.e. the MME entity manages the subscriber mobility and the S-GW forwards the data traffic). Similarly, the security functions namely authentication, access control and privacy has been moved from the SGSN to the MME entity.

The mapping of the mobile access architectures to the generic model shows that, unlike in Fixed access, the NSGw performs other network functions besides giving access to PDN networks. For instance, the NSGw is responsible for allocating IP addresses, enforcing network policies and managing UE mobility at IP level.

Table 2.4 summarize the security and mobility services ensured in both of Fixed and Mobile accesses. We note that the network connectivity in Fixed accesses includes basic mechanisms such as simple authentication mechanism and IP address allocation. However, the network connectivity in Mobile accesses ensures more advanced network services (e.g. privacy service, mobility services).

The multi-access 3GPP system is the most advanced architectures in terms of inter-working between heterogeneous technologies. This model was given by the 3GPP standards and is therefore naturally focused on the cellular core network. The multi-access 3GPP is not achieving a real network convergence. In fact, the interworking between heterogeneous access networks are rather designed to drive the traffic of mobile devices towards the cellular core network (i.e. EPC core) when they are connected to fixed IP networks. In return, with the increase of the data traffic, these solutions have several limitations that may cause problems in terms of

Network services	Fixed access	Mobile access
Security Services		
Authentication	various mechanisms (e.g. line-based, web-based, EAP-AKA, EAP-TLS, etc.)	same mechanism UMTS/EPS-AKA
Access Control	simple (authorized/denied)	advanced (i.e. according to subscriber profile)
Data Traffic Protection	performed in Wireless accesses (i.e. radio link protection)	always ensured
Privacy	no specific mechanism	natively ensured
Mobility Services		
Session Continuity	not ensured	natively ensured (seamless/hard handover)
Nomadism	ensured with 802.1X and web-based authentication	natively ensured
Reachability	not ensured	natively ensured

Table 2.4: Network services comparison between Fixed and Mobile Accesses

complexity, network dimensioning and necessarily associated costs for the operator. Some of these issues are as follows:

- At control plane: the security services such as authentication are still access-specific (i.e. SGSN for 3G access, MME for LTE access and AAA server for non-3GPP accesses) leading to function redundancy.

- At transfer plane: the P-GW should handle all traffic coming from the different accesses and can therefore become a real bottleneck [BBB09a]. IP tunnels should be used between network equipments to manage UE mobility. This may generate signaling and transmission overhead and add latency to UE sessions [MP08].

Moreover, the 3GPP system is not flexible enough to face the the very changing circumstances. In fact, this system ensures a full network connectivity (i.e. includes the above security and mobility services) for each subscriber systematically, independently of the real needs.

2.4 Conclusion

In this chapter, we defined and presented the network services that we considered during this thesis. Then, we used a generic model to analyze of network connectivity management in different access network architectures. From this analysis, we noted that existing network connectivity models are not optimized enough to cope with the real subscriber needs. In fact, the network connectivity in multi-access 3GPP systems activates the network mechanisms for all subscribers independently of the subscriber context (e.g. ensuring mobility for static user). In addition, the network connectivity management is still access-specific leading to functional redundancy (e.g. redundant authentication). Therefore, it would be interesting to provide a unified and context-aware connectivity manager for such environment.

Based on several network scenario usage, the next chapter examines the network connectivity behavior in multi-access 3GPP system and points out the technical challenges and main requirements for more flexible network connectivity.

Chapter 3

Network Connectivity Analysis in Multi-Access Context

3.1 Introduction

Managing the network connectivity in multi-access architectures becomes a critical issue as these architectures should be able to host heterogeneous technologies and to support various network usages. Within these architectures many network services are always provided to each user independently of their real needs. In this chapter, we use network use case scenarios to discuss the technical aspects which make the network connectivity more adaptable to changing circumstances.

First, we discuss the technical issues that are brought by the new ecosystem. Then, we elaborate several network scenario usages to analyze the network connectivity behavior in multi-access 3GPP systems. Each time, we bring to light the main requirements that make the network connectivity more adaptable to changing circumstances. Finally, we formulate the problem statement and give the related work.

3.2 Ecosystem Challenges

Building an ubiquitous access network that will connect billions of devices with various application types and different users behaviors is a truly challenge for network operators (Figure 3.1). This section describes the new ecosystem that multi-access architecture shall face.

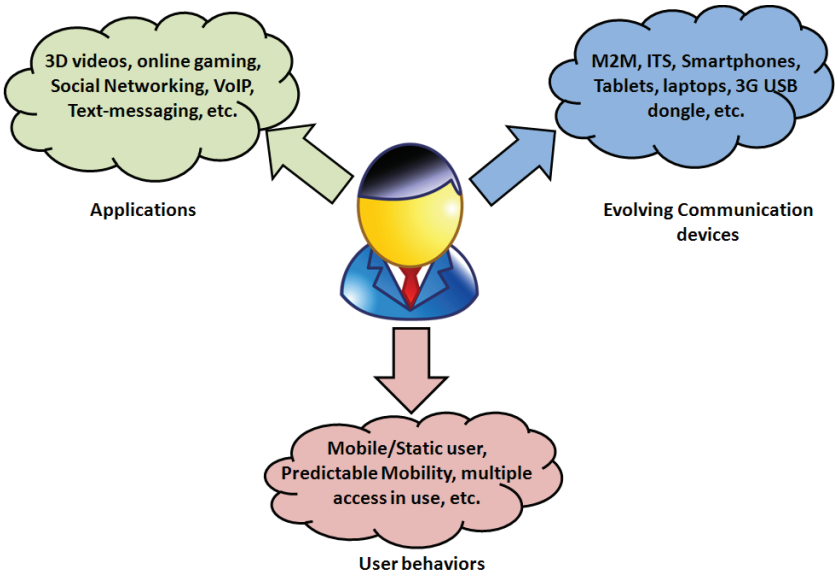


Figure 3.1: Ecosystem

3.2.1 Evolving communication devices

Each year, new devices in different capabilities with increased intelligence are being introduced in the market. For instance, Ericsson [Eri11] foresees more than 50 billion connected devices by 2020. This includes the multi-interface mobile devices, Machine-to-Machine (M2M) devices and Intelligent Transportation Systems (ITS). The feature common to all of these evolving devices is their reliance on ubiquitous networks to transmit information and distribute services to end users.

Nowadays, many mobile devices are equipped with multiple interfaces to support different wireless access technologies, typically cellular and WiFi access. These new devices could generate a large amount of data traffic. In 2012, it was noted that the typical Smartphone generates 50 times more data traffic than the typical basic-feature cell phone [Cis13]. Using 3G USB dongle, laptops can also use cellular accesses. In fact, there were 161 million laptops on the mobile network in 2012, and each laptop generated 7 times more traffic than the average Smartphone [Cis13].

In addition, several categories of smart objects will surround us in large numbers (i.e. Smart Cities). For instance, small devices in our homes will continuously monitor the home situation in terms of lighting, air conditioning and energy consumption. Patients will be equipped with small sensors that will monitor the state of patient's health. In addition, many cameras will be installed in streets to perform video surveillance. The Annual Cisco VNI Forecast expects that, by 2017, mobile networks hosts 1.7 billion M2M connections (e.g. sensors for medical applications, tracking systems in shipping, GPS systems in cars, etc.)[Cis13]. Each time, this

new category of devices requires a network connectivity to upload/download data periodically or following an event-trigger to/from their service platforms. However, hosting such great number of devices is challenging for multi-access architectures [TK12]. Supposing a scenario where power outage takes place or network equipment fails. Millions of connected devices will simultaneously re-connect, bringing down the access network. In fact, authenticating and setting up the connectivity for this number of connected devices will considerably increase the processing load in network equipments.

Moreover, Intelligent Transportation Systems (ITS) are advanced applications which provide innovative services relating to different modes of transport and traffic management. Thus, various advanced devices (e.g. intelligent cars, speed camera, etc.) are needed to realize these ITS applications. For instance, the traffic enforcement camera system consists of a camera and vehicle-monitoring device. This machine detects traffic regulation violations and automatically record them in digital file. This system need network connectivity (e.g. via LTE or WiFi accesses) to send this digital file to the processing system of the relevant police district [Spe11].

The evolving explosion in the number of connected devices and the disparity of their behaviors drive obviously new constraints in the current access network architectures. First, future access network architectures should support the exponential growth in the number of devices. Second, this kind of architecture should provide network connectivity adapted to device needs and capabilities. For example, network operator should select and activates the security mechanisms in accordance with battery-powered sensors capabilities. In fact, this kind of devices cannot offer their service if their energy is depleted [BZ09].

3.2.2 New Application profiles

Subscribers are more and more using a variety of applications (e.g. Email, Web Browsing, VoIP, Online Gaming, Social Networking, Machine-type communications, etc.) that may use a mix of voice, video and text-messaging. These new applications brings new challenges for access networks.

Nowadays, the use of Text-Messaging applications (e.g. SMS, Google Talk, Skype, Whatsapp, Facebook Messenger, etc.) is increasing. A survey forecasts 27.7 trillion messages by 2016 [Cro13]. In the IP-based access networks, the Text-Messaging service will use IP packets to deliver subscriber messages. Even the SMS messages will be transported over IP packets in IP-based access networks [3GP13c][3GP13d]. The data sent or received by the Text-Messaging service has two main features: short and sporadic. Establishing and maintaining tunnels (e.g. mobility or security tunnels) in the data plane to just transport a short and infrequent message may increase the load in access network control plane.

The traffic generated by Machine-type communication (MTC) presents a different pattern than the traffic due to human activity. Figure 3.2 is taken from [TNO10] and shows that the MTC includes different applications with different characteristics and various requirement. For example, the eCall applications [eEP], rarely send data and are device-originated only (i.e. no need to be reachable). The fleet management applications rarely send data but should be reachable for triggering. Metering applications regularly request a connection to send a small amount of data. Surveillance cameras will need a high bandwidth to transmit records. MTC applications that are based on remote control need an always-on connection to send and receive data. After recovering from network failure, MTC devices may reconnect immediately (i.e. for initialization or synchronization) putting strain on network equipment.

Specific QoS guarantees should be ensured for real-time MTC applications. Real-time network-based tele-operation systems involving two distant robots through IP-based wide area networks such as Telesurgical Robot Systems (TRSs) [MSP07] have become an emerging technology. Ensuring the security for this kind of communication is challenging since the security support may cause deadline misses or unacceptable QoS degradation.

Characteristics / Example Applications	Data volume	Quality of Service	Amount of signaling	Time sensitivity	Mobility	Server initiated communication
Smart energy meters	low	low	very low	very low	no	yes
Road charging	low	low	low	low	yes	no
eCall	very low	very high	very low	very high	yes	no
Remote maintenance	low	low	high	high	no	yes
Fleet management	low	low	very high	intermediate	yes	yes
Photo frames	intermediate	low	high	low	no	yes
Asset tracking	low	low	very high	high	yes	yes
Mobile payments	intermediate	low	high	very high	yes	no
Media synchronisation	high	low	high	intermediate	yes	yes
Surveillance cameras	very high	very high	low	very high	no	yes
Health monitoring	high	high	high	very high	yes	yes

very low low intermediate high very high

Figure 3.2: Machine-Type Communication examples

With the variety of applications running over access networks, the requirements in terms of network services (e.g. mobility management, QoS control, and security services) differ strongly. For instance, a static camera that surveys the street and uploads video periodically does not need the mobility services.

Some applications based on Session Initiation protocol (SIP), such as VoIP, could not be impacted by the IP address change as the session continuity is ensured at application layer. Therefore, the Session Continuity service provided by the access

network (i.e. using the mobility support at IP level) is not required. Some other applications are not able to manage the subscriber mobility, like video streaming, the Session Continuity service should be ensured by the access network in that case. Some application need just the Reachability and Nomadism services such as MTC applications with remote control. Moreover, sessions that are already secured (e.g. SSL sessions established between medical sensors and their server, VPN sessions established between the traveling employees laptops and their corporate Intranet, etc.) need neither confidentiality nor integrity protection (i.e. the Data Traffic Protection service) at the access network.

Even applications of the same category have different requirements. For example, some "Always-ON" applications such as medical monitoring need a constant network connectivity independently of the network conditions. Some other "Always-ON" applications such as social networking, emails and chat clients send small packets (i.e. keep-alive messages) to their servers periodically. The interval between the "heartbeat" messages related to these applications might be between 10s-120s[Mob12]. There are typically several parallel "Always-ON" applications running in the same user device. These periodic application updates leads to exchanging a lot of control messages between the subscriber device and the network [NM13]. Such unexpected/extra signaling traffic may lead to periodic congestion in network and even network equipment failures [Kar11]. Generally, access network have no control over such kind of applications. However, it may block or reschedule the connectivity related to such applications in case of serious network conditions degradation.

There are more complex cases where application needs depend on the subscriber context. For instance, the online gaming or video streaming need the mobility services when the subscriber is mobile (e.g. subscriber in train). These services are needless when the subscriber runs the same applications at home before going to bed.

3.2.3 Subscriber behaviors

Every subscriber follows a daily routine, imposed by their daily habits. Studies on Human mobility [DGP12] [SDK+06] [SQBB10] shows that certain places are visited by the same user periodically, and that people are likely to return to places they have visited before. A recent study [PSBD11] confirms that a large fraction of subscribers have limited and predictable mobility. For instance, when subscribers are in public transportation (e.g. bus, subway, etc.), their trajectory can be determined from bus trajectory and subscriber destination.

The mobility profile differs from one subscriber to another. We find the extremely mobile subscribers without any typical location. This could be due to subscriber's speed (i.e. traveling by train). We find also mobile subscribers with typical location

(e.g. home, bar, office, etc.) and fairly stationary subscribers (e.g. a grandmother staying at home all the time). Therefore, the mobility-related contextual information (e.g. mobility history, next location, velocity, e.g.) may enable some optimization in future multi-access architectures. For instance, the mobility support can be activated only for subscribers who move during an ongoing session [BBB09b]. In fact, in current approaches, mobility mechanisms are designed to be always activated for all kinds of sessions. These mechanisms do not take into consideration that the subscriber will probably not move during a given session (84% of subscribers are either static or nomadic during a session and 16% of subscribers are mobile users [TRKN09]).

Furthermore, the subscriber may desire to use multiple access technologies simultaneously. Therefore, IP flows belonging to the same or different applications may be spread along different accesses simultaneously to enhance subscriber experience [AADB10]. Moreover, a seamless movement of selected IP flows between different network interfaces is likely to happen. However, this approach may consume more network resources compared to using the same access for all applications. In fact, this new usage introduces functional redundancy in the architecture. For instance, multi-access 3GPP system should authenticate the same subscriber and protect his traffic in each access. Scenario 3.3.2 will illustrate the challenges brought by using multiple access simultaneously.

3.3 Network Usage Scenarios

In this section, we elaborate several scenarios to analyze the network connectivity behavior in multi-access 3GPP systems and drive new requirements for future multi-access architectures.

In each scenario, we suppose that the subscriber device (e.g. Smartphone, PC with USB dongle) is equipped with two wireless interfaces adapted for two wireless access technologies, such as LTE and WiFi. In addition, we suppose that the network operator acts as a network and service access provider at the same time. We assume that the network architecture in use follows the 3GPP standards and the VoIP sessions are handled by the IP Multimedia Core Network Subsystem (IMS).

3.3.1 Scenario A: Application requirement

3.3.1.1 Description

In this scenario, we analyze the network connectivity in different accesses for SIP-based session such as VoIP sessions. Each time, we point out the systematic activation of security and mobility mechanisms in different accesses. We assume that the subscriber is in the office and launches a VoIP session. In this situation, two

kinds of accesses are available namely the office WiFi AP (i.e. Untrusted non-3GPP access) and the LTE network.

Before initiating the VoIP call, the subscriber should register with the Proxy-Call Session Control Function (P-CSCF) and Serving-Call Session Control Function (S-CSCF) servers in the IMS platform. The subscriber's profile is stored in the HSS database. The subscriber has three subscription identities: IMSI, IM Private Identity (IMPI) and IM Public Identity (IMPU). The IMSI identifies the subscriber within the access network and the couple (IMPI, IMPU) within the IMS platform.

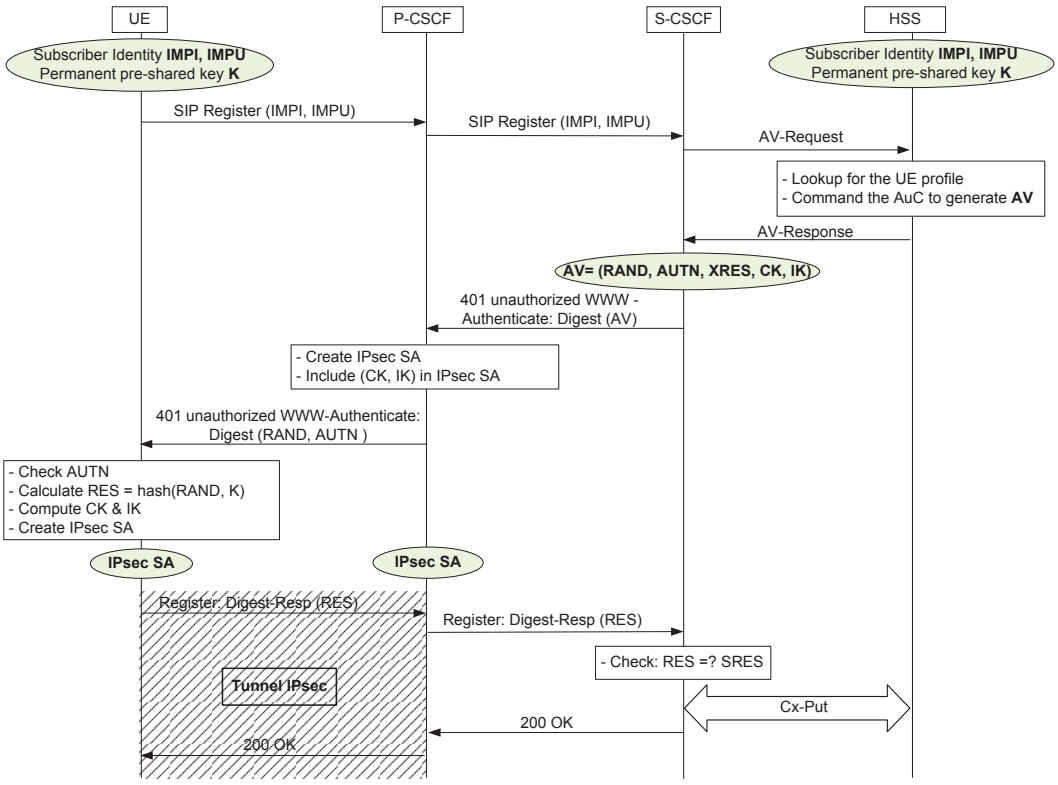


Figure 3.3: IMS registration and authentication

The IMS registration and authentication procedure is depicted in Figure 3.3. The IMS-AKA was designed to be access independent. Thus, it is performed even if the subscriber has been already authenticated at the access network of the same operator. We should note that the communication between the UE and the P/S-CSCF is based on the SIP Protocol. This application-layer protocol is able to manage the UE mobility by handling any change in IP address [SW00]. In addition, the IMS registration leads to establishing IPsec tunnel between the subscriber device and the P-CSCF server to secure the data traffic.

3.3.1.2 Analysis

Generally, between the LTE and untrusted WiFi accesses, the UE selects the LTE access for its VoIP session as it requires good QoS [GJ03]. However, if we examine the use case more deeply, we note 3 alternative connections (Figure 3.4) to reach the IMS platform:

- *Connection 1*: the UE sessions goes through the LTE access. In this connection, the mobility services are natively ensured in LTE.
- *Connection 2*: the UE sessions goes through the Untrusted WiFi access. The sessions are anchored to the core network via the ePDG.
- *Connection 3*: the UE sessions reach the IMS platform directly via the Untrusted WiFi access without passing through the core network.

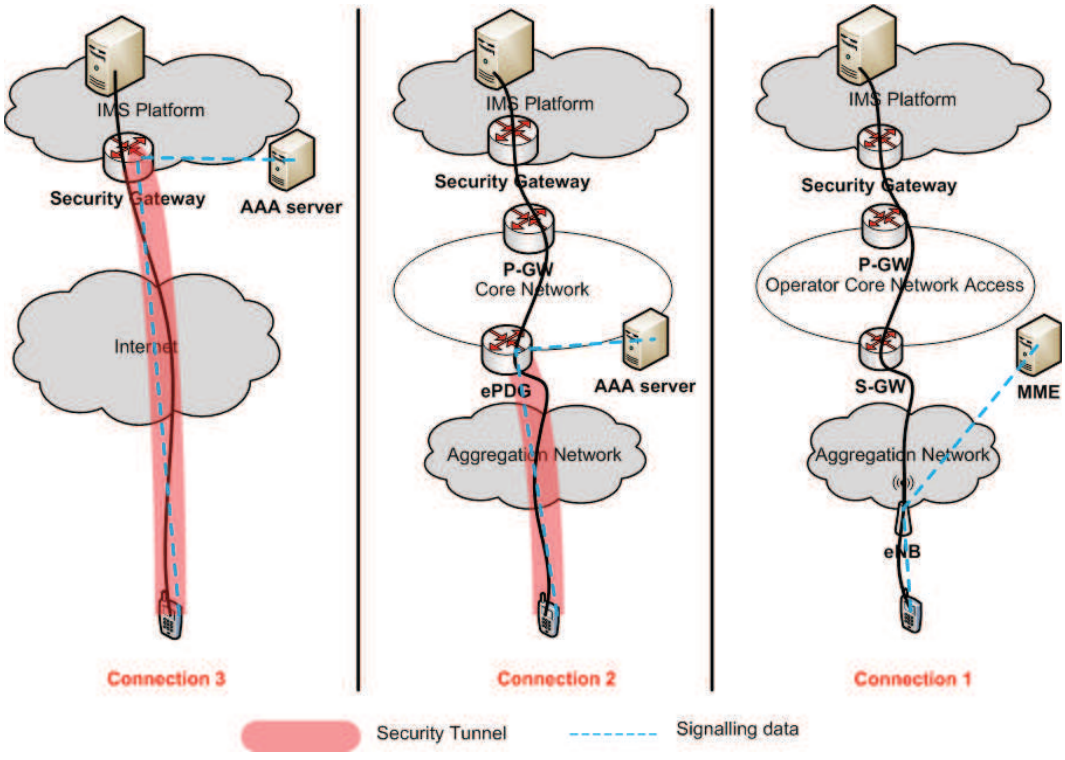


Figure 3.4: Possible Connections in Scenario A

In the following, we analyze the network connectivity in different use cases. We show that, by considering contextual information, network resources can be used more efficiently.

Use Case 1: The UE uses Connection 1

We start our analysis by examining the activated security mechanisms in such a case. First, the UE (i.e. the subscriber Smartphone) is authenticated and authorized by the MME to use the LTE access (i.e. the authentication and access control services). The DTP and privacy services are systematically setup while attaching the UE to EPC core. After that, the UE registers to the IMS platform leading to rerunning all security services. In fact, the UE is authenticated and authorized by the S-CSCF (i.e. authentication and access control services) as shown in Figure 3.3. In addition, an IPsec tunnel is established between the UE and P/S-CSCF to protect their communication (i.e. DTP and privacy services). Finally, an authentication at the VoIP server can be even required [SM06].

As we can see, the security mechanisms are performed multiple times for the same user before getting the required service. For instance, the user should run the authentication procedure three times before connecting to the targeted application via the LTE access. As the access network, the IMS platform and the related-applications belong to the same network operator, the systematic activation of the authentication procedure at different levels for the same UE is redundant. Network operators can avoid such functional redundancy by examining the application-related context before setting up the network connectivity. In fact, knowing that the IMS platform and the access network in use belong to the same network operator, the security mechanisms can be bypassed at the access network to be completely performed at the IMS platform. The second alternative consists in running the security mechanisms at the access network and bypassing them at the IMS platform as was proposed in [NXS10].

As regards the mobility mechanisms, we have the same issue. To enable the reachability service within access networks, an IP address is systematically allocated by the PGW during the UE attachment and the UE location is always updated. Similarly, the SIP protocol ensures the UE reachability by updating the UE location periodically. Such functional redundancy can be avoided if future access networks are aware of any application-related contextual information.

Use Case 2: The UE uses Connection 2

Generally, the UE should anchor its sessions to the core network via the ePDG when an Untrusted access is in use. This enables network operators to ensure session mobility between heterogeneous access or offer additional service such as parental control. Supposing that the UE uses *Connection 2* for its VoIP session.

First of all, the UE is authenticated and authorized by the AAA server (i.e. the authentication service). The ePDG acts as the authenticator. Therefore, the authentication exchanges are transported via Internet Key Exchange (IKEv2) protocol between the UE and the ePDG and then via Diameter protocol between the ePDG and AAA server (Figure 3.5). The first step of the authentication exchanges consists

in establishing a security association (IKE SA) at both of UE and ePDG to secure IKEv2 exchanges. Then, the EAP-AKA messages are exchanged between the UE and the AAA server. A successful authentication results in setting up IPsec security associations (IPsec SA) at the UE and the ePDG to secure the UE-ePDG interface (i.e. the DTP service). Similarly to *Connection 1*, the UE should be authenticated again at the IMS platform. A second IPsec tunnel is established between the UE and P-CSCF (i.e. the DTP service at IMS platform). As we can note, the authentication and DTP services are run several times for the same UE. This latter should be able to derive several security associations and applies the encryption and integrity protection algorithms on the VoIP session.

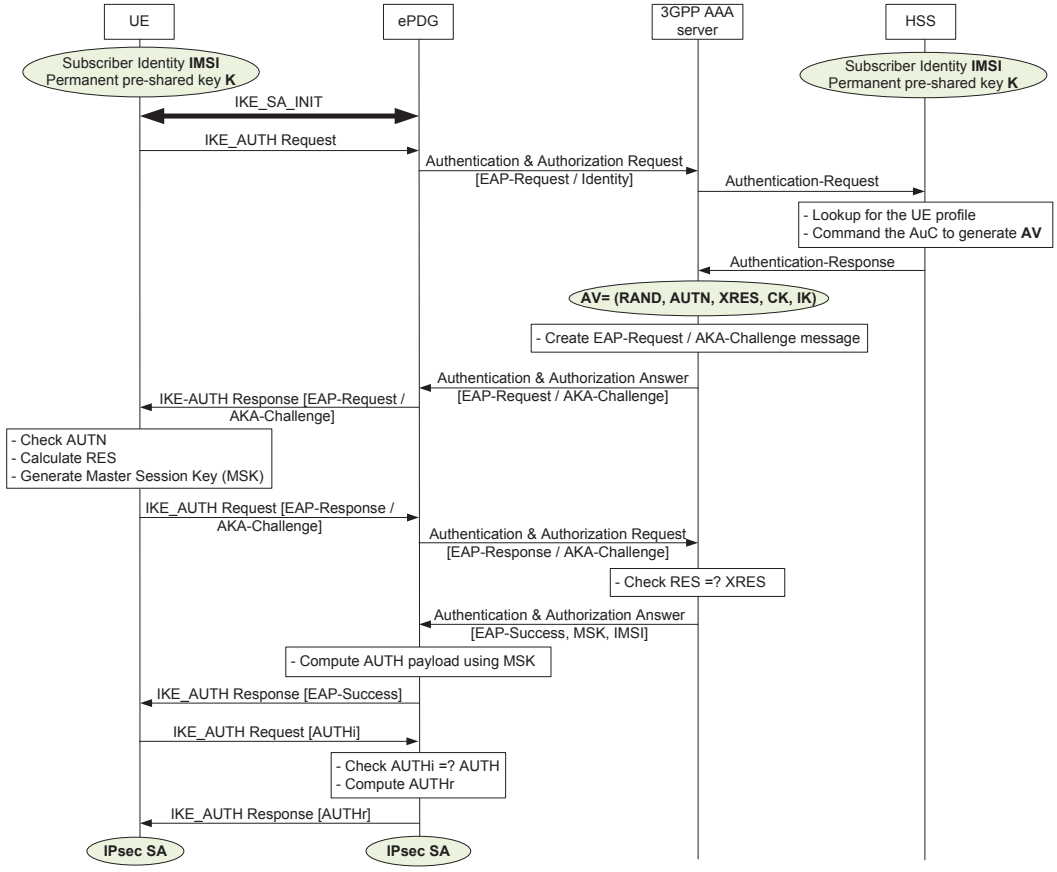


Figure 3.5: Authentication procedure in Untrusted non-3GPP access

Similarly to *Use Case 1*, the mobility mechanisms are systematically activated in this access. In fact, a tunnel between the ePDG and PGW is systematically established (i.e. GTP or Proxy Mobile IP (PMIP) [GLD+08] tunnel). Moreover, the IKEv2 Mobility and Multihoming Protocol (MOBIKE) [Ero06] is activated during the attachment phase to avoid IPsec tunnel interruption during UE mobility.

At the same time, the VoIP session is able to handle the UE mobility as it is based on the SIP protocol. As we can see, the UE mobility is managed twice: at access level via IP mobility protocol and at IMS platform via the SIP protocol. Such functional redundancy burden the UE sessions impacting the connectivity latency and throughput. In fact, a tunnel overhead is added to each packet reducing the throughput and an additional processing load is generated in both of ePDG and PGW increasing, then, the latency. However, knowing that it is highly unlikely that the subscriber moves (i.e. it is 3 PM and the subscriber is at his office), the mobility mechanisms at the access level can be deactivated. Generally, this will not impact the UE quality of experience as the SIP protocol is able to handle any change in IP address.

Use Case 3: The UE uses Connection 3
In this use case, the UE is only authenticated and authorized by the IMS platform. The VoIP session is only protected one time via the IPsec tunnel between UE and P-CSCF. In addition, as we discussed beforehand, the reachability service is already ensured at the IMS platform. It is clear that using *Connection 3* in such context saves at least some processing capacity within access networks.

Conclusion
Through this deep analysis of the network connectivity in different accesses, we derive **Requirement 1** and **Requirement 2**.

Requirement 1	The network connectivity in multi-access architectures need to be context-aware (i.e. aware of any information that may influence network connectivity behavior. Examples of context elements are given in Chapter 5.)
Requirement 2	The network connectivity in multi-access architectures should be able to adjust the network mechanisms according to the real needs. To this end, the network connectivity should be modular, where the network mechanisms should be easily activated, deactivated or configured.

The entity that will decide the network mechanism activation/deactivation within future multi-access architecture is another point that deserves to be examined. Shall the UE decide the required network mechanisms as most of the contextual information are located at the user side? Or shall we keep such decisions within access networks?

Let we examine the first alternative (i.e. the UE decides the required mechanisms). The new category of devices (e.g. Smartphones, tablets, etc.) can be effectively used to capture user activity (e.g. application usage, calling behavior, mobility history, etc.) and to record other contextual information (e.g. current location, session type, etc.) [DGP12]. Such information enables the UE to foresee the required network mechanisms. However, contextual information related to the network (e.g. available network resources, congestion status, network topology, etc.) are not easy to retrieve due to operators' security boundaries and privacy issues. Therefore, the UE could not have a global vision about the network resources constraints.

Recently, the 3GPP specifications presented the Access Network Discovery and Selection Function (ANDSF) [3GP10a][3GP12a] to be included within multi-access 3GPP systems. The purpose of the ANDSF server is to assist the UE to discover non-3GPP access networks (e.g. Wi-Fi, WIMAX, etc.) to be used in addition to 3GPP access. This function provides the UE with network policies related to each access network. Generally, the ANDSF server provides network policies (e.g. the available accesses with their priorities, etc.) to the UE. Based on these policies the UE chooses the adequate access for its session. Such solution is passive as it provides just assistance for the UE to select access. However, no adaptation within access networks (e.g. the activation of the encryption and integrity protection mechanisms) are performed.

It is difficult to imagine that the UE will decide the adequate network policies (e.g. the network mechanisms activation/deactivation) that optimize the use of network resources while providing a good QoE. In fact, the priority in the decision-maker algorithm will be for terminal-related contextual information as the UE main goal is to optimize the use of its own resources. Even when operators sends the network-related contextual information the the UE, these parameters should be simple and stable to not occupy too many network resources during their distribution to subscriber devices. In addition, the decision-maker algorithm should not occupy too much computing resources within the device. Indeed, running too many algorithms such as decision algorithms needs a special processing capacity at the UE side and increases complexity.

In the second alternative, the access network should include entities that decide the required services and activate the related mechanisms. As such entities have the access right to manage and configure network equipments, enforcing the decided adaptation is more easier. Obviously such alternative has some limitations such as the lack of the contextual information related to the subscriber and the

application in use when taking decision. However, the UE may assist the decision making process by making available some necessary contextual information to the access network, such as the device type, the available interfaces and their status, radio signal strength, the user day travel and the application QoS requirement. We conclude this brief discussion with **Requirement 3** that future multi-access architecture should satisfy.

Requirement 3	The network connectivity adaptation in multi-access architectures should be decided at the network side. The network mechanisms that will be activated should be selected and orchestrated by a trusted network entity. The subscriber may assist the decision by providing the required contextual information.

3.3.2 Scenario B: Simultaneous use of multiple accesses

3.3.2.1 Description

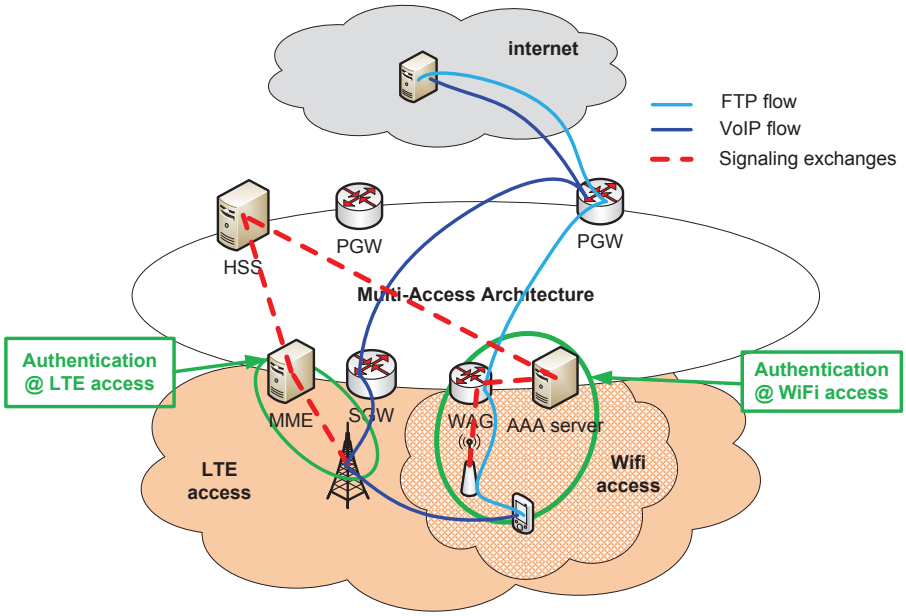


Figure 3.6: Scenario B

The purpose of this scenario is to analyze the impact of simultaneous use of different accesses on network resources usage. This new kind of connectivity was proposed in [AADB10] [KMH12] [DlOBC+11] and was adopted in 3GPP standards. We suppose that the subscriber connects to the LTE and WiFi Hotspot (i.e. trusted WiFi) accesses. The subscriber launches an FTP and VoIP sessions in parallel. We

suppose that the FTP session was assigned to the WiFi hotspot to benefit from the available bandwidth and the VoIP sessions to LTE as shown in Figure 3.6.

3.3.2.2 Analysis

For each of the heterogeneous accesses, the multi-access 3GPP system includes a technology specific authentication mechanism as shown in Table 2.3. Therefore, the UE is authenticated twice at the same time; It is authenticated by the MME in the LTE access and by the AAA server in the WiFi access. It is clear that devoting separate authentication servers (i.e. the MME and AAA server) for each access lead to redundant authentication. This issue can be resolved by unifying the authentication server in multi-access architecture.

Moreover, this use case highlights a second challenge related to the access control service: the UE profile download procedure is performed twice for the same user. In fact, upon successful authentication, the MME and the AAA server download the UE profile related the same subscriber. This lead to redundant signaling load between the serving network and the home environment and a redundant access to the HSS database. Similarly, this issue can be resolved by unifying the authentication server. Therefore, future multi-access architecture should satisfy **Requirement 4**

Requirement 4	A unified connectivity manager is required in multi-access architectures.

Regarding the mobility management, the multi-access 3GPP system activates the mobility mechanisms in each access systematically (e.g. GTP tunnels in LTE and PMIP tunnel in trusted accesses). Therefore, the mobility services are ensured for each flow of the same UE. Such kinds of usage increases the mobility signaling and processing costs. In fact, several mobility tunnels are maintained for the same UE (e.g. two tunnels in our scenario). Moreover, an UE with several flows spread over different access may require several horizontal or vertical handovers during UE mobility (i.e. in our scenario, the VoIP session could be handed off to new eNB and FTP session to new WiFi AP or eNB). In addition, the UE location is updated twice (i.e. one update at each access) instead of once. Such issues can be avoided, if the multi-access architecture has a unified connectivity manager that activates mobility mechanisms according to contextual information.

3.3.3 Scenario C: Always-On Applications

3.3.3.1 Description

To ensure the Session continuity service during UE mobility, the LTE/EPC architecture uses mobility tunnels at the data plane (i.e. GTP tunnels). These tunnels are managed (i.e. established, updated and released) by the MME. Each tunnel includes several bearers as shown in Figure 3.7. These bearers are used by network equipments to ensure the QoS.

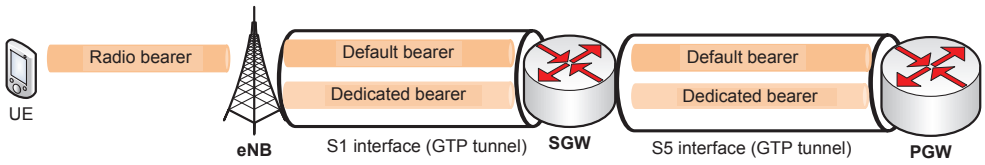


Figure 3.7: LTE/EPC data plane

When the UE needs to send or receive data through LTE/EPC access, the data plane (i.e. the data bearers between UE, eNB, SGW and PGW) should be established and the UE should move to CONNECTED state. In order to avoid excessive resource usage, the radio data bearer (i.e. between the UE and the eNB) and the S1 data bearer (i.e. between the eNB and the SGW) are released when the UE is in IDLE state. The S5 data bearer (i.e. between the SGW and the PGW) is maintained as long as the UE is registered to the network. The data traffic forwarding between the eNB, the SGW and the PGW is based on the GPRS Tunneling Protocol (GTP). Due to the generic property of 3GPP LTE/EPC architecture, the same procedures (i.e. data bearer establishment and release) are used for any type of applications.

The aim of this scenario is to analyze the effectiveness of such procedures in LTE for "Always-ON" applications.

3.3.3.2 Analysis

In this scenario, we are interested in three main procedures related to the management of mobility tunnels (i.e. GTP tunnels) at the data plane namely initial attachment, access bearer setup, and access bearer release.

- *Initial attachment procedure*: it enables the UE registration to the network during the UE initial power on [3GP11b] (see Figure 3.8). The UE sends the Attach Request message to the eNB, which includes its identity. The eNB forwards this message to MME. After a successful authentication, the MME starts the default bearer setup by sending creation session request to SGW which creates a new entry in its table and sends the same message to PGW.

Similarly, the PGW creates a new entry in its table and returns to the SGW a Create Session Response message. The SGW updates the entry related to this UE and sends the same message to MME. Therefore, the S5 data bearer is setup. Then, the MME sends to the eNB the Attach Accept message piggybacked with the Initial Context Setup Request message to setup the S1 data bearer. The eNB establishes the radio data bearer, forwards the Attach Accept message to the UE and sends back to the MME the Initial Context Setup Response message. At this time, the MME sends to the SGW the Modify Bearer Request message. In case the UE hands over from non-3GPP access such as Wifi access to the LTE access, the MME should insert Handover Indication in the Modify Bearer Request message and the SGW should inform the PGW through the Modify Bearer Request message. This will prompt the PGW to start forwarding the UE packet to current LTE access. At the end of this procedure, the UE is in CONNECTED state.

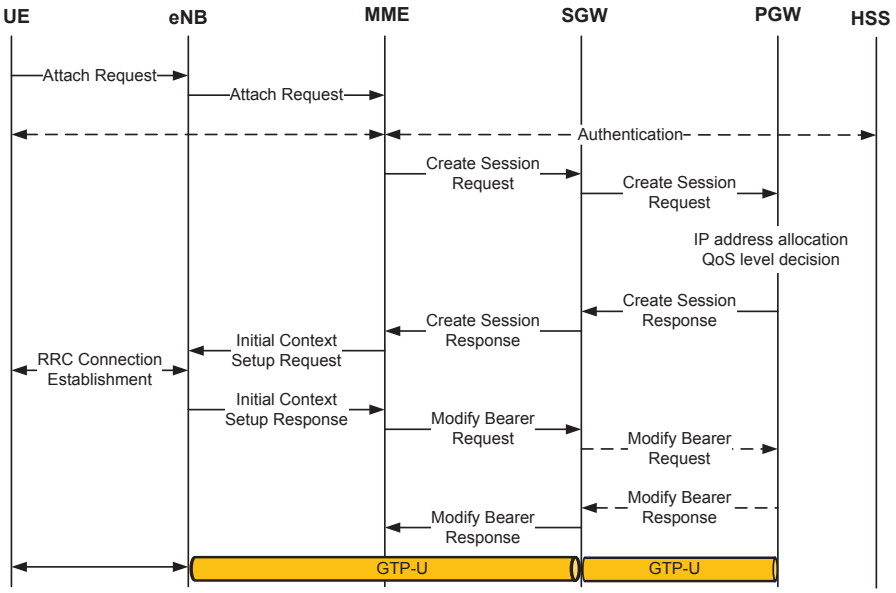


Figure 3.8: Initial attachment procedure.

- *Access bearer setup*: it is performed when the UE wants to move from IDLE to CONNECTED state (see Figure 3.9). As the S5 data bearer is maintained even when the UE is in IDLE state, this procedure will just establish the radio and S1 data bearers. Upon receiving NAS Service Request message from UE, the eNB transparently relays the NAS message to the MME. This latter initiates the UE authentication when no UE context exists. The MME sends to the eNB the Initial Context Setup Request message. The eNB establishes the radio

data bearer and sends back to the MME the Initial Context Setup Response message. The MME and the SGW exchanges the Modify Bearer Request and Response messages. Therefore, this procedure results in establishing radio and S1 data bearers.

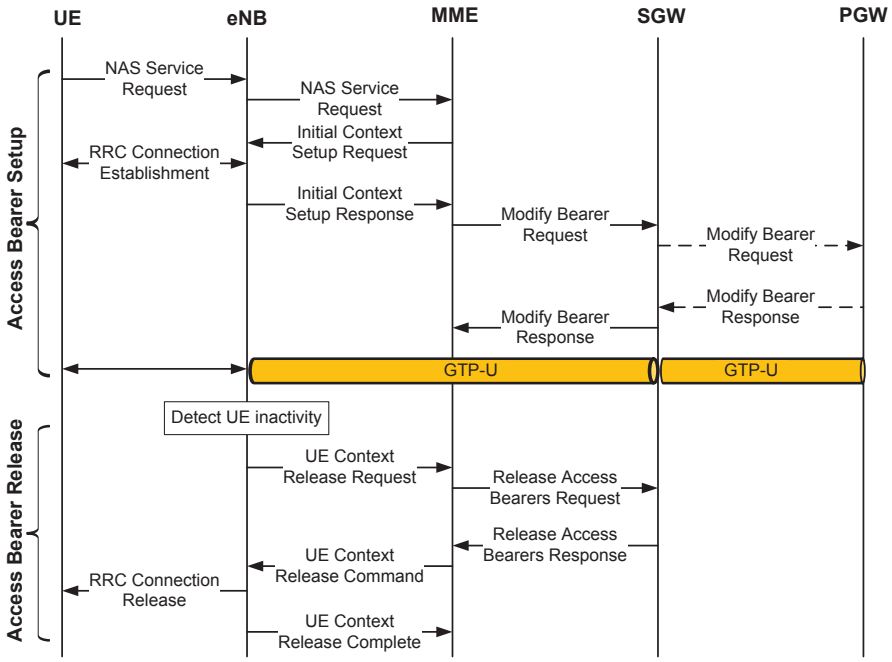


Figure 3.9: Access bearer setup and release procedures

- *Access bearer release*: Detecting the UE inactivity at the eNB triggers this procedure as shown in Figure3.9. The eNB sends to the MME the UE Context Release Request message. The MME requests the SGW to release the S1 data bearer by sending the Release Access Bearers Request message. Upon receiving the SGW response, the MME commands the eNB to release the Radio and S1 data bearers. Therefore, this procedure results in releasing the access bearer (i.e. radio and S1 data bearers) and putting the UE in IDLE state.

These LTE/EPC data plane management procedures presents several drawbacks. First, the initial attachments lead to a systematic establishment of the data plane (i.e. GTP tunnel between eNB, SGW and PGW) even when there is no data traffic to be sent.

Second, the data plane parameters are unaware of the session type. For instance, the UE inactivity timer is locally pre-configured in the eNB and has static value. However, configuring the same value for all types of sessions is not adequate. In fact, as it was mentioned beforehand, some "Always-ON" background applications

connect periodically to the network. For instance, the Chat applications sends notifications periodically to update contacts' status such as in Skype or Whatsapp. The Email application connects periodically to their server to get new emails. If the UE inactivity timer expires before the application reconnection, an extra signaling load is generated to setup again the data plane. At the same time, maintaining the data plane for applications that rarely connect to the network is a waste of network resources. Therefore, the UE inactivity timer should be adapted to the session profiles.

This scenario leads to a requirement that supplement the *Requirement 2*.

Requirement 2'	Within a given network connectivity, network resources and the maintained states should be adjusted according to the sessions needs.

3.3.4 Scenario D: Resiliency and Load balancing

3.3.4.1 Description

The purpose of this scenario is to analyze two aspects of the current network connectivity management in 3GPP LTE/EPC architecture: *(i) resiliency*, i.e. restoring active sessions when one critical equipment fails, and *(ii)load balancing*, i.e. spreading the traffic load across multiple network equipments and links during traffic peaks.

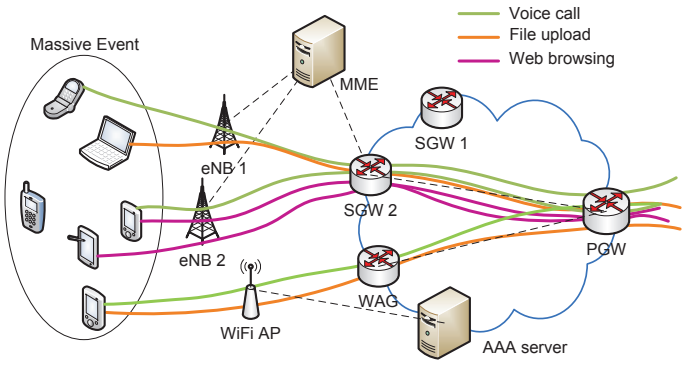


Figure 3.10: Scenario D.

We consider a Massive Event such as Olympic Gaming to illustrate network operators'need for flexible and optimized use of network resources in multi-access architectures. Supposing that the Olympic stadium receives around 200,000 visitors at peak time, keeping this number of spectators communicating represent a challenge

for the network operator. Many spectators eagerly update their Facebook or Twitter status from the Olympic Stadium. Users such as athletes, organizers and media need surely to make calls, share scores, upload photos in their cloud storage and send footage in real time using various devices. Moreover, unprecedented flows of video are streamed live to the Internet by many service broadcasting corporation. This scenario involves various kinds of network usages as shown in Figure 3.10 that may lead to bottlenecks in the network or poor Quality of Experience (QoE). It is clear that any network equipment failure in such scenario brings tremendous strain on network operators because it may lead to temporary service outage.

3.3.4.2 Analysis

We start by examining the resiliency aspect. In LTE/EPC architecture case, the 3GPP restoration procedures upon equipment failures were discussed in [3GP12e] and specified in [3GP12f]. To understand how these procedures work, let we suppose that SGW 2 goes down in our scenario (Figure 3.10). This results in automatic interruption of all ongoing sessions that are processed by this SGW.

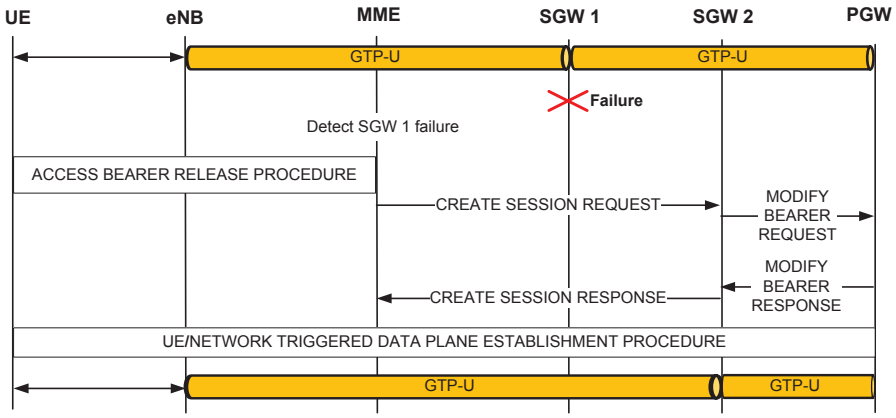


Figure 3.11: 3GPP restoration procedure after an SGW failure.

The 3GPP restoration procedure [3GP12f] related to the SGW failure is depicted in Figure 3.11. This failure can be detected by the MME or PGW via the incremented counter in the GPRS Tunnel Protocol (GTP) echo messages. Upon detecting the SGW 2 failure, the MME initiates the access bearer release procedure for active sessions that go through this failing SGW. Then, the MME assigns SGW 1 to the impacted users (i.e. idle and active users who were assigned to failing SGW) and triggers this new SGW to update all the impacted S5 bearers. Obviously, the MME informs the SGW 1 about the current PGW address and the PGW-Tunnel Endpoint Identifier (PGW-TEID) value related to each impacted S5 bearer. The

PGW-TEID value are allocated locally by the PGW and should be used by the SGW when relaying packet from eNB to PGW. The SGW 2 allocates in turn new SGW-TEID values for the S5 bearers and sends them to the PGW. The SGW-TEID parameter will be used by the PGW when relaying packet to targeted SGW. The MME should receive the Service Request message from the UE or Downlink Data Notification message from the SGW 1 to complete the restoration of active sessions.

As we can see, the 3GPP restoration procedure related to the SGW failure is not transparent as it cuts off active sessions and waits till the UEs initiate the service request procedure again. Moreover, the session re-establishment of impacted users may generate significant amount of signaling because new GTP tunnels should be established. For instance, when the SGW relocation procedure takes place, new SGW-TEID values should be allocated by the target SGW and notified to the PGW and the eNB. The TEID allocation is a key function in the GTP tunnel establishment.

Now, we examine the load balancing aspect of the current network connectivity management in multi-access 3GPP system and especially in LTE/EPC access. In the current 3GPP LTE/EPC architecture, SGW and PGW selection is performed by the MME and based on Weight Factors (WF) that are downloaded from the Domain Name Server (DNS) [3GP11b]. The WF is set according to the gateway capacity compared to the capacity of concurrent gateways. For instance, the SGW-WF is set according to the capacity of a SGW relative to other SGWs serving the same area. As the MME consider the SGW capacity before assigning the UE traffic to the appropriate SGW, we can say that it performs the proactive load balancing. Although this preventive approach, the SGW may experience periods of congestion as the current load balancing mechanism does not consider the SGW load in real-time. Indeed, an overload situation takes place when the packet arriving rate at the SGW is higher than the SGW service rate. If 200,000 users, which are assigned to the same SGW, start sending data traffic simultaneously during the peak hour, the SGW will be overloaded. Without knowing the load of SGWs in real-time and the session type, the current MME keeps assigning users to the same SGW leading to bottlenecks. In addition, the 3GPP standards specify no mechanism to temporarily free the overloaded SGW by moving some sessions seamlessly to another SGW in the same domain.

However, the knowledge about contextual information (e.g. network equipment load, application requirements in terms of latency, UE mobility profile, location, etc.) enables network operator to spread the load across multiple network equipment as shown in Figure 3.12, thereby preventing bottlenecks and equipment failures.

This scenario leads to a new requirement related to the network resources usage in future multi-access architecture (Requirement 5).

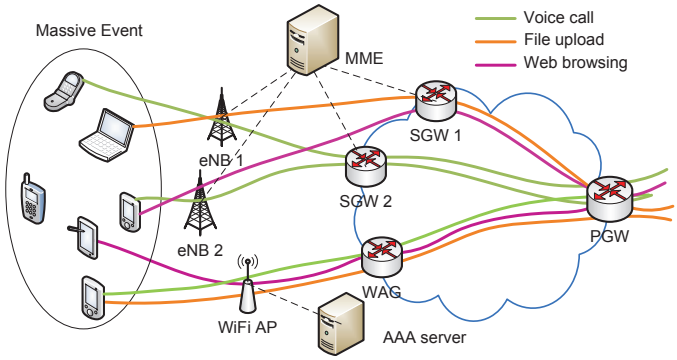


Figure 3.12: LTE/EPC data plane with load balancing.

Requirement 5	The network connectivity should be able to use network resources in a flexible and smart manner to mitigate network equipment failure and overload cases.

3.4 Problem statement and Related Work

As it was shown in Section 3.2, multi-access architectures face the challenge to host several categories of subscribers (e.g. fairly stationary subscribers, subscribers with high mobility, subscribers with small amount of data, etc.). Through the above network usage scenarios, we showed that, with the traditional network connectivity, network mechanisms such as mobility management and security mechanisms are designed to be activated in a systematic manner leading to rising the network operating cost. These mechanisms are isolated and unaware of several contextual information such as subscriber's location, subscriber's device, running application, network status, time of the day, etc. In addition, a recent study [PSBD11] shows that a large fraction of subscribers have limited mobility. Moreover, it was established in the same study that the subscriber mobility is predictable. Therefore, offering just the necessary network services for each application according to the user's environment, the application type, and the network status may help in reducing network costs. This may alleviate the control plane load as well as the data plane load while providing an adequate QoE to the subscribers.

The above network usage scenarios highlights also the lack of coordination among network mechanisms within multi-access architectures that are working at the same or different levels (e.g. link, network, application). As an example of the lack of coordination between mechanisms across levels, the authentication at link level and the authentication at application level usually do not exchange any information that could harmonize their logic and take advantage of each other's knowledge of the subscriber authentication state. Regarding the mechanisms of the same level,

the handover mechanism is not coordinating with the authentication mechanism to select the adequate authentication method that enables seamless handover. This lack of coordination results in a functional redundancy.

The problem of connectivity adaptation to ever-changing context has gained great interest over the last few years. It was motivated by the increased number and disparity of devices connected to the network, the variety of application's requirements, the variety of access network technologies, and the rapid change of the network availability. The context-awareness paradigm dates back to Dey's paper [Dey01]. There, the author defines the context as *"any information that can be used to characterize the situation of an entity"*. Bellavista et al. [BCG11] have done pioneering work reviewing many proposals on the connectivity management in heterogeneous wireless access. They asserted that the selection of the adequate connectivity should depend on the user's context (preferences, location, etc.) and the environment where the connectivity is operating (time, resource state, network bandwidth availability, etc.).

Several research papers highlighted the need for a more adapted network connectivity in Mobile access networks. For instance, [SBL11] proposed to adapt the connectivity in LTE/EPC to short sporadic messages (e.g. SMS) by sending such messages through the control plane instead of establishing the data plane each time. Indeed, this proposal reduces the signaling load related to the data plane establishment but it goes against the control and data plane separation principle that was introduced in LTE/EPC architecture specification. The paper [XLH+09] proposed an inter-gateway load balancing based on the gateway load situation. According to this paper, the overloaded SGW triggers the ongoing session handover to a less loaded SGW. To do that, the SGWs of the same service area should periodically exchange load information. However, this periodic exchange between SGWs can lead to a signal storm. In fact, if we have 10 gateways in the same service area and each gateway should periodically inform the rest about its load, 90 messages about the load will be sent periodically in this domain. In addition, a new interface should be specified between these gateways which leads to CAPEX and OPEX increase.

The paper [CVVM13] proposed to introduce a management layer function in the core network. This function is called SelFit function and has as main role to provide customized parameters independently for each subscriber.

Even the network connectivity in Fixed access need to be more adaptive and flexible [GAB+09]. For instance, [YPP+10] [YPSM11] [YPMM12] proposed a policy-based framework for end-to-end QoS control and resource management. This Framework interacts with the transport network to enable dynamic policy-based resource control that ensures the required QoS for each application.

With the promising advances in context sensing and modeling systems [Pop12] [BBH+10] [CPRW03] [CK+00], it is obvious that contextual information will be an

integral key component in managing the network connectivity in multi-access architectures. In fact, recent trends of mobile computing and communication technologies have highlighted the significant role of the context-awareness paradigm in mobile applications and services [BTSB11]. Moreover, adapting the network connectivity to the contextual information is considered as a way to improve the subscriber Quality of Experience (QoE) [CMVW10] [ZJZ10] [GAB+09].

In conclusion, we aim at designing an adaptive network connectivity tailored to a particular user, place, time and event. Contextualization is crucial in transforming network connectivity with static behavior into flexible and dynamic one - where network mechanisms are activated according to users real needs. Within access networks, there are many network mechanisms that need to be adapted. Among these mechanisms, we were particularly interested in challenging: *(i) Security mechanisms*, and *(ii) Mobility management mechanisms*. In the following, we recall the security and mobility management challenges regarding the multi-access 3GPP systems. After that, we outline the related European projects.

3.4.1 Challenge 1: Adaptive Security

The evolving explosion in communicating entities numbers, types and behaviors drives obviously new constraints in security implementation in multi-access 3GPP systems. In fact, research efforts conducted to overcome the problem of security in access network have lead to expensive and unnecessary strong and, sometimes, unwanted security solutions. For instance, the tunnel IPsec with its heavy cryptographic operation is required when an untrusted access is in use. Such strong security guarantees can not be affordable to all categories of devices.

In the current approach, security mechanisms are designed to be always activated, managing all the traffic in the same way. This leads to high processing load [TST10] at both of user devices and network equipments and generates additional signaling load [HCWzL09] [NXS10].

Moreover, current security implementations in multi-access 3GPP systems are more adapted to human communication and Smartphone capabilities. The sensors cannot support extensive security operation like keys generation and ciphering. [BZ09] highlighted the need for a tailored security where a trade-off should be made between the level of security guarantees and the amount of energy-consumption in sensors. [AM05] and [AC10] asserted that novel security mechanisms are required. These mechanisms should be able to deal with the ubiquity and the rapid change of such environments. [Hag04] designed a context-aware security framework that chooses the adequate encryption algorithm in each situation.

With many devices that connect at the same time, several points of congestion are expected. For instance, the MME is the first entity that can become overloaded

[TS11] in the LTE/EPC access. With a traditional attach procedure, the MME should perform several security operations before setting up the data plane. For instance, the MME should authenticate the connected device, generate several security keys, choose the ciphering and integrity protection algorithms, download the corresponding UE profile from the HSS, setup bearers between the eNB, the S-GW and the P-GW. Similarly, the ePDG represents the first point of congestion in untrusted non-3GPP accesses. In fact, the number of the tunnel IPsec at the ePDG increases with the increasing number of the devices that camps in an untrusted access.

Several papers have been concerned about the impact of security mechanisms on network performances. In [KK06], authors studied the energy consumption pattern of the encryption algorithms. They proposed to use the key size and the number of operational rounds in the encryption algorithm as parameters for adjusting the security level. [FKC+05] showed the higher the security level the higher the latency in WLAN access. Johnson et al. [JIFW06] proposed a practical system model to determine the adequate authentication level based on necessary trade-off between security and performance. This model has mainly three modules namely Information Collection, Decision Making, and Establishment modules. The Decision Making module uses Analytic Hierarchy Process (AHP) to decide about the adequate authentication mechanism after collecting the required information. [UFFGPC+11] adopted the cross-layer solution to avoid applying the encryption mechanism several times on the same byte; thereby, saving energy. According to their approach, the transport layer should notify the IP and physical layer whenever the data traffic is encrypted with the SSL protocol. Generally, extensive security measures can impact the network performances in several way:

- increase energy consumption due to cryptographically operations
- decrease data rates and increase the amount of signaling overhead
- increase the latency into data transmission due to the processing time reserved for each packet
- severely limit the benefit of other mechanisms such as packet compression. For instance, it is forbidden to apply the compression mechanism on encrypted packets.

The type of devices that could connect to multi-access 3GPP system is constrained by the authentication method that was specified. In fact, the same authentication method (i.e. AKA) is run whatever the access in use. The USIM card is a key component in the AKA method as it contains the UE identity and the pre-shared security key. As a result, the access will be restricted only to equipments that integrate a support for the USIM card. The rapid USIM card growth induces operating

cost that can be a serious concern for the network operator. Moreover, for each USIM card, a valid entry must be entered in the HSS database [Sysb]. Actually, the network operator should generate N USIM card for N connected devices. However, several devices may belongs to the same subscriber [Cis13]. The same user might have Smartphone, Tablet, laptop with 3G USB dongle, vehicular network with several multimedia devices, home network with sensors, etc. Each device is considered a subscription apart with a required authentication credential such as USIM card.

In some cases, the security implementation at the target access needs the intervention of the subscriber making the interworking less transparent to the subscriber. For example, the handover from the LTE access to a WiFi hotspot requires that the subscriber introduces the adequate security credentials such as the login/password or the pre-shared secret. If the WiFi hotspot is configured with a web authentication mechanism, the subscriber should introduce the login/password credential. Such act during the handover mechanism may disrupt the session continuity.

As we can see, the security in multi-access architecture is a critical issue as a complex diversity of security requirements are combined. Alternatively, security mechanisms tailored to contextual information seems to be more promising solutions. For instance, taking into consideration that the energy consuming cryptographic computations such as encryption and keys generation will deplete batteries and make devices out of service rapidly, access networks may deactivate the encryption for these categories of devices. This kind of adaptation can be advantageous for the access network as it saves at least the processing load associated to security keys computations. However, the missing security features have to be compensated at higher or lower layer for security purpose.

3.4.2 Challenge 2: Adaptive Mobility management

As we said beforehand, the multi-access 3GPP system should host multiple subscribers with different mobility pattern. As, this architecture was designed with the mobility management in mind, the mobility mechanisms such as mobility tunnels and location update are always activated managing the traffic in the same manner. For instance, the location update is always performed even for static objects such as speed camera. Moreover, as it was shown in scenario 3.3.1, the lack of coordination between the mobility mechanisms at IP level (e.g. GTP, PMIP) and application (e.g. SIP) level increases the signaling costs and results in functional redundancy.

Moreover, the sensitivity to the handover procedure varies from one session to another. For instance, the conversational sessions such as VoIP call are very sensitive to the handover. Thus, an optimized authentication procedure within the mobility management is required to avoid the intolerable latency. However, the background sessions such as contact status update in chat application are less sensitive to the

handover. As [SQBB10] showed that up to 93% of human movements can be predicted with the right prediction algorithm. Using such algorithm, network operator can predict the subscribe trajectory and select the adequate access points that can serve this subscriber during his journey. For instance, to save the latency related to the security setup, the current security context [LNPK05] from the old access point to the next access point or a new security context can be setup in the next hop before handing off the session [OWZ10].

Several handover mechanisms or interface selection algorithms have been proposed for mobile terminals in multi-access architecture. Their main aim was to select the optimal network interface or access point for each user. However, with the emergence of heterogeneous access technologies, traditional handover decision-making process based on the quality of link are no more adapted to the vertical handover mechanisms. Taking into account the contextual information has contributed in improving the handover decision-making process [ZJZ10][VCP04] and network interface selection [NVAGD08]. The ultimate goal of such a work was to improve subscriber QoE by contextualizing handover or interface selection algorithms. On the other hand, several researchers have been concerned about the challenges caused by mobile data traffic increase regarding the network operator costs [WBLR09][CYX+11] [BZR12] [TSF12]. For instance, [CYX+11] [BZR12] considered that having a centralized mobility anchors (i.e. the PGW in multi-access 3GPP system) is one of these challenges. Therefore, a distributed mobility management can be a solution for such challenge. However, it is necessary to select the adequate mobility anchor for each session that ensure the better QoE [TSF12] and to reduce the unnecessary anchor relocation [KTS10]. The offload is another solution that reduce the mobility anchor load [TSS11].

3.4.3 European Projects

There are several European projects that were interested in contextualizing the network connectivity. Among these projects, we cite:

- Scalable and Adaptive Internet Solutions (SAIL) [Sca]: is a European project that started in August 2010, and ended February 2013. According to this project, end-users need to be able to address content directly, rather than addressing servers to get the closest copy, application providers need to be able to move applications and content around in the network quickly and automatically to fulfil the varying demand, and finally the network needs to adapt rapidly to connect applications and end-users, and take advantage of all available resources. The Work Package C of this project proposed a generic framework (Open Connectivity Services (OConS)) that manages and provides connectivity services (e.g. multi-path service, multi-protocol service, Delay

Tolerant (DTN) routing, etc.) in a coordinated and consistent manner. In addition, this framework facilitates easy integration of new and enhanced mechanisms.

- Mobile Cloud Networking (MCN) [MCN] is a ongoing European Project, which aims at transforming current mobile network architecture to fully cloud-based mobile communication systems. In fact, MCN aims at extending cloud computing to support on-demand and elastic provisioning of novel mobile services and providing service orchestration with guaranteed end-to-end QoS across multiple heterogeneous technological domains - wireless, mobile core and data centers.

- Connectivity management for eneRgy Optimised Wireless Dense networks (CROWD) [mfeOWDnC]: is an ongoing European project that promotes a paradigm shift in the future Internet architecture towards global network cooperation, dynamic network functionality configuration and fine, on demand, capacity tuning. The project targets very dense heterogeneous wireless access networks and integrated wireless-wired backhaul networks. One of the CROWD key goals is guaranteeing mobile user's quality of experience by designing smarter connectivity management solutions.

3.5 Conclusion

This chapter discussed trends driving the need for adaptive network connectivity. Through detailed analysis of several network scenario usage, we showed that future multi-access architecture need to focus more on contextual information such us user behavior or application profile and activate the adequate network mechanisms accordingly. The above analysis enabled us to formulate different requirements that have to be fulfilled by future multi-access architectures. It has shown that, even though some solution may exist, they are not very efficient as the implemented network mechanisms are systematically activated leading to high signaling and processing cost. The above analysis concluded that multi-access 3GPP systems fall short of meeting the flexibility and adaptive aspects. In fact, the current network connectivity management should be reviewed, mainly the network mechanisms that should be activated for a given user.

More generally, multi-access architecture should be more reactive to context changes and to activate network mechanisms in a clever manner. To reach this level of awareness, we need context-aware connectivity that use the context-awareness features to provide a smart management of network resources. In the following chapter, we analyze and assess the security services costs in 3GPP LTE/EPC ar-

chitecture before proposing a new connectivity management model for multi-access architectures.

Chapter 4

Security Analysis in LTE/EPC Access

4.1 Introduction

The security is one of the connectivity management key components in multi-access architecture. Network operator should take into consideration the signaling overhead and processing load induced by the systematic activation of security mechanisms.

In this chapter, we evaluate the security-related signaling, processing and transmission costs quantitatively in the mono-access scenario namely the LTE/EPC architecture. To do that, we identify the signaling exchanges, processing operations and transmission overheads in each security service, considering one static UE. Then, we provide the related analytic formulation. Based on several assumptions and numerical results, we analyze the security service deactivation impact on total security-related signaling, processing or transmission costs. At the end of this evaluation section, we use several scenarios to show how far the security-related signaling and processing load can be saved when security services are adaptive.

4.2 System Model, Parameters and Methodology

4.2.1 System Model

The network model is illustrated in Figure 4.1. It is composed of a centralized home environment and distributed serving networks. The HSS database is located at the home environment. The serving network includes eNB, SGW, MME and PGW. We denote by $H_{x,y}$ the hop distance, i.e. the number of hops, between two network entities x and y. The hop distance is assumed to be symmetric ($H_{x,y} = H_{y,x}$).

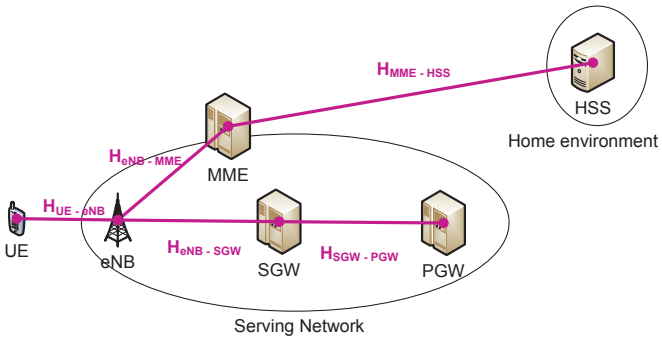


Figure 4.1: System model

4.2.2 Traffic Model

We assume that the session arrival to an UE follows a Poisson process with mean rate λ_s. Each session involves bursty sequences of packets. The inter-arrival times between subsequent packets in a session are exponentially distributed. The sizes of user packets are modeled by an independent and identically distributed (i.i.d.) random variable that follows the truncated Pareto distribution [ETS98].

4.2.3 Methodology

In our analysis, we assume that security services are successfully setup and no failure arises in any mechanism. First, we calculate the unit cost related to each security service. Then, we use these costs to discuss several scenarios where the security services are activated for a percentage of users.

The **signaling cost** is calculated as the product of the transmitted message size and the traveled hop distance. It is usually calculated in this way like in [LEC10]. The evaluation is done per security service and per interface connecting two network entities. Therefore, the security-related signaling load is evaluated at three different interfaces separately:

- UE-eNB interface: corresponds to the radio interface. It is based on the Radio Resource Control (RRC) and Packet Data Convergence Protocol (PDCP) protocols for the transport of the UE signaling and data traffic respectively.
- eNB-MME interface: corresponds to the IP-interface between the eNB and the MME. It is based on the S1-AP protocol.
- MME-HSS interface: corresponds to the IP-interface between the serving network and the home environment. It is based on the Diameter protocol.

The SC_X^Y denotes the cumulative security signaling overhead related to the security service X at the interface Y. The security messages size in Table 4.1 are determined

from the 3GPP specification [3GP11d][3GP12b][3GP11c]. The M_{aia} size corresponds to the transfer of one authentication vector as recommended by the 3GPP specification [3GP11a]. The M_{lua} size depends on the UE profile size denoted by S_{prof} which varies from an UE to another.

The **processing cost** is calculated as the sum of the required processing time at the network entity to complete processing operations (e.g. message transmission, key derivation, etc.) during the security service running. The evaluation is performed per security service and per network entity. The processing time depends on the security algorithms and the processing speed in use. The PC_X^Y denotes the cumulative security processing load related to the security service X at the network entity Y. For convenient analysis, we define the set of processing cost parameters in Table 4.2.

The **transmission cost** represents the security-related overhead that is added during packet delivery of an ongoing session. It is calculated as the product of the transmitted security-related overhead size and the traveled hop distance. The evaluation is done per interface connecting two network entities. The TC_X^Y denotes the cumulative security transmission overhead related to the security service X at the network interface Y. The required parameters for the transmission cost evaluation are defined in Table 4.3.

4.3 Security Cost Formulation

4.3.1 Identification and Authentication service cost

The subscription identity (IMSI) allocation generates neither signaling exchanges nor processing load during the network connectivity establishment, update and release. Indeed, the IMSI generation and storage in the USIM card and the HSS database is performed once (i.e. after signing the subscription contract). We do not consider the processing cost related to this operation in our analysis.

The authentication procedure is depicted in Figure 4.2. Upon receiving a Non-Access Stratum (NAS) message (e.g. Attach Request, Tracking Area Update Request, Service Request, etc.), the MME determines the IMSI by either retrieving it directly from the NAS message or requests it from the UE using the Identity Request message as shown in Figure 4.2. After that, the MME fetches N_{av} authentication vectors from the HSS database using the IMSI. Then, the MME challenges the UE by sending the User Authentication Request message. The UE answers the MME challenge through the User Authentication Response. Then, the MME checks the correctness of the answer. A successful authentication results in creating a security association between the UE and the MME. This security association contains the IMSI, the UE security capabilities and the security key master K_{ASME}.

Message	src-dst	Notation	Size (bytes)
	Authentication		
Identity Request	MME - eNB	$M_{ireq'}$	67
	eNB - UE	M_{ireq}	8
Identity Response	UE - eNB	M_{irsp}	13
	eNB - MME	$M_{irsp'}$	85
User Authentication Request	MME - eNB	$M_{uar'}$	100
	eNB - UE	M_{uar}	41
User Authentication Answer	UE - eNB	M_{uaa}	23
	eNB - MME	$M_{uaa'}$	95
Authentication Information Request	MME - HSS	M_{air}	307
Authentication Information Answer	HSS - MME	M_{aia}	359
	Access Control (UE profile download)		
Location Update Request	MME - HSS	M_{lur}	401
Location Update Answer	HSS - MME	M_{lua}	$234 + S_{prof}$
	Confidentiality and integrity protection for the UE-MME interface protection		
NAS Security Mode Command	MME - eNB	$M_{NASsmcd'}$	78
	eNB - UE	$M_{NASsmcd}$	19
NAS Security Mode Complete	UE - eNB	$M_{NASsmcte}$	11
	eNB-MME	$M_{NASsmcte'}$	83
	Confidentiality and integrity protection for the UE-eNB interface protection		
AS Security Mode Command	eNB - UE	M_{ASsmcd}	25
AS Security Mode Complete	UE - eNB	$M_{ASsmcte}$	24
	Security key (K_{eNB}) transfer		
Initial Context Setup Request	MME - eNB	$M_{icsrq'}$	326
Initial Context Setup Response	eNB - MME	$M_{icsrp'}$	89
UE Context Modification Request	MME - eNB	$M_{cmrq'}$	109
UE Context Modification Response	eNB - MME	$M_{cmrsp'}$	64
	Privacy		
Attach Accept	MME - eNB	$M_{aa'}$	424
	eNB - UE	M_{aa}	273
Attach Complete	UE - eNB	M_{ac}	81
	eNB - MME	$M_{ac'}$	130
Tracking Area Update Accept	MME - eNB	$M_{taua'}$	154
	eNB - UE	M_{taua}	98
Tracking Area Update Complete	UE - eNB	M_{tauc}	11
	eNB - MME	$M_{tauc'}$	83
GUTI re-allocation Command	MME - eNB	$M_{grcd'}$	75
	eNB - UE	M_{grcd}	27
GUTI re-allocation Complete	UE - eNB	M_{grcte}	11
	eNB - MME	$M_{grcte'}$	83

Table 4.1: The security messages and their sizes

Symbol	Description
C_m^j	Transmission/reception cost at the network entity j
C_g^j	Security key generation cost at the network entity j
C_v^j	Verification cost at the network entity j
C_h^j(i)	Hash computation cost of the message i at the network entity j
C_e^j(i)	Encryption cost of the message i at the network entity j
C_d^j(i)	Decryption cost of the message i at the network entity j
C_{encap}^j(i)	IPsec encapsulation cost of the message i at the network entity j
C_{decap}^j(i)	IPsec decapsulation cost of the message i at the network entity j

Table 4.2: Processing-related Notation

Symbol	Description
λ_s	sessions arrival mean rate
D_s	average session duration
S_p	average packet size (bytes)
N_p	average number of packet in an ongoing session (packets)
ρ^{up}	the ratio between the number of packets sent in uplink direction to the total number of packets sent during the session ($0 \leq \rho^{up} < 1$)
Bl	Block size in the encryption algorithm (16 bytes for AES)

Table 4.3: Transmission-related Notation

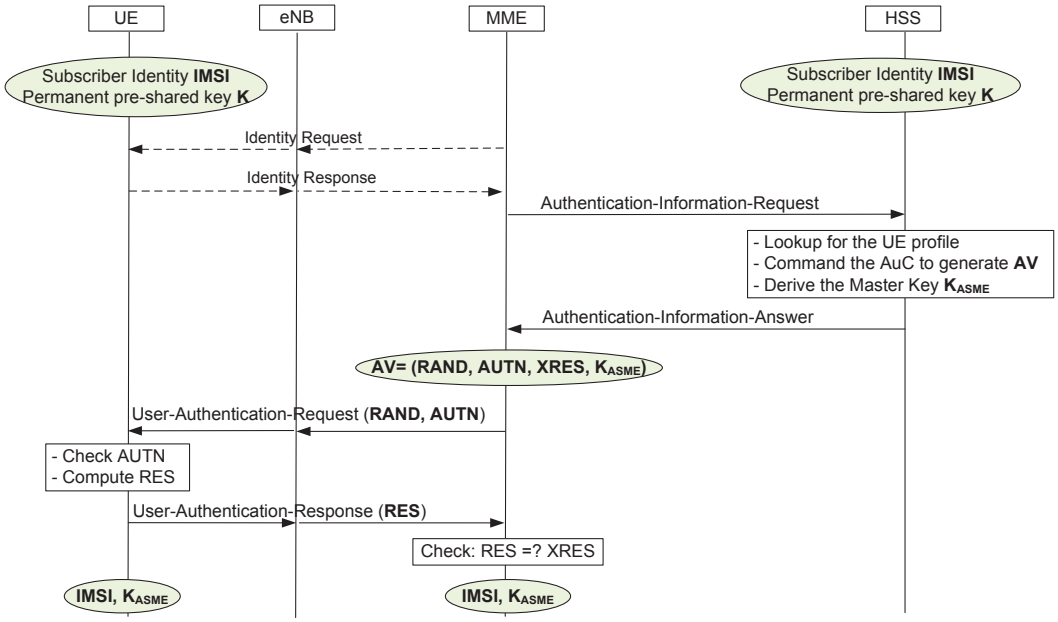


Figure 4.2: The authentication procedure in LTE/EPC access

4.3.1.1 Signaling Cost

From Figure 4.2, the unit authentication signaling cost at the UE-eNB interface is given by

$$SC_{auth}^{ue,enb} = (M_{uar} + M_{uaa} + P_{imsi}(M_{ireq} + M_{irsp}))H_{ue,enb} \tag{4.1}$$

Where P_{imsi} denotes the probability that the MME requests the IMSI parameter from the UE. Similarly, the authentication signaling cost at the eNB-MME interface is given by

$$SC_{auth}^{enb,mme} = (M_{uar'} + M_{uaa'} + P_{imsi}(M_{ireq'} + M_{irsp'}))H_{enb,mme} \tag{4.2}$$

Two signaling exchanges between the MME and the HSS take place during UE authentication. Therefore, the signaling cost at the MME-HSS interface is given by

$$SC_{auth}^{mme,hss} = (M_{air} + M_{aia})H_{mme,hss} \tag{4.3}$$

4.3.1.2 Processing Cost

The authentication service involves four entities namely HSS, MME, eNB and UE.

- **HSS**: Upon receiving the "Authentication Information Request" message from the MME, the HSS looks up for the IMSI in its database. We assume that the HSS uses a tree-based data structure to build its database. This results in a lookup cost proportional to the total number of subscribers (N_{ue}^{hss}) in the log scale[Sha11]. Then, based on the Authentication Vector (AV) retrieved from the Authentication Center (AuC), it derives the security key master K_{ASME}. After that, it sends the AV back to the MME. In this analysis, we do not consider the AV generation cost at the AuC entity. Therefore, the computational cost at the HSS includes the master key derivation cost only.

$$PC_{auth}^{hss} = 2C_m^{hss} + log(N_{ue}^{hss})C_v^{hss} + C_g^{hss} \tag{4.4}$$

- **MME**: Upon receiving an Attach Request message, the MME looks up for the subscriber security parameters in its local database. Similarly, a tree-based data structure algorithm results in a lookup cost proportional to the total number of subscribers under the same MME (N_{ue}^{mme}) in the log scale. If there is no local security context for the subscriber, the MME requests the AV from the HSS to authenticate the subscriber.

$$PC_{auth}^{mme} = 2(2 + P_{imsi})C_m^{mme} + log(N_{ue}^{mme})C_v^{mme} \tag{4.5}$$

- **eNB**: relays 2 messages. The related processing load is given by

$$PC_{auth}^{enb} = 2(1 + P_{imsi})C_m^{enb} \tag{4.6}$$

- **UE**: Upon receiving the challenge from the MME, the UE compute the key master and checks the network identity (i.e. AUTN value). Then, the UE computes the response value (i.e. RES value) by applying the one-way hash function on the challenge and the permanent security key. We assume that the response computation cost is equal to the key generation cost. The processing load is given by

$$PC_{auth}^{ue} = 2(1 + P_{imsi})C_m^{ue} + C_v^{ue} + 2C_g^{ue} \tag{4.7}$$

4.3.2 Access control service cost

The access control service in the LTE/EPC architecture is based on predefined UE profiles which are permanently maintained by the HSS database. The UE profile contains the subscriber rights such as the authorized QoS, the handover restriction lists, the authorized services, etc. The access control consists of 3 steps (see Fig4.3):

- Download the UE profile from the HSS database

- Decide whether the UE is authorized to use the requested service
- Enforce the access control rules within the serving network such as the communication path setup between the UE and the internet.

In the access control cost analysis, we focus on the UE profile download procedure. In fact, this procedure can be bypassed in some cases such as for emergency sessions. Therefore, we think that this procedure can be more adaptive if the contextual information (e.g. session type) is considered. However, the other 2 steps should be always executed independently of the contextual information to protect network resources.

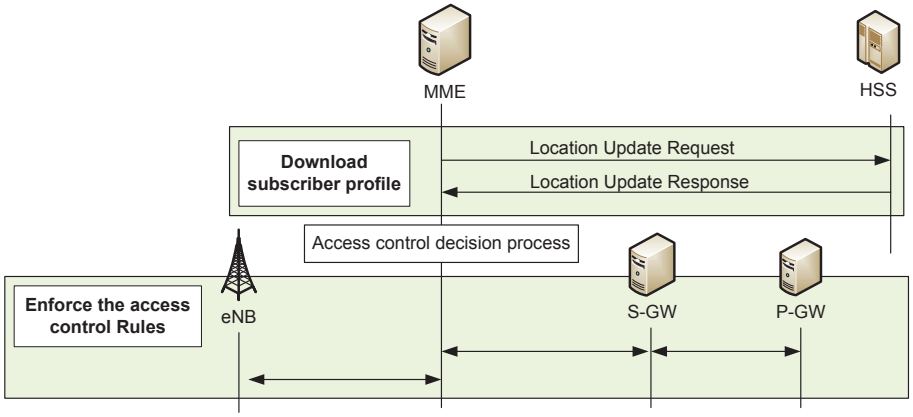


Figure 4.3: The access control procedure in LTE/EPC access

4.3.2.1 Signaling Cost

The UE download procedure cost (Figure 4.4) is evaluated as follows

$$SC_{ac}^{mme,hss} = (M_{lur} + M_{lua})H_{mme,hss} \tag{4.8}$$

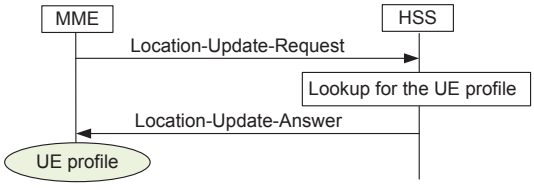


Figure 4.4: The UE download procedure

4.3.2.2 Processing Cost

Only the MME and the HSS are involved in the UE download procedure. No cryptographic operations are performed during this procedure.

- **HSS**: Upon receiving the Location Update Request message from the MME, the HSS looks up for the UE profile in its database and sends it back to the MME. The processing load is calculated as follows

$$PC_{ac}^{hss} = 2C_{m}^{hss} + log(N_{ue}^{hss})C_{v}^{hss} \tag{4.9}$$

- **MME**: exchanges 2 messages with the HSS. The processing load is given by

$$PC_{ac}^{mme} = 2C_{m}^{mme} \tag{4.10}$$

4.3.3 Data Traffic Protection service cost

The Data Traffic Protection (DTP) service is ensured through the activation of the encryption and integrity protection mechanisms at the data plane. In LTE/EPC architectures, the DTP service is provided in a hop-by-hop fashion. The first hop (i.e. UE-eNB interface) is protected via the encryption and integrity protection algorithms that are specified in 3GPP TS 33.401. The UE and the eNB should agree on security keys and algorithms before setting up the data plane. The other hops, within the serving network (i.e. between eNB and PGW), are protected via IPsec tunnels as specified in 3GPP TS 33.210 [3GP12d]. The encryption and integrity protection mechanisms within these tunnels are systematically activated and independent of the ongoing traffic type.

The Access Stratum (AS) Security Mode Command procedure takes place between the UE and the eNB to setup the UE-eNB interface protection. However the UE and the eNB do not possess a pre-shared key. Therefore, the MME derives the security key K_{eNB} from the master key K_{ASME} and transmits it with the UE security capabilities parameter to the eNB. The K_{eNB} may be transferred via

- the Initial Context Setup Request message whenever the UE switches from IDLE to CONNECTED state,
- or the UE Context Modification Request message when the MME decides to refresh the K_{eNB} and the UE is still in CONNECTED state.

The DTP service incurs two types of costs: setup-related and operation-related costs. The setup-related cost consists of the signaling and processing load that are generated during the DTP service setup. The operation-related cost includes the transmission overhead and the processing load that are generated during the DTP operation (i.e. when there is an ongoing session). We denote this second cost by the Data Protection Cost (DPC).

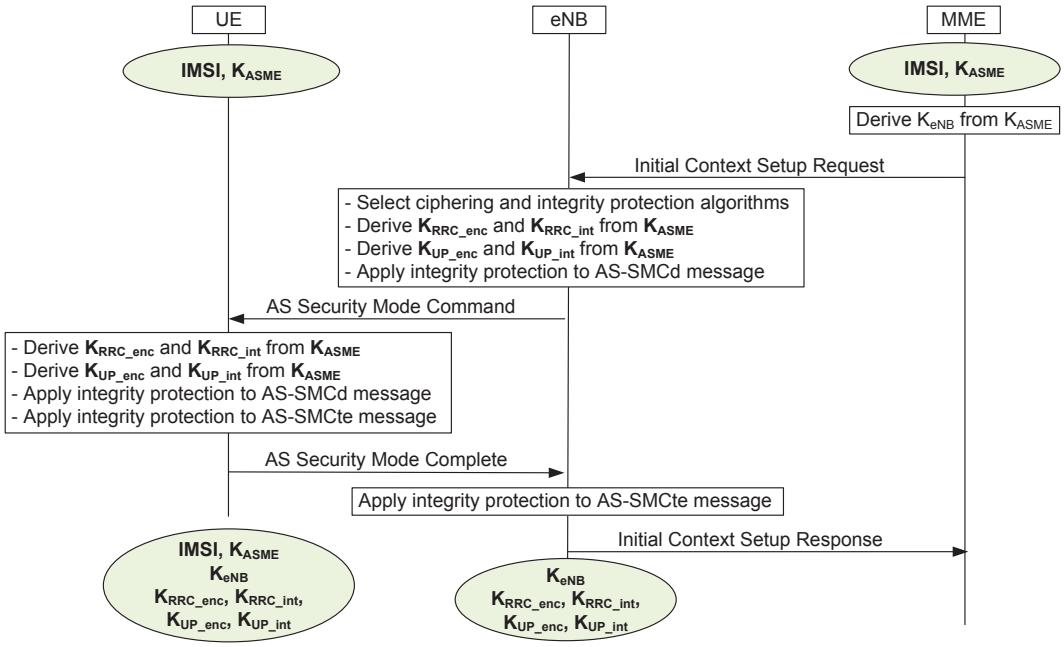


Figure 4.5: The Data Traffic Protection setup

4.3.3.1 Signaling cost

The signaling cost of the K_{eNB} transfer from the MME to the eNB is calculated as follows

$$SC_{dtp}^{enb,mme} = (P_{idle}(M_{icsrq'} + M_{icsrsp'}) + P_{connected}(M_{cmrq'} + M_{cmrsp'}))H_{enb,mme} \tag{4.11}$$

where P_{idle} and $P_{connected}$ denotes the probabilities that the UE is in IDLE and CONNECTED state, respectively. The signaling cost related to the AS Security Mode Command procedure is given by

$$SC_{dtp}^{ue,enb} = (M_{ASsmcd} + M_{ASsmcte})H_{ue,enb} \tag{4.12}$$

4.3.3.2 Processing cost

The DTP service involves three entities namely the MME, eNB and UE.

- **MME**: derives the security key K_{eNB} from the security key master and sends it to the eNB. The processing cost is given by

$$PC_{dtp}^{mme} = 2C_m^{mme} + C_g^{mme} \tag{4.13}$$

- **eNB**: derives 4 security keys from the received K_{eNB}: $(K_{RRC-enc}, K_{RRC-int})$ for protecting the RRC exchange and (K_{UP-enc}, k_{UP-int}) for data traffic protection. The eNB applies the integrity protection on the AS Security Mode Command (AS-SMCd) and AS Security Mode Complete (AS-SMCte) messages.

$$PC_{dtp}^{enb} = 3C_{m}^{enb} + 4C_{g}^{enb} + C_{h}^{enb}(M_{ASsmcd}) + C_{h}^{enb)}(M_{ASsmcte}) \tag{4.14}$$

- **UE**: derives the same 4 security keys upon receiving the AS-SMCd message. It applies the integrity protection algorithm on the AS-SMCd and AS-SMCte messages.

$$PC_{dtp}^{ue} = 2C_{m}^{ue} + 4C_{g}^{ue} + C_{h}^{ue}(M_{ASsmcd}) + C_{h}^{ue}(M_{ASsmcte}) \tag{4.15}$$

4.3.3.3 Data Protection Cost

The DPC represents the cost of protecting an ongoing session (i.e. running encryption and integrity protection mechanism). This cost includes the Transmission Cost (TC) as well as operation-related Processing Cost (PC). We start by evaluating TC within the serving network. Then, we assess the PC in each network equipment involved in session protection.

Transmission Cost

At UE-eNB interface, both encryption and integrity protection increase the final size of transmitted packets thereby participating in Transmission cost. Actually, the original packet should be padded to make its size a multiple of the basic block size of the encryption algorithm [XLMS06]. The extra pad represents the encryption overhead and is algorithm and payload specific. On the other hand, the integrity protection algorithm adds an authentication data field to the original message (i.e. Message Authentication Code (MAC-I) field in Figure 4.6). Therefore, the DTP-related TC at UE-eNB interface is calculated as follows:

$$TC_{dtp}^{ue,enb} = N_p[TC_{encryption}^{packet} + TC_{integrity}^{packet}] \tag{4.16}$$

where $TC_{encryption}^{packet}$ and $TC_{integrity}^{packet}$ represents the unit transmission costs related to encryption and integrity protection mechanisms, respectively. They are expressed as follows:

$$TC_{encryption}^{packet} = (\lceil \frac{S_p}{Bl} \rceil Bl - S_p)H_{ue,enb} \tag{4.17}$$

$$TC_{integrity}^{packet} = S_{mac}H_{ue,enb} \tag{4.18}$$

$\lceil x \rceil$ represents the smallest integer bigger than or equal to x.

Figure 4.6: A protected PDCP message

Within the serving network, the IPsec suite [KS05] and especially the ESP in tunnel mode is used to secure IP traffic between pairs of network equipment[3GP12d]. [XLMS06] and [TST10] provided an analytical formulation of the overheads imposed by IPsec and the associated cryptographic algorithms. These formula were checked with simulations and test-beds. We use these formulas to assess the transmission overhead (TC) during the data delivery within the serving network. The encryption mechanism adds an overhead to each packet (i.e. padding and ESP trailer in Figure 4.7). In addition, the integrity protection mechanism adds an authentication data (i.e. ESP auth). An ESP and new IP headers should be added to each packet to ensure the tunnel mode. Therefore, the TC within the serving network is expressed as follows:

$$TC_{dtp}^{enb,pgw} = N_p[TC_{encryption}^{packet} + TC_{integrity}^{packet} + TC_{tunnel}^{packet}] \tag{4.19}$$

where $TC_{encryption}^{packet}$, $TC_{integrity}^{packet}$ and TC_{tunnel}^{packet} represents the unit transmission overheads related to encryption, integrity protection and tunneling mechanisms, respectively. They are expressed as follows:

Figure 4.7: A protected packet with the ESP protocol

$$TC_{encryption}^{packet} = (\lceil \frac{S_p + S_{ESP\text{-}trailer}}{Bl} \rceil Bl - S_p + S_{ESP\text{-}trailer}) H_{enb,pgw} \tag{4.20}$$

$$TC_{integrity}^{packet} = S_{ESP\text{-}auth} H_{enb,pgw} \tag{4.21}$$

$$TC_{tunnel}^{packet} = (S_{ESP\text{-}header} + S_{\text{IP-header}}) H_{enb,pgw} \tag{4.22}$$

Processing Cost

Now, we examine the operation-related processing cost. To protect an ongoing session, the network equipment involved in the data delivery applies encryption/decryption, integrity protection and tunneling mechanisms on data packets as shown in Figure 4.8. Therefore, an additional processing load is generated in

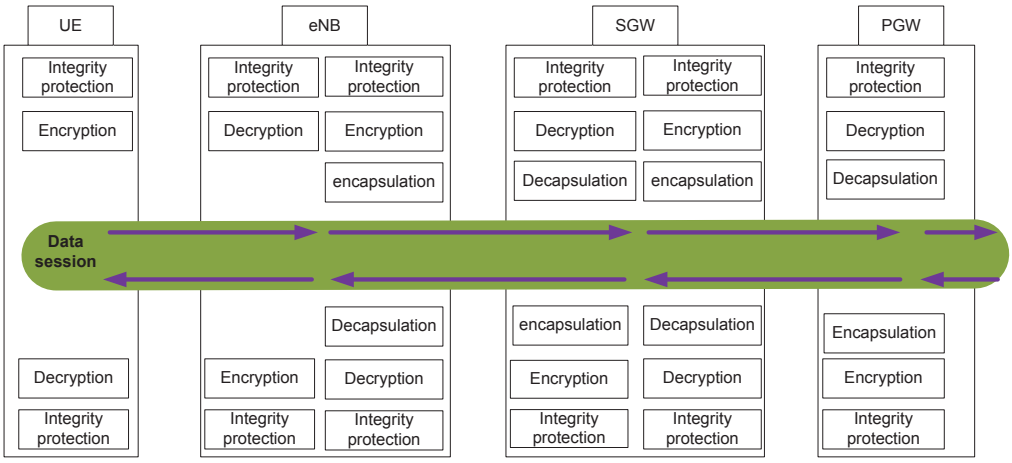


Figure 4.8: Security processing tasks during an ongoing session

these network equipment during packet delivery. In the following, we evaluate the processing cost of protecting one session in each network equipment.

- **PGW**: runs the decapsulation, decryption and integrity protection mechanisms upon receiving packets in the uplink direction. In addition, PGW runs the integrity protection, encryption and encapsulation mechanisms upon receiving packets in downlink direction. Therefore, the DTP-related processing cost in PGW during packet delivery is given by

$$PC_{dtp}^{pgw}(session) = N_p[\rho^{up}(C_{decap}^{pgw} + C_d^{pgw}(S_p) + C_h^{pgw}(S_p)) + (1-\rho^{up})(C_{encap}^{pgw} + C_e^{pgw}(S_p) + C_h^{pgw}(S_p))] \quad (4.23)$$

- **SGW**: decapsulates and decrypts packets coming from the PGW. Also, it checks their integrity protection. Then, SGW hashes, encrypts and encapsulates these packets before relying them to eNB. The same mechanisms are applied on packets sent in uplink direction (i.e. from eNB to PGW). The DTP-related PC in SGW during packet delivery is given by

$$PC_{dtp}^{sgw}(session) = 2N_p[C_d^{sgw}(S_p) + C_e^{sgw}(S_p) + 2C_h^{sgw}(S_p) + C_{decap}^{sgw} + C_{encap}^{sgw}] \quad (4.24)$$

- **eNB**: applies the decapsulation, decryption and integrity protection mechanisms on packets coming from the SGW. Then, eNB encrypts and hashes these

packets before sending them through radio link. In the uplink direction, the eNB decrypts packets coming from UE, checks their integrity and prepares them to be sent through IPsec tunnel (i.e. runs IPsec related encryption, integrity protection and encapsulation mechanisms). The DTP-related PC in eNB during packet delivery is given by

$$PC^{enb}_{dtp}(session) = N_p[2C^{enb}_{d}(S_p) + 2C^{enb}_{e}(S_p) + 4C^{enb}_{h}(S_p) + \rho^{up}C^{enb}_{encap} + (1-\rho^{up})C^{enb}_{decap}] \quad (4.25)$$

- **UE**: ciphers and hashes packets before sending them (i.e. packets in uplink direction). It deciphers and checks the integrity protection of packets received from eNB (i.e. downlink direction). The DTP-related PC in UE during packet delivery is given by

$$PC^{ue}_{dtp}(session) = N_p[\rho^{up}C^{ue}_{e}(S_p) + (1-\rho^{up})C^{ue}_{d}(S_p) + 2C^{ue}_{h}(S_p)] \quad (4.26)$$

4.3.4 Privacy service cost

The Privacy service in LTE/EPC access is responsible for protecting the signaling traffic (i.e. by applying the encryption and integrity protection to the signaling exchange between the UE and the MME) and the subscription identity (i.e. by allocating a temporary identity (GUTI)).

The Non-Access Stratum (NAS) Security Mode Command (SMC) procedure takes place between the UE and the MME to setup the confidentiality and integrity protection services for the UE-MME exchanges as shown in Figure 4.9.

In order to protect the UE subscription identity, the MME allocates the temporary identity Globally Unique Temporary Identity (GUTI) to the UE (Figure 4.10). Henceforth, the UE will use this temporary identity in any request sent to the MME instead of the permanent identity IMSI. The GUTI parameter may be transferred from the MME to the UE via

- the Attach Accept message as an answer to an Attach Request message,
- the Tracking Area Update Accept message as an answer to a Tracking Area Update Request message,
- or the GUTI Re-allocation Command message.

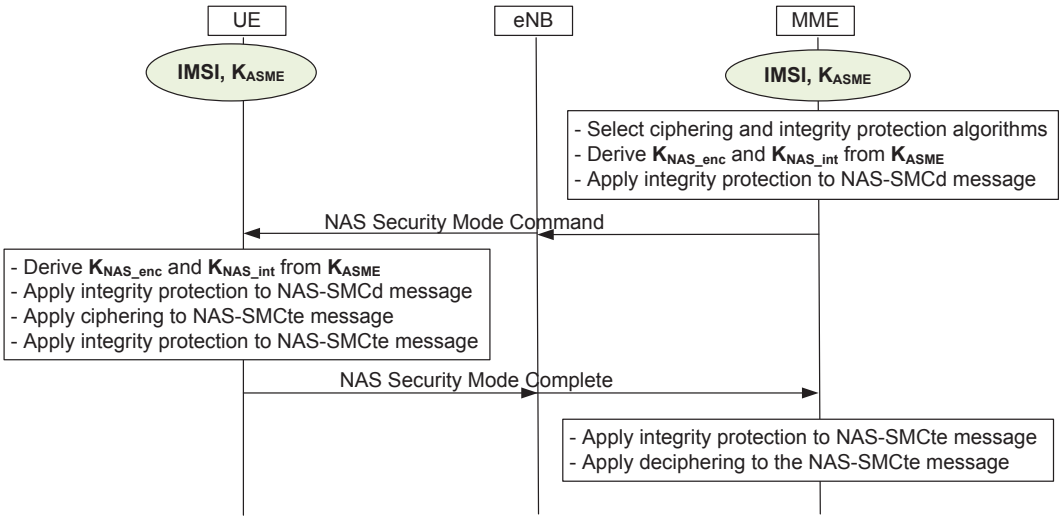


Figure 4.9: The signaling traffic protection

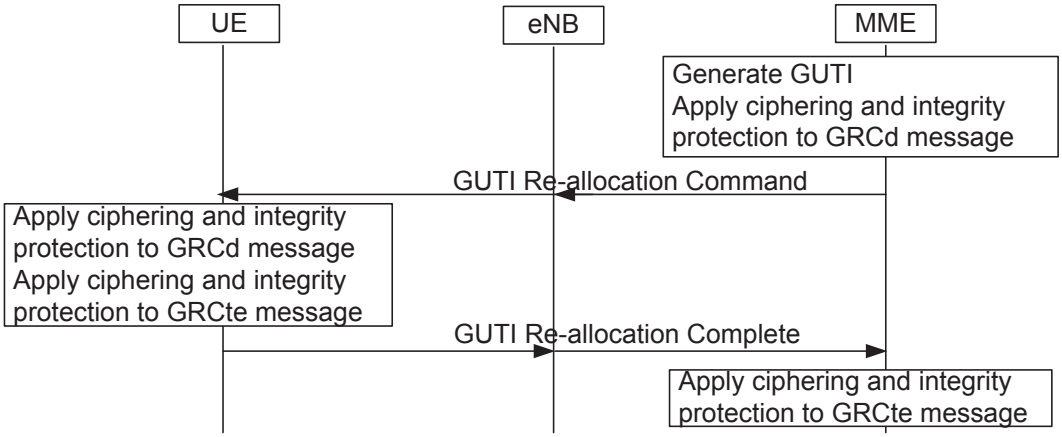


Figure 4.10: The subscriber identity protection

4.3.4.1 Signaling Cost

From Figure 4.9, two exchanges are required to setup the UE-MME protection. In addition, two exchanges are required to transfer the GUTI parameter from the MME to the UE (Figure 4.10). The signaling costs at the UE-eNB and eNB-MME interfaces are given by

$$\begin{aligned} SC_{privacy}^{ue,enb} &= SC_{signaling\ traffic\ protection}^{ue,enb} + SC_{GUTI\ allocation}^{ue,enb} \\ &= (M_{NASsmcd} + MS_{NASsmcte} + P_1(M_{aa} + M_{ac}) \\ &+ P_2(M_{taua} + M_{tauc}) + P_3(M_{grcd} + M_{grcte}))H_{ue,enb} \end{aligned} \tag{4.27}$$

Where P_1, P_2, and P_3 denote the probabilities that the GUTI is transferred via the Attach Accept, Tracking Area Update Accept, and GUTI Re-allocation Command messages, respectively. Similarly, the signaling cost of the privacy service at the eNB-MME interface is given by

$$\begin{aligned} SC^{enb,mme}_{privacy} &= SC^{enb,mme}_{signaling\ traffic\ protection} + SC^{enb,mme}_{GUTI\ allocation} \\ &= (M_{NASsmcd'} + M_{NASsmcte'} + P_1(M_{aa'} + M_{ac'}) \\ &+ P_2(M_{taua'} + M_{tauc'}) + P_3(M_{grcd'} + M_{grcte'}))H_{enb,mme} \end{aligned} \tag{4.28}$$

4.3.4.2 Processing Cost

- **MME**: selects the encryption and integrity protection algorithms and derives two security keys from the security key master K_{ASME}. Then, it notifies the selected algorithms to the UE. The MME applies the integrity protection algorithm on the NAS Security Mode Command (NAS-SMCd) and NAS Security Mode Complete (NAS-SMCte) messages. As the UE encrypts the NAS-SMCte message, the MME should decrypt it. In addition, the MME generates GUTI and transfers it to the UE. The MME applies the confidentiality and the integrity protection services on the GUTI Re-allocation Command (GRCd) and GUTI Re-allocation Complete (GRCte) messages. Therefore, the processing load at the MME is calculated as follows

$$\begin{aligned} PC^{mme}_{privacy} &= 4C^{mme}_{m} + (2+\omega)C^{mme}_{g} + C^{mme}_{e}(M_{grcd}) \\ &+ C^{mme}_{d}(M_{NASsmcte}) + C^{mme}_{d}(M_{grcte}) + C^{mme}_{h}(M_{NASsmcd})) \\ &+ C^{mme}_{h}(M_{NASsmcte} + C^{mme}_{h}(M_{grcd}) + C^{mme}_{h}(M_{grcte}) \end{aligned} \tag{4.29}$$

  where ω represents the weighting factor for GUTI generation cost. In fact, the time required to generate the temporary identity may be equivalent to the key generation time. For instance, the encryption algorithm such as AES-ECB may be applied on the IMSI to generate the temporary identity [3GP13a].

- **eNB**: relays the 4 exchanges. The processing load at the eNB is given by

$$PC^{enb}_{privacy} = 4C^{enb}_{m} \tag{4.30}$$

- **UE**: derives the same two security keys upon receiving the NAS-SMCd message. The UE decrypts and controls the integrity of the received GRCd message. Moreover, it runs the encryption and integrity protection algorithms on

the GRCte message. The processing cost is expressed as follows

$$\begin{aligned} PC^{ue}_{privacy} = 4C^{ue}_{m} + 2C^{ue}_{g} + C^{ue}_{e}(M_{smcte-nas}) + C^{ue}_{e}(M_{grcd}) \\ + C^{ue}_{d}(M_{grcte}) + C^{ue}_{h}(M_{NASsmcd}) + C^{ue}_{h}(M_{NASsmcte}) \\ + C^{ue}_{h}(M_{grcd}) + C^{ue}_{h}(M_{grcte}) \end{aligned} \tag{4.31}$$

4.4 Numerical Results and Discussions

In this section, we present and discuss the numerical results showing the impact of the context-awareness property on security signaling, processing and transmission costs. For this purpose, we make several assumptions and define default values for the parameters used in the security costs formulated above.

4.4.1 Assumptions and Default Values

The average number of hops between network entities, $H_{x,y}$, depends on network topology. The hop distances are assumed as follows: $H_{ue,enb} = 1$, $H_{enb,mme} = 2$, $H_{mme,hss} = 6$, $H_{enb,sgw} = 2$, $H_{sgw,pgw} = 6$.

From Table 4.1, we note that the $M_{icsrq'}$ size is 3 times higher than the $M_{cmrq'}$ size. In fact, the first message should include information about the data bearers that should be setup in addition to the security information. The second one transports only information that should be updated like the security key. In order to decorrelate the signaling load related to security key transfer and the signaling load related to bearer setup, we assume that the K_{eNB} is always transferred via the UE Context Modification message (i.e. $P_{idle} = 0$ and $P_{connected} = 1$).
Similarly, the Attach Accept and the TAU messages sizes are 4 and 2 times, respectively, higher than the GUTI re-allocation command message size. In fact, the Attach accept and TAU messages includes information related to the session and mobility management. In order to decorrelate the signaling load related to the privacy service and the signaling load related to session and mobility management procedures, we suppose that the GUTI is always allocated through the GUTI re-allocation command message (i.e. $P_1 = 0$, $P_2 = 0$, and $P_3 = 1$). We assume that the average UE profile size equals 518 bytes[3GP11c].

Based on mathematical formulas presented in [TST10], we computed the processing time required to cipher, decipher and control integrity of the following messages: NAS-SMCd, NAS-SMCte, AS-SMCd, AS-SMCte, GRCd and GRCte messages. To ensure that, we calculated their sizes [3GP11d][3GP12b]. We assume that each entity uses the AES algorithm for encryption/decryption and the HMAC-SHA-1 algorithm for integrity protection. Generally, the UE processing speed ranges from 50 to 200 Millions of Instructions Per Second (MIPS) [XLMS06]. For a UE with

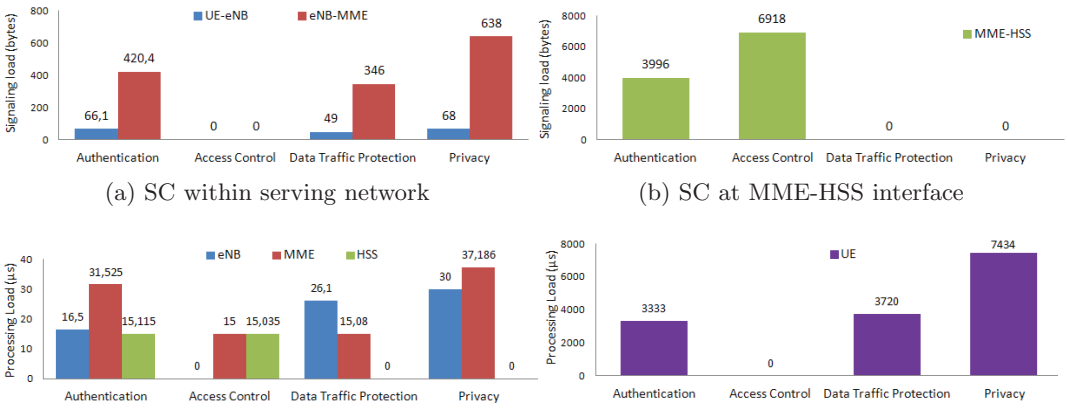
Parameter	Value
General parameters	
P_{imsi}	0.1
ω	0.2
N_{ue}^{hss}	$10,000,000\, subscribers$
N_{ue}^{mme}	$100,000\, subscribers$
Processing parameters for UE (Processing Rate = 45 MIPS)	
C_m^{ue}	$1500\,\mu s$
C_g^{ue}	$16\,\mu s$
C_v^{ue}	$1\,\mu s$
C_h^{ue}	$328\,\mu s$
C_e^{ue}, C_d^{ue}	$30\,\mu s$
Processing parameters for NE (Processing Rate = 9000 MIPS)	
C_m^{ne}	$7.5\,\mu s$
C_g^{ne}	$0.08\,\mu s$
C_v^{ne}	$0.05\,\mu s$
C_h^{ne}	$1.64\,\mu s$
C_e^{ne}, C_d^{ne}	$0.15\,\mu s$
C_{encap}^{ne}, C_{decap}^{ne}	$3.5\,\mu s$

Table 4.4: Default Values

45 MIPS as processing rate (e.g. sensor equipped with ARM Cortex-M processor [CM]), our calculations shows that the average encryption/decryption time is equal to $30\,\mu s$ ($c_e^{ue} = c_d^{ue} = 30\,\mu s$). Also, the average hashing time (c_h^{ue}) is evaluated to $328\,\mu s$. We assume that the average time needed to compare two values (c_v^{ue}) is equal to $1\,\mu s$. As the Key Derivation Function (KDF) was not specified in 3GPP standards, we assume that the HMAC-SHA-256 function is used to compute the security keys. The average processing time related to this function (c_g^{ue}) is evaluated to $16\,\mu s$ approximately [**?**]. The average signaling message transmission cost at the UE is assumed to be $1500\,\mu s$.

We assume that each Network Equipment (NE) (i.e. eNB, MME, SGW, PGW or HSS) are equipped with the Intel pentium 4 processor [Eng]. Therefore, each entity has 9000 MIPS as processing rate. The cost related to the processing operation in NE is proportional to the cost evaluated at the UE (e.g. $c_m^{mme} = \alpha c_m^{ue}$). The coefficient α is calculated as follows

$$\alpha = \frac{processing\ rate\ at\ UE}{processing\ rate\ at\ the\ network\ equipment}$$



(a) SC within serving network

(b) SC at MME-HSS interface

(c) PC at network equipment

(d) PC at UE

Figure 4.11: Security services setup costs

The unit processing costs values in NE are shown in 4.4. The IPsec encapsulation/decapsulation is assumed to be $3.5\,\mu s$ [SSGC05].

For the data traffic model, we consider an average packet size (S_p) equal to $480\,bytes$ and an average number of packet per session (N_p) equal to 1000 [ETS98].

4.4.2 Discussions

The signaling and processing load related to security services setup are shown in Figure 4.11. While Figure 4.11a focuses on the signaling load within the serving network, Figure 4.11b shows the signaling load between the serving network and the home environment (i.e. MME-HSS interface). The processing time consumed by the security services setup at both of network equipment and UE is depicted in Figure 4.11c and 4.11d, respectively.

We define the **Signaling Load Saving (SLS)** parameter as the percentage of the additional signaling load due to the activated security service. Similarly, we define the **Processing Load Saving (PLS)** parameter as the percentage of the additional processing load due to the activated security service. These parameters are relevant for the cost-based comparative analysis of the security services.

$$SLS = \frac{security\ service\ signaling\ load}{baseline\ security\ scenario\ signaling\ load}$$

$$PLS = \frac{security\ service\ processing\ load}{baseline\ security\ scenario\ processing\ load}$$

The baseline security scenario corresponds to the scenario where all security services are activated for the UE. Table 4.5 and Table 4.6 summarize the security services SLS and PLS values, respectively.

We note that deactivating the authentication mechanism saves 36%, 30% and 37% of the security signaling load at UE-eNB, eNB-MME, and MME-HSS interfaces respectively. In addition, this saves 23%, 23%, 32% and 50% of the overall security processing load at UE, eNB, MME and HSS, respectively. The confidentiality and integrity protection setup for the data traffic (i.e. DTP service setup) generates 27% and 25% of the security signaling load at UE-eNB and eNB-MME interfaces, respectively. This service represents 26%, 25%, and 17% of the security-related processing time at UE, eNB, and MME respectively. At MME-HSS interface, a large portion of the security signaling load is generated by the access control service (63%). This is due to the subscriber profile size.

Security services	Signaling Load Saving		
	UE-eNB	eNB-MME	MME-HSS
Authentication	36%	30%	37%
Access control	-	-	63%
Data Traffic Protection	27%	25%	-
Privacy	37%	45%	-

Table 4.5: Signaling Load Saving.

Security services	Processing Load Saving			
	UE	eNB	MME	HSS
Authentication	23%	23%	32%	50%
Access control	-	-	15%	50%
Data Traffic Protection	26%	36%	15 %	-
Privacy	51%	41%	38%	-

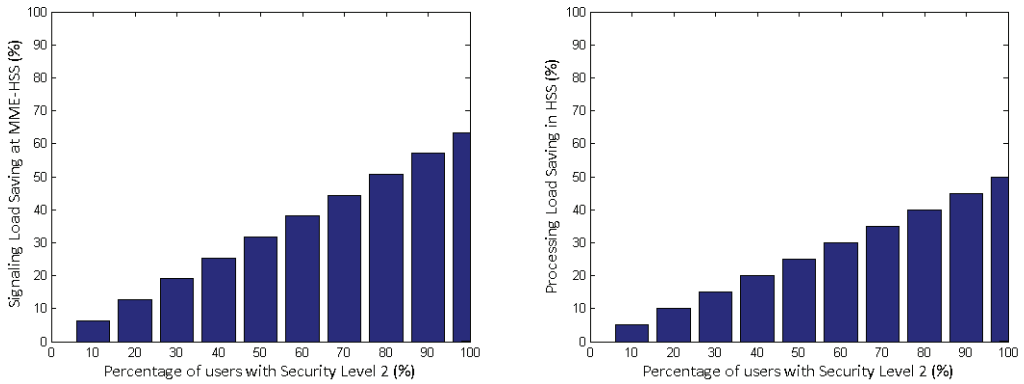
Table 4.6: Processing Load Saving.

In the following, we analyze the impact of having adaptive Access Control (AC), Data Traffic Protection (DTP) and Privacy services on the overall security costs. We considered three scenarios. In each scenario, we assume that the MME can decide between two security levels for each subscriber: Security Level 1 where all security services are activated and Security Level 2 where one security service (e.g. AC or DTP or Privacy) is deactivated.

Scenario 1: Adaptive AC service

In this scenario, we propose to evaluate the impact of adapting the UE profile download procedure in the AC service to contextual information. Before authorizing

an UE to use the LTE/EPS access, the MME should download the associated profile from HSS. However, this procedure may be bypassed; especially for a set of sensors that send periodically small packets of fixed size. We can imagine that the MME maintains a typical profile for this category of UEs. Static UEs are another category of subscribers where the UE profile download procedure is not required.



(a) SLS at MME-HSS interface (b) PLS in HSS

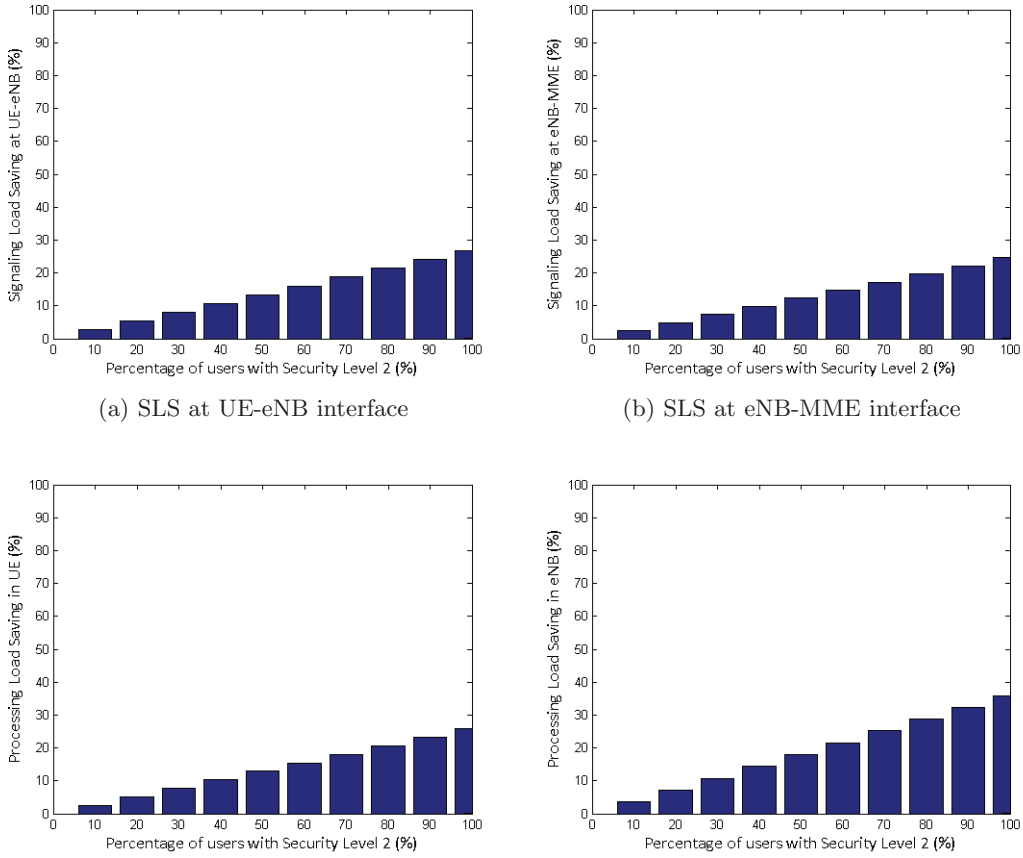
Figure 4.12: SLS and PLS in Scenario 1

We assume that the UE profile download procedure is deactivated in Security Level 2. Upon receiving an attach request, the MME uses the contextual information to decide whether the UE profile should be downloaded (i.e. security level 1 or security level 2). We vary the percentage of subscribers using security level 2 and calculate the corresponding SLS and PLS values. The results of SLS values at the MME-HSS and PLS values in HSS are shown in Figure 4.12a and Figure 4.12b, respectively. We note that approximately 13% of the security signaling load is saved at the MME-HSS interface when 20% of the subscribers uses Security Level 2. Similarly, 10% of the security processing load is saved in HSS when 20% of the subscribers uses Security Level 2. Moreover, as expected, when the percentage of UEs with Security Level 2 increases, the SLS and PLS proportionally increases.

Scenario 2: Adaptive DTP service

In this scenario, we propose to assess the impact of adapting the DTP service activation to the contextual information. This scenario is motivated by the fact that this service is systematically ensured for the UE data traffic in LTE access networks. However, traffic that is already secured like VPN sessions need no additional protection. Moreover, several M2M connections where the information sent is not sensitive do not require this kind of protection.

We assume that the DTP service is deactivated in Security Level 2. In this



(a) SLS at UE-eNB interface

(b) SLS at eNB-MME interface

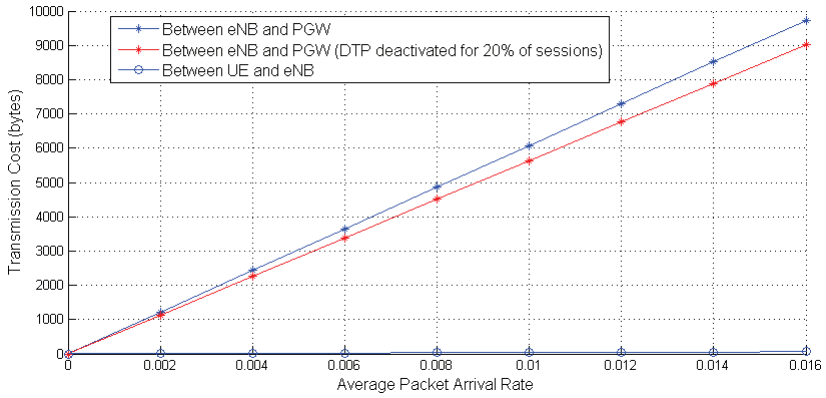
(c) PLS in UE

(d) PLS in eNB

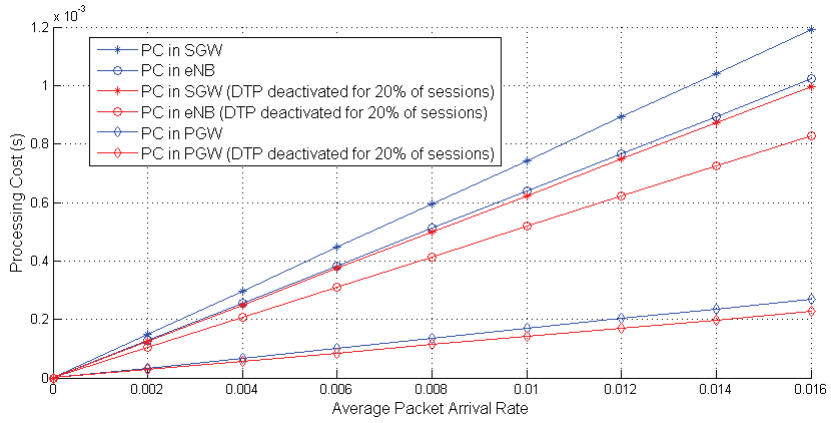
Figure 4.13: SLS and PLS in Scenario 2

level, network equipment apply no additional security mechanisms on data packets. Based on the session nature, MME decides of the security level to be deployed. For instance, the MME deploy the security level 2 for VPN sessions. The percentage of UEs using security level 2 (i.e. with deactivated DTP service) varied and the results of SLS and PLS values are shown in Figure 4.13. We note that if the DTP service is deactivated for 40% of UEs, around 10% of the security signaling load is saved at UE-eNB and eNB-MME interfaces (Figure 4.13a and Figure 4.13b). Moreover, around 10% and 15% of the security processing load can be saved in UE and eNB, respectively (Figure 4.13c and Figure 4.13d).

Now, we examine the Data Protection Cost (DPC) when the DTP service is activated. We vary the average session arrival rate, λ_s, from 0 to 60/3600 per second. Fig 4.14a and Fig 4.14b show the variation of the TC and PC as a function of λ_s. As we expected, the DPC (i.e. TC and PC) increases linearly with λ_s. In



(a) Security Transmission Cost



(b) Processing Cost in network equipment

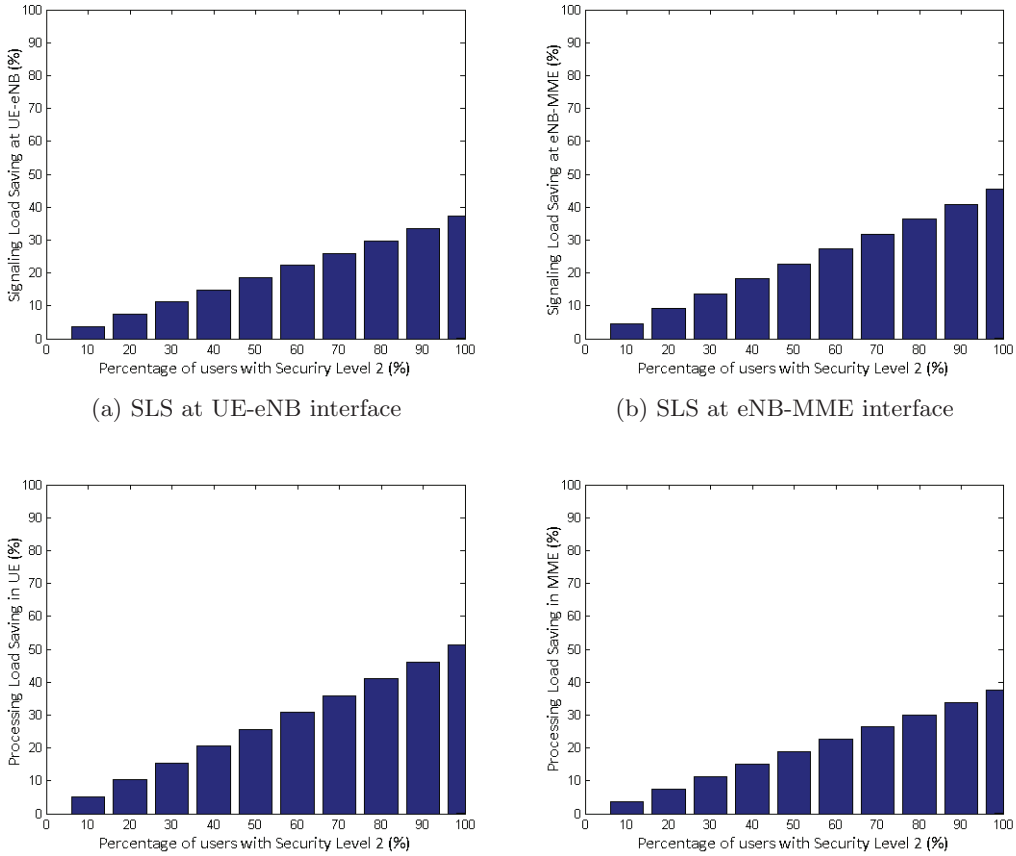
Figure 4.14: DPC in Scenario 2

fact, as λ_s increases, the average number of active sessions increases and hence the DPC increases. We note that SGW presents the highest PC during data delivery as it applies more security mechanisms on the ongoing sessions than eNB and PGW.

We assessed the DPC in two cases: DTP is activated for all sessions (blue lines) and DTP is activated for 80% of sessions only (red lines). We note that the security TC within the serving network is higher than that at the UE-eNB interface. This is due to the overhead imposed by the IPsec protocol. In fact, IPsec is a network layer protocol that protects traffic on a per connection basis between network equipment. It is independent from the nature of sessions that run above it. Deactivating the encryption and integrity protection mechanisms within IPsec for 20% of sessions reduced the transmission overhead (red line in Fig 4.14a). Similarly, the processing load in the network equipment is reduced when the encryption and

integrity protection mechanisms are not applied for 20% of sessions. It is clear that adapting the DTP service to the contextual information saves the processing load in network equipment and reduces the signaling load and transmission overhead within the serving network.

Scenario 3: Adaptive Privacy service



(a) SLS at UE-eNB interface

(b) SLS at eNB-MME interface

(c) PLS in UE

(d) PLS in MME

Figure 4.15: SLS and PLS in Scenario 3

In this scenario, we try to assess the impact of adapting the privacy service to the contextual information. In the current LTE/EPC network, the Privacy service is systematically ensured for each UE independent of the real needs. However, some UEs need neither protection for its signaling traffic nor for its subscription identity. For instance, static devices connect to the network just to get bandwidth and exchange no mobility-related signaling. For this kind of usage, it is needless to setup protection for the signaling traffic and subscription identity.

We assume that the Privacy service is deactivated in Security Level 2. Based on the contextual information, MME decides whether the Privacy service should be deployed. The percentage of UEs using Security Level 2 (i.e. with deactivated Privacy service) varied and the SLS and PLS results are depicted in Figure 4.15. We note that the deactivation of the privacy service for 30 % of subscribers reduces around 10 % of the security-related signaling load at both of UE-eNB and eNB-MME interfaces. In addition, 15 % and 10 % of the security-related processing load can be saved in UE and MME, respectively.

4.5 Conclusion

In this chapter, we analyzed the security services in LTE/EPC accesses. To evaluate the signaling, processing and transmission costs related to activation of these services, we made several simplifying assumptions. In fact, it was difficult to get the operational values of several parameters such as key generation cost. However, these assumptions does not impact our conclusions as we are comparing the results to a common relative references. From this evaluation, we showed that any added security measure introduces additional signaling and processing load. Moreover, the data traffic protection service incurs transmission overhead within the serving network and high processing load in network equipment during data delivery.

The security services in LTE/EPC accesses are designed to be activated in a systematic manner independently of the real needs. In the discussion section, we showed that adapting these services to the current context (e.g. session type, device type, etc.) can reduce the security costs. Therefore, we believe that security implementation in future access network architectures should offer greater adaptability without necessarily incurring high network operating costs. The multi-level property is a major challenge in future security implementations. Actually, one static security level (i.e. where all the security services are ensured) does not fit all user situations. Hence, the security implementation should be able to provide different levels of security depending on the contextual information. For instance, a connectivity with a simple authentication can be offered to the sensors that cannot support extensive cryptographic operations. On the other hand, a connectivity with a high security level (i.e. includes strong authentication and protects the data traffic) can be offered to the subscribers that send sensitive data.

In the next chapter, we recall the main connectivity requirements in future multi-access architectures before proposing a new architecture model for the network connectivity management.

Chapter 5

Context-Aware Connectivity Management (CACM) Model

5.1 Introduction

In previous chapters, we analyzed the connectivity management in multi-access 3GPP system and highlighted the related issues using several network usage scenarios. Particularly, we have been interested in the security and mobility services and the cost of their "always-on" activation. In addition, we presented various aspects of research related to the context-awareness paradigm. Nowadays, it has become imperative that access networks react to different elements of context by adapting the network connectivity behavior.

We propose in this chapter to introduce a network-based connectivity manager in the control plane of multi-access architectures. The main role of this entity consists in providing customized network connectivity for each subscriber according to the contextual information (e.g. profile information, mobility pattern, application requirements, the network resource usage pattern, etc.). The proposed solution aims to enable an efficient usage of network resources.

First, we recall the requirements that future multi-access architectures should fulfill. Then, we propose a network-based connectivity manager for multi-access architecture and provide a detailed description of the functional architecture. Finally, we evaluate the proposal functionally by applying it on the network usage scenarios that were introduced in Chapter 3.

5.2 Architecture Requirements

As was discussed in Chapter 3, the design of our connectivity manager should comply with the following requirements:

5.2.1 Context-aware connectivity (Req 1)

The network connectivity in multi-access architecture should be context-aware. In fact, the context-awareness paradigm is essential as it provides intelligence to multi-access architectures, allowing them to make appropriate and timely decisions on behalf of subscribers. The task of offering a global vision of the context is very complex and is not the purpose of this thesis. Even though contextual information can be theoretically limitless, we are primarily interested in any information that may influence network connectivity behavior. Thus, we define the context as ***"the set of personal and environmental states that either determines the network connectivity behavior or in which an update of the network connectivity should occur and is interesting to the user"***. We classified the contextual information into four categories:

- User-related context: enables the system to infer the user profile and states. It includes the user mobility pattern (static or highly mobile user), user activity (e.g. in meeting), device in use (characteristics and capabilities), traffic pattern, etc. Knowing user's mobility patterns can be used in various areas such traffic congestion control and network bandwidth provisioning [SDK+06].

- Network-related context: describes the network status and constraints. It includes static information such as network topology and network equipment capacities. Also, it includes dynamic information such as delay, bandwidth and network load.

- Application-related context: describes the application profile as well as its environment. The application profile refers to the application type (streaming or real-time), QoS requirements. The application environment describe where the application is running (e.g. in physical server or virtual machine). The application environment gives an idea about the dynamics of the application.

- Environmental context: describes the surrounding area. It includes user location (e.g. home, work, on bus, etc.), time of the day as well as the number of people in close physical proximity to the user.

Table 5.1 provides some examples of contextual information.

5.2.2 Adaptive network connectivity (Req 2)

The network connectivity should make use of the available context information to adapt itself dynamically regarding the context change. Actually, the system's changing environment provides a source of dynamism in future multi-access architectures,

Context Category	Context Elements
User-related	Location, Geographic position, Mobility status (static, highly mobile), User's speed (high, low), User's trip, User's preferences, User's profile, Terminal type (Smartphone, PC, iPAD, etc.), Energy consumption (alternate current (AC) input, battery life, etc.)
Network-related	Nearest AP features (IP address, trust level (trusted, untrusted), Load indicator (high, low), geographic position, AP load history), network congestion indicator, network load statistics
Application-related	Session type (Conversational, Streaming, Interactive, Background), Data sensitivity (high, medium, low)
Environmental	Time of the day, Geographic Map, Social Events

Table 5.1: Context Categories

when the same connectivity request may be processed differently depending on the situation.

As we showed in Chapter 3, one static connectivity where all the security and mobility mechanisms are systematically activated does not fit all network usage situation. For example, IPsec tunnel can be affordable for Smartphone or Computer, but it is heavy for sensors. Moreover, the systematic activation of network mechanisms impacts network performances. For example, Chapter 4 has shown that any added security mechanism introduces additional signaling and processing load. Several scientific papers have shown that the systematic activation of the mobility mechanisms induces signaling, processing and transmission costs (e.g. [LEC10], [AAOBL13] ,etc.). Obviously, when new network mechanisms are designed, network operators have to consider that these mechanisms will be activated in a systematic manner and thus limit several network parameters (e.g. number of simultaneous connections, bandwidth, dedicated CPU, etc.) to realize an acceptable overall network costs. However, if we have an architecture that adapt the network connectivity behavior in each situation, the problem of network costs does not arise in the same way.

Therefore, future multi-access architectures should be able to provide a network connectivity with different security levels and mobility profiles to improve network performances. Each security level is associated to a given number of security services/mechanisms. Similarly, each mobility profile includes a given number of mobility services. Future multi-access architecture will be in charge of selecting the adequate security level and mobility profile depending on the contextual information. To this end, the network connectivity should be modular where network mechanisms can be easily orchestrated, activated, deactivated and configured.

In addition, within a given network connectivity, the network resources in use and the maintained states should be adjusted according to the sessions needs and to the network status.

Table 5.2 provides some examples of context elements and the corresponding adaptations that could be decided in the network.

Context Element	Possible action
User-related context	
Location (office), Geographic position	Find the adequate NAGw and NSGw
Mobility status (subscriber will be on move in 5/10 mn), subscriber's trip (from office to home), Movement velocity (by bus)	Perform the initial authentication and prepare the handover (pre-authentication/ context transfer / generate a ticket)
Terminal type (smartphone), Power supply (battery lifetime)	Choose the security mechanisms (authentication method, encryption and integrity protection algorithms) that consume less energy
Network-related context	
Available access (LTE access, office wifi access), network load indicator	Choose the access that offers the better QoS
Application-related context	
Session type (conversational)	Optimized fast authentication schemes if mobility
Data sensitivity: session already secured (e.g. VPN session)	Encryption mechanism is deactivated for the data traffic, Encryption and integrity protection mechanisms are activated for the signaling traffic
Environmental context	
Time of the day (5 PM)	Select the next serving AP that will not be under heavy load

Table 5.2: Context Elements and Possible actions

5.2.3 Network-side adaptation decision (Req 3)

The network connectivity adaptation in multi-access architectures should be decided at the network side. The network mechanisms that will be activated should be selected and orchestrated by a trusted network entity. The subscriber may assist the decision by providing the required contextual information.

5.2.4 Unified connectivity management (Req 4)

As motivated in Chapter 2 and Chapter 3, a large number of network mechanisms have been proposed and adopted in current access networks to provide specific features or to overcome specific deficiencies. To be able to combine different mechanisms together, a unified connectivity management is required. This allows multi-access architecture to harmonize network mechanisms across the multiple access technologies.

Instead of designing for each new deployed access technology dedicated control functions as is commonly done, multi-access architecture should present a common control plane that is able to support the integration and cooperation of different technologies. Therefore, an architecture with unified connectivity management enables the interconnection of the different access technologies closer to users and avoids, therefore, the functional redundancy.

5.2.5 Flexible use of network resources (Req 5)

The dynamic nature of the current environment requires a dynamic network connectivity that ensures the flexible use of network resources. In fact, network resource availabilities may vary as new connections are added and older connections terminate. Suppose, for example, that a subscriber requests a reliable and secure connectivity. In a particular situation such as limited network resources, this kind of connectivity cannot be ensured, but less secure connectivity can be provided. A static policy could reject this request because the required security level cannot be ensured. However, even a less secure connectivity might be good enough to transport the subscriber's traffic.

Therefore, future multi-access architecture should provide a network connectivity with a security level and mobility profile adaptable to situations with a scarce resources. At the same time, the same connectivity should be able to evolve and provide a strong security and mobility guarantees when more resources are available.

5.3 Context-Aware Connectivity Management (CACM) Architecture

5.3.1 Overview

In this section, we propose a Context-Aware Connectivity Management (CACM) architecture to address the above requirements. The CACM is a control framework that provides the capability to orchestrate a set of network services within a network connectivity, running on one or more network equipments. We recall that a network connectivity is formed by a specific combination of network services. A brief overview

of the CACM architecture is presented in this section (Figure 5.1), with its key functional components detailed in further sections.

In order to achieve the objective in having context-aware network connectivity (i.e. **Requirement 1**), a built-in context management system or an interface to external context management system is required. This system provides multi-access architectures with the required parameters to decide the adaptation that should be done in the network connectivity. Therefore, our control plane is made up of two logical subsystems (see Figure 5.1): *Context Management Subsystem (ContextMS)* and *Connectivity Management Subsystem (ConnectMS)*. Together these two subsystems provide the policy-based infrastructure to enable adaptive network connectivity in multi-access architecture.

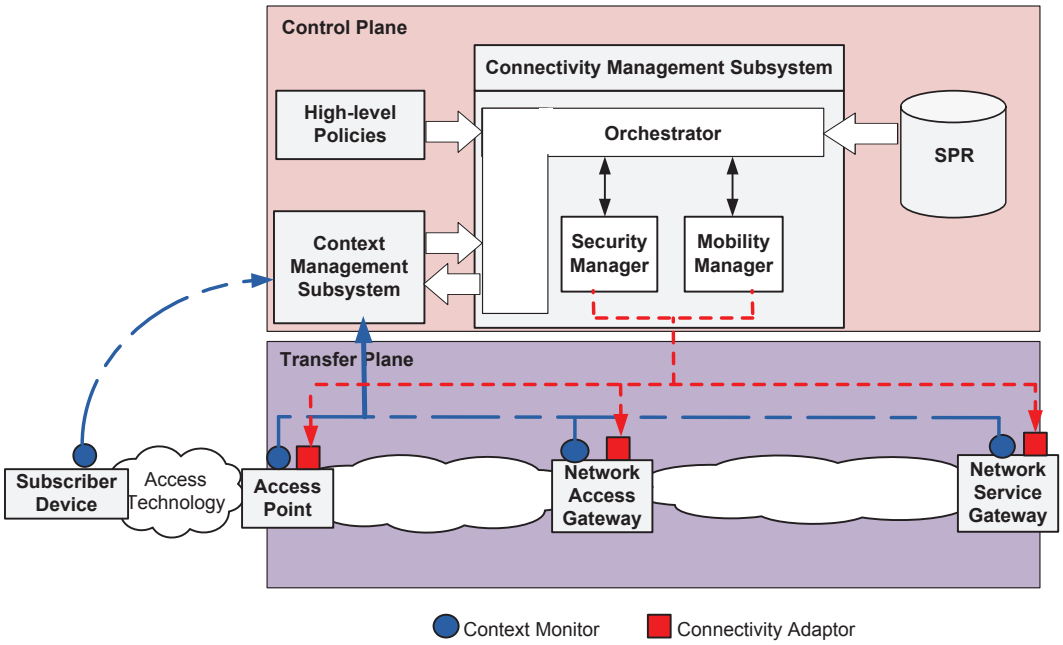


Figure 5.1: Context-Aware Connectivity Management Model

The *ContextMS* is a principal requirement to build context-aware network services as it is responsible for collecting and processing the contextual information. The ContextMS should present the contextual information in an intelligible format for the ConnectMS modules.

The *ConnectMS*, in turn, is responsible for setting up, monitoring and updating the network connectivity (i.e. **Requirement 2**). The high-level policies govern the behavior of the ConnectMS modules. The ConnectMS is designed to be generic and not specific to an access technology. This enables the ConnectMS to connect and manage multiple access at the same time (i.e. **Requirement 4**).

The control plane includes also a Subscriber Profile Repository (SPR). This database acts as a central database of all subscriber-specific information such as access restrictions, the subscribed QoS profiles and the subscriber preferences. The HSS is a typical example of the SPR.

In our proposal, the CACM considers the packet flows separately (i.e. it decides for each packet flow the adequate network services). We adopt the same definition as in [RAJC11] where the packet flow is defined as a sequence of packets that are sent from a particular source to a particular destination. These packets related to the same flow can be identified, at the data plane, with a 5-tuple (i.e. source IP address, destination IP address, source port number, destination port number, and the protocol in use).

The *Orchestrator* is the main module in the ConnectMS. It is responsible for receiving and processing connectivity requests. In addition, it is able to map the high-level policies (i.e. represents business-level objectives) to lower-level operational policies that are intelligible by the network equipment at the transfer plane. The Orchestrator includes several specialized modules that generate decision regarding the activation of specific network mechanisms (i.e. **Requirement 3**). For instance, *Security Manager* and *Mobility Manager* modules generate decisions regarding security and mobility mechanisms, respectively.

Upon receiving a connectivity request, the *Orchestrator* triggers the Security and Mobility Managers to specify an ordered list of security and mobility mechanisms that should be activated. This specification takes into consideration high-level network policies, contextual information from the ContextMS and user profile from the SPR. For example, the Security Manager may decide to deactivate Data Traffic Protection service for flows that are already protected such as those related to VPN sessions.

The *Orchestrator* launches the decided network mechanisms. Moreover, it configures the network connectivity parameters (e.g. tunnel timers) in accordance with high-level network policies and contextual information. For instance, based on the application profile and the user-related context state, the ConnectMS can predict the user IDLE period value during which the network connectivity is maintained at the transfer plane. Therefore, instead of having static IDLE timer configured locally in network equipments, our *Orchestrator* decides an IDLE timer value adapted to the application profile and, then, configures network equipments with this value.

When the context change, the *Orchestrator* adapts the network connectivity to this change by triggering the adequate modules to make the right decision. For example, when the subscriber traffic passes from unprotected (e.g. HTTP) to secured (e.g. HTTPS), the Context Monitor in the NAGw sends an alert to the Orchestrator. As the alert is related to security, the Orchestrator triggers the Security Manager to decide the adequate action (e.g. deactivate the DTP service). After that, the

Security Manager executes the decided action (e.g. stop the encryption and integrity protection mechanisms in the transfer plane). In Chapter 6, we implemented a test-bed that reproduces this example.

In the Transfer plane, we will need the presence of: *Context Monitor* to provide the ContextMS with up-to-date contextual information, and *Connectivity Adaptor* to enable the Orchestrator to adapt the network connectivity according to the decided policies.

The *Context Monitor* controls and registers any change in the network equipment status, environment or user context. Examples of context monitors are WLAN access points (APs) that provide information about the network load status or the flow status. The Context Monitors may be a set of sensors that gather information about temperature, user mobility patterns, travel trajectory/destination, planned activities in users' personal calendars, or try to identify people in the same meeting room. The ContextMS configures Context Monitors to control specific flows and relay any change in the flow status.

The *Connectivity Adaptor* is an active entity listening for commands sent by the ConnectMS and communicating with software and hardware resources for the realization of network policies. This entity is responsible for adapting and configuring managed network equipments such as access points, routers, or gateways. While the decision process can be conceived as a centralized module in the considered architecture, the Connect Adaptors are typically distributed.

The Context Monitor and the Connectivity Adaptor collaborate to ensure the flexible use of network Resources (i.e. **Requirement 5**). For instance, upon detecting an overload situation, the Context Monitor relays the information to the ContextMS. This latter triggers the ConnectMS to make the right decision. In such case, the ConnectMS may decide to free the overloaded network equipment by moving some flows to other available network equipments.

5.3.2 Context Management Subsystem (ContextMS)

In order to make the contextual information available for the *ConnectMS*, a *ContextMS* is required. Several works have developed context-aware systems that collect, process and provide the contextual information [DA00] [IOMS10] [TL11]. The ContextMS collects network-related information in real-time from Context Monitors that are deployed throughout access networks. Also, it gathers user-related information from sensors that are installed in user's surrounding and application-related information from the Application Providers. After that, the ContextMS classifies, analyzes, processes and generates a value-added information that can be delivered to the ConnectMS.

The ContextMS architecture is not the purpose of our study, but it may in-

clude the following modules: context acquisition, context processing and context publishing modules.

- Context acquisition: is responsible for gathering the required contextual information from different context sources such as user devices, network equipments, and other sensors.
- Context processing: is responsible for analyzing and aggregating the raw contextual data that are collected from various entities. This module will infer a situation from the raw data. It requires reasoning and inference methods to infer higher level information from lower level contextual information. This module should be able to represent the contextual data in a format that is easy to understand by other entities.
- Context publishing: is responsible for publishing the contextual information to the ConnectMS modules. The Mobility Manager may subscribe to the network load information. Therefore, the context publishing module will alert the Mobility Manager about any change in the network load.

Prediction-based approaches are of key relevance for context-aware and recommendation systems. Having such approaches enables the ContextMS to determine several dynamic contextual information such as subscriber mobility and traffic patterns. In fact, [DGP12] showed that the human mobility pattern can be learned from contextual variables. For instance, [SK05] presents a mechanism that predicts user mobility, traveling trajectory and destination using static contextual information such as user preferences and goals, and spatial conceptual maps. Therefore, we assume that the ContextMS includes a mechanism for gathering contextual information, performing online predictions, and distributing relevant prediction results to the ConnectMS.

5.3.3 Security Manager

The analysis in Chapter 4 showed that any added security mechanism introduces additional signaling and processing load. In this section, we propose a Security Manager (SM) module (see Figure 5.2) that makes a trade-off between the activated security services and the amount of the security signaling and processing load. For example, taking into consideration that the energy consuming cryptographic computations such as encryption and keys generation will deplete batteries and make sensors out of service rapidly, the SM module can deactivate the encryption and integrity protection mechanisms for sensors with non-sensitive data. This kind of adaptation is advantageous for the access network as it saves at least the processing load associated to security keys computations.

The SM main task consists in determining the security services to be ensured that corresponds to the current context. Obviously, the SM should verify that these services can be offered within performance and CPU resource availability constraints. In addition, the SM should verify that adapting the security services to the contextual information does not introduce new security threats.

We define the security level as the set of security services that should be ensured for a given connectivity depending on the contextual information. For each IP flow, the Security Manager matches the required security level and actives the appropriate security mechanisms accordingly. Table 5.3 provides a non-extensive list of security levels that can be implemented in future multi-access architectures.

Security Level	Active Security Services
level 1	No security services
level 2	Identification and Authentication
level 3	Identification and Authentication Access control
level 4	Identification and Authentication Access Control Privacy
level 5	Identification and Authentication Access Control Data Traffic Protection
level 6	Identification and Authentication Access Control Data Traffic Protection Privacy

Table 5.3: Possible Security Levels.

The SM functional architecture is shown in Figure 5.2. It comprises a multi-criteria **Security Decision-Maker** process to decide the proper security level according to the current context and the corresponding mechanisms (e.g. select the login/password mechanism to ensure the authentication service). In the literature, we find several efficient algorithms for multi-criteria decision-making such as Fuzzy logic [ZJZ10] or Analytic hierarchy process [JIFW06]. The Security Decision-Maker component varies security levels to remain within user-specific range, while adapting to changing context (e.g. changing location or CPU resource availabilities).

The SM or the Orchestrator should inform subscribers with the decided security level related to their current network connectivity. For example, the SM may decide

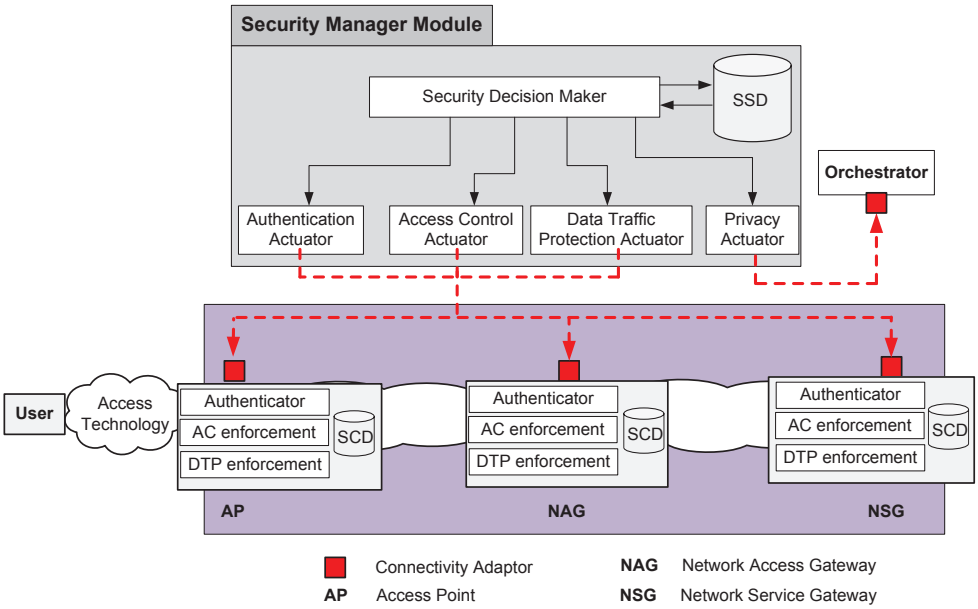


Figure 5.2: Security Manager

to temporarily deny data traffic protection service at some access points for energy saving (i.e. according to network operator policies). In that case, the operator should notify their subscribers. Therefore, the subscribers can take the necessary precautions such as activating security at application or IP layer.

The SM module includes also a set of **actuators** which represents the enforcement points related to different security services. The actuators take as input the security actions decided by the Security Decision-Maker process to drive their enforcement, by communicating with the adequate Connectivity Adaptors situated in the network equipment at the transfer plane. As the security actions may involve different layers, the security actuators interact via proper interfaces with the different connectivity adaptors.

Regarding the privacy actuator, the corresponding Connectivity Adaptor is located in the Orchestrator. Recall that the privacy service includes the generation of temporary identity (i.e. to protect the subscriber permanent identity). It includes also the security algorithm selection and the security keys generation (i.e. to protect the signaling exchange). First, the privacy actuator should set the security algorithm and keys in the orchestrator to protect any signaling exchange between the subscriber and the Orchestrator. Then, the privacy actuator relays the temporary identity to the Orchestrator. This latter sends this identity to the subscriber.

According to the selected level, the Security Decision-Maker triggers the appropriate actuators. For example, the user request a network connectivity for a VPN

session with the Intranet. The Security Decision-Maker component decides to deactivate the Data Traffic Protection service for this kind of use (i.e. security level 4). Thus, the security action is *"deactivate encryption and integrity protection at the transfer plane"*. The Security Decision-Maker communicates its decision to the Data Traffic Protection actuator. The latter enforces the decided security action in the transfer plane by asking the appropriate Connectivity Adaptors to stop encryption and integrity protection processes.

The SM module includes a **Security Status Database (SSD)** that maintain the security services status of each packet flow. In fact, the Security Decision-Maker queries this database to know the current activated security services and, therefore, decides whether any adaptation is required. In the transfer plane, the network equipments includes a Security Context Database (SCD) where the security parameters (e.g. security algorithms and keys) for a given packet flow are stored.

5.3.4 Mobility Manager

The Mobility Manager (MM) module main task consists in selecting and activating the mobility services that fulfill the connectivity requirements. For example, the MM may decide to just ensure the reachability service for a camera mounted on a bus. Therefore, the camera can be reached by the public transport staff at any time. In addition, as was motivated in Chapter 3, the SIP sessions can survive without having the session continuity service at the access network. Therefore the MM module can deactivate the session continuity for such sessions. The MM tasks include also the selection of the adequate mobility anchor in the transfer plane that corresponds the current context. For example, for a highly mobile subscriber, the MM can anchor the sessions in the NSGw entity to reduce the mobility anchor relocation.

We define the mobility profile as the set of mobility services that should be ensured for a given connectivity depending on the contextual information. Therefore, for each IP flow, the MM matches the adequate mobility profile and actives the appropriate security mechanisms accordingly. Table 5.4 provides a non-extensive list of mobility profiles that can be implemented in future multi-access architectures.

The MM functional architecture is shown in Figure 5.3. The MM module has the same principle as the SM module. The **Mobility Decision-Maker** is the main component in the MM module. This component takes into consideration the contextual information to decide the mobility profile that should be ensured for a given connectivity.

The **Mobility Status Database (MSD)** maintains the mobility profiles that were specified for the ongoing network connectivity.

Similarly to the SM module, the MM module includes a set of **actuators** that acts as the enforcement points for the mobility service. In fact, these actuators com-

Mobility Profile	Active Mobility Services
Profile 1	No mobility services
Profile 2	Nomadism
Profile 3	Reachability
Profile 4	Nomadism Reachability
Profile 5	Reachability Session Continuity
Profile 6	Nomadism Reachability Session Continuity

Table 5.4: Possible Mobility Profiles.

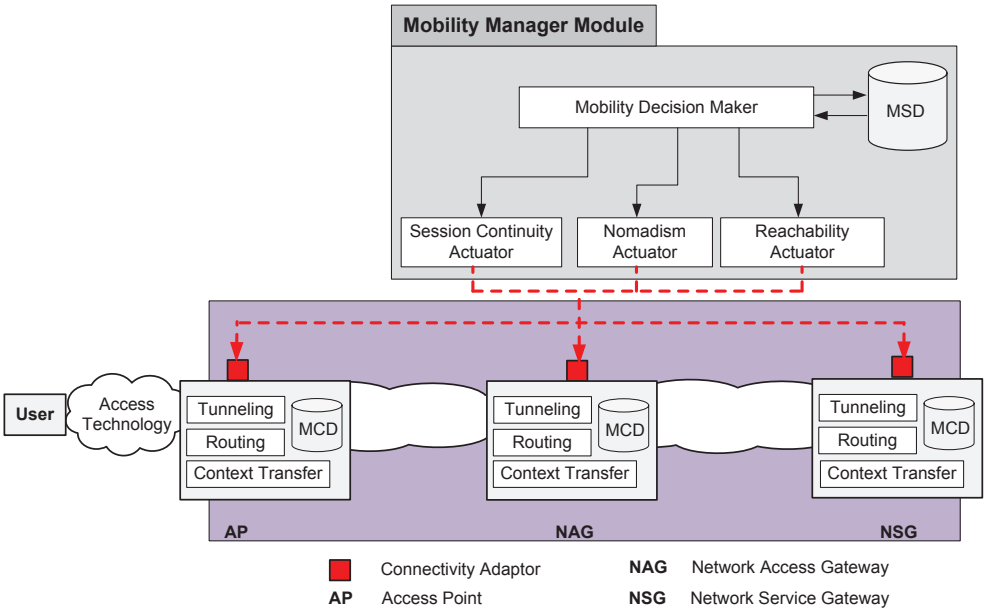


Figure 5.3: Mobility Manager

municate with the adequate Connectivity Adaptors to execute the decided adaptations (e.g. activate/deactivate a given mobility mechanism). In addition, they are in charge of configuring the decided connectivity parameters such as the connectivity lifetime and IDLE timer. Moreover, these actuators are responsible for transferring the mobility context from old Connectivity Adaptor to the new one. Actually, each Connectivity Adaptor manages a **Mobility Context Database (MCD)** where the mobility contexts are maintained (e.g. tunnel identifier, IDLE timer, etc.).

5.4 Qualitative Evaluation

A way to check whether our proposal fulfill the architecture requirements is to use network usage scenarios. Therefore, in this section, we revisit the network usage scenarios that were introduced in Chapter 3 and analyze the behaviour of the proposed architecture in each scenario.

Context	Multi-access 3GPP system approach	CACM approach
SIP-based session through LTE	The DTP service is systematically ensured.	The ConnectMS deactivates the DTP service because the SIP-based session is already secured with the IPsec tunnel. Therefore, the network equipments in the transfer plane do not apply the encryption and integrity protection mechanisms.
	The Session Continuity service is natively ensured for any types of session.	The ConnectMS decides to not ensure the Session Continuity service as the SIP protocol is able to manage any change in the IP address. Therefore, the handover, context transfer and data forwarding mechanisms at the LTE/EPC access are deactivated.
SIP-based session through Untrusted WiFi	All the subscriber sessions goes through the IPsec tunnel between the subscriber device and the ePDG. Therefore, all security services are ensured for all sessions. In addition, all the mobility services are ensured as the sessions are anchored to the PGW.	The ConnectMS decides to not ensure the security and mobility services at the access network level. Therefore, the SIP-based session uses Connection 3 (see Figure 3.4) where the subscriber connects directly to the IMS platform.

Table 5.5: Scenario A - Comparison between multi-access 3GPP system and the CACM approaches.

Scenario A highlighted the systematic activation of security and mobility mech-

anisms without considering the real connectivity needs. This kind of behavior leads to functional redundancy. In the proposed architecture, such problem is avoided as the ConnectMS takes into consideration the different contextual information before deciding the mechanisms to be activated for a given connectivity. In this scenario, we assumed that the subscriber is in the office and will not move for a while (i.e. the UE mobility historic). We assumed also that we have a SIP-based session. By having the knowledge about the UE mobility historic and the application nature, the ConnectMS decides to deactivate the security and mobility mechanisms at the access network. Table 5.5 provides a comparison between the current approaches and the CACM approaches related to this scenario.

Scenario B underlined another type of functional redundancy when the subscriber uses two different accesses simultaneously. In this scenario, the same network mechanisms are performed in each access. The proposed architecture overcomes this issue by having a unified connectivity management. In fact, the ConnectMS is generic and not specific to an access technology. On the contrary, it is able to manage multiple accesses at the same time avoiding, then, the functional redundancy. For instance, knowing that the subscriber has been authenticated in one access, the ConnectMS will not repeat the authentication in the second access. Moreover, the ConnectMS specifies the connectivity requirements in terms of network services and, then, selects the corresponding mechanisms that are in adequacy with the access in use. Table 5.6 compares between the multi-access 3GPP system and the CACM approaches related to this scenario.

Context	Multi-access 3GPP system approach	CACM approached
Sessions through LTE and Trusted WiFi	The authentication and access control (i.e. EU profile download) mechanisms are executed twice for the same subscriber.	The ConnectMS authenticate the subscriber in LTE and download the related UE profile. Then, the ConnectMS decides to not perform the authentication at the Trusted WiFi. In this case, the ConnectMS can use the Generic Bootstrapping Architecture (GBA) technology [3GP10b] to avoid the unauthorized use of the access network.

Table 5.6: Scenario B - Comparison between multi-access 3GPP system and the CACM approaches.

Scenario C showed that several connectivity parameters in the transfer plane (e.g. timers) are configured locally with constant values. These values are not

adapted to all types of applications. In the proposed architecture, the ConnectMS is able to decide an adapted values for the connectivity parameters that corresponds to the application profile. For instance, for Always-ON application, the ConnectMS may decide to maintain the network connectivity (i.e. tunnels and network parameters) within the access network Transfer Plane and to deactivate the IDLE timer. This results results in reducing the signaling costs related to the tunnels release (see Annex **??**). Then, the ConnectMS enforces the decided value in the transfer plane via the Connectivity Adaptors. Table 5.7 compares between the multi-access 3GPP system approach and the CACM approach related to this scenario.

Context	Multi-access 3GPP system approach	CACM approach
Applications through LTE	The IDLE timer is configured locally in the eNB with a static value for all kinds of applications.	The ConnectMS decides the adequate value for the IDLE timer based on the application pattern. Then, it enforces this value in the concerned network equipments.

Table 5.7: Scenario C - Comparison between multi-access 3GPP system and the CACM approaches.

Context	Multi-access 3GPP system approach	CACM approach
Unpredicted overload situation in the SGW	The MME is unaware of the situation and keeps allocating the new incoming sessions to this SGW.	The ConnectMS is alerted about the situation by the Context Monitors . Thus, it decides to allocates the new incoming sessions to other NAGws. In addition, it decides to temporarily free the overloaded NAGw by moving the delay-tolerant sessions to other NAGws.

Table 5.8: Scenario D - Comparison between multi-access 3GPP system and the CACM approaches.

Scenario D emphasized the need for a network connectivity that uses the network resources in a more flexible way. By taking into consideration the contextual information and particulary the network-related context, the ConnectMS is able to use network resources in an efficient manner. For instance, upon detecting that a

given NAGw is overloaded, the ConnectMS decides to free momentary this NAGw by transferring delay-tolerant sessions to other NAGws. Chapter 7 proposes an example of implementation that addresses this scenarios. Table 5.8 compares between the multi-access 3GPP system approach and the CACM approach related to this scenario.

Finally, Table 5.9 summarizes how the CACM model ensures the architecture requirements.

Req 1	The ContextMS provides contextual information to the ConnectMS enabling, then, a context-aware connectivity
Req 2	The Orchestrator and the decision-maker modules (e.g. Security and Mobility Managers) decides and orchestrates the network mechanisms to be activated. This decision takes into consideration the connectivity requirement within the specific context
Req 3	The Orchestrator and the decision-maker modules are located in the network side.
Req 4	The ConnectMS is generic and not specified to a specific access. This can ensure the unified management of the network connectivity in multi-access architectures
Req 5	When the network-related context (e.g. congestion indicator) changes, the ConnectMS adapts the network connectivity accordingly. This ensures the flexible use of network resources

Table 5.9: Requirements and how they are addressed in the CACM model

The CACM model presents a modular connectivity where the network mechanisms are activated/deactivated according to the real needs. In addition, the CACM model enables flexibility as it adapts the network connectivity to the current situation. However, such a model may face several challenges that should be studied. Among these challenges, we cite:

- Mutiple technologies and mechanisms in the same architecture: The CACM model does not only select the network services to be activated, but it also selects the adequate corresponding network mechanisms. Therefore, for the same network service, the access network architecture should implement various network mechanisms. Such architecture may grow in complexity on a large-scale.

- Privacy implication: In the CACM model, the subscriber should assist the connectivity manager by pushing several user-related contextual information to the Context Management Subsystem. However, such behavior can be accepted depending on the country as it raises privacy issues. In fact, some subscriber

personal information (e.g. subscriber activities, mobility pattern, etc.) may be disclosed. Therefore, the proposed model should implement the required mechanisms to protect the subscriber privacy (e.g. the communication between the subscriber device and the Context Management Subsystem).

- Security consideration: In the CACM model, the Security Manager (SM) activates/deactivates security mechanisms according to the contextual information. However, the SM should check that the deactivation of a given security mechanism will not introduce new security threats and impact, therefore, the overall network security.

- Level of granularity: In the CACM model, it is established that the decisions taken by the SM and MM modules are per-flow granularity (i.e. they decide the required security and mobility mechanisms for each flow separately). Recall that a given network connectivity can include several flows. However, in the current architectures, the security and mobility mechanisms are activated/deactivated for all flows of the same connectivity without exception. For example, when the encryption and integrity protection mechanisms are activated in the network connectivity, they are applied for all flows. Therefore, to implement our proposal, the Connectivity Adaptor should install filters inside the network connectivity to separate flows and apply the corresponding network policies each time. Even more, the Orchestrator may decide to establish several simultaneous connections for the same subscriber. The flows with the same needs are bundled in the same connections. However, it may be expensive for network operators to install these filters or maintain several simultaneous connections of the same subscriber. Therefore, the SM and MM modules may adjust the granularity level according to the network resources availability and network costs.

5.5 Conclusion

In this chapter, we proposed a connectivity management framework (CACM) for multi-access architectures. Then, a detailed description of the functional architecture was provided. The proposed architecture was evaluated through network usage scenarios.

The CACM is designed to fulfill the architecture requirements that were highlighted in Chapter 3. First, the CACM is context-aware as it selects, orchestrates and launches security and mobility mechanisms according to the contextual information. Secondly, the CACM is adaptive as it decides the most appropriate actions (e.g. activating/deactivating network mechanisms, move flows from an equipment

to another, etc.) when a change in the context occurs. Thirdly, the adaptation decision are taken at the network-side. Fourth, the CACM is able to manage multiple accesses at the same time and reduces, therefore, the functional redundancy when the subscriber uses several access simultaneously. Finally, the CACM is designed to use network resources efficiently as it is aware of the network-related context.

In the following chapters, we propose two concrete applications of the proposed model. Chapter 6 presents an experimental test-bed that implements adaptive Data Traffic Protection (DTP) service in Untrusted access. Chapter 7 proposes an implementation of adaptive mobility services using the OpenFlow protocol.

Chapter 6

Implementing Adaptive DTP service in non-3GPP access

6.1 Introduction

To demonstrate the feasibility of our proposal, we have implemented a prototype that reproduce the trusted and untrusted non-3GPP accesses with WiFi as access technology. Using this prototype, we have developed and experimented with a novel security control where the Data Traffic Protection (DTP) service is activated/deactivated according to the application requirements.

First, we give the test bed goal. Then, we propose mechanisms that enable adaptive DTP service in trusted as well as untrusted non-3GPP accesses. After that, we validate our proposal on an experimental test bed. We end the chapter by giving the challenges that we met during implementing this test bed.

6.2 Motivation

The aim of the test bed is twofold: create an experimental platform that reproduces the trusted and untrusted non-3GPP access and implement an adaptive DTP service. Through this test bed, we demonstrate the gain that access network can achieve by contextualizing the Data Traffic Protection (DTP) service - activating/deactivating the service according to the session type.

The test bed is composed of a AAA server, trusted Access Point (T-AP), untrusted Access Point (U-AP), access router (AR) and a client as shown in Figure 6.1. The authentication and access control in the T-AP follows the 802.1X protocol. The U-AP is an open access point. The AR acts as an ePDG for untrusted access networks and as a WAG for trusted access networks. The red boxes represents the functions that we intend to implement in the test bed to realize the adaptive DTP

mechanism. The test bed components as well as the adaptive DTP mechanism will be detailed in the following sections.

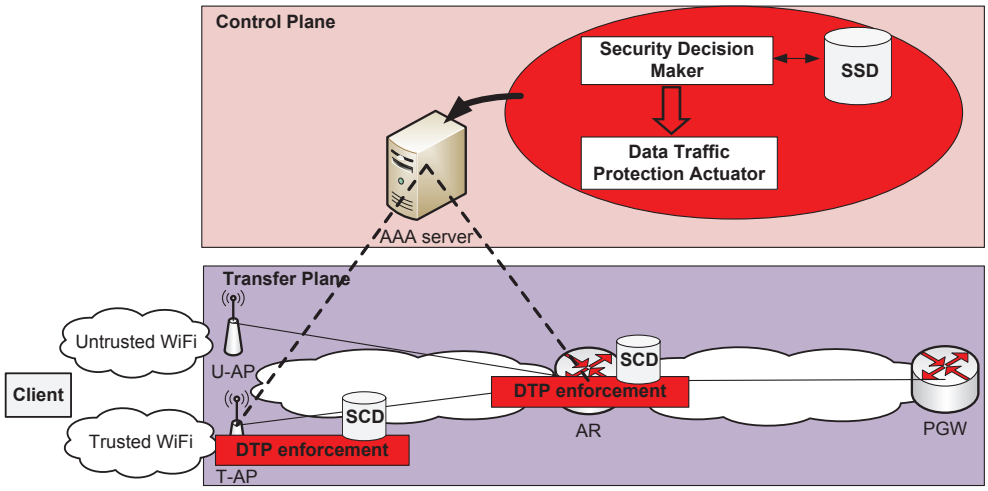


Figure 6.1: Architecture to be tested

Our test bed address the following scenario. The network operator offers three options:

- Option 1: connectivity where the DTP service is "Always ON" independently of the application requirements.
- Option 2: connectivity where the DTP service is "Always OFF" independently of the application requirements.
- Option 3: connectivity where the DTP service is activated only when needed. For instance, the service is deactivated when the traffic is already secured.

The subscriber can select and subscribe to one of these options. In Option 1, the DTP service is activated upon signing the contract with the operator and stays activated for all subscriber traffic. Similarly, in Option 2, the DTP service is deactivated upon signing the contract with the operator and is never ensured. In both of these options, there is no additional procedures when the subscriber is using the access network.

In Option 3, the network operator offers an adaptive DTP service. In this option, the access network should detect the application requirements and acts according to these requirements (i.e. DTP is deactivated when the application is already secured such as HTTPS traffic or VPN sessions).

6.3 Adaptive DTP service specification

To realize the above scenario, we create three subscriber groups in the AAA server: *(Gr 1) DTP always ON*, *(Gr 2) DTP always OFF*, and *(Gr 3) DTP when needed*. In addition, we defined two modes of connectivity that could be ensured for the subscriber in the data plane: eap_clear and eap_ciphered. In the first mode, the DTP service is deactivated. In the second one, the DTP service is activated.

First, we present the test bed functional architecture. Then, we specify mechanisms that should be added to the test bed to enable adaptive DTP service.

6.3.1 Test bed functional architecture

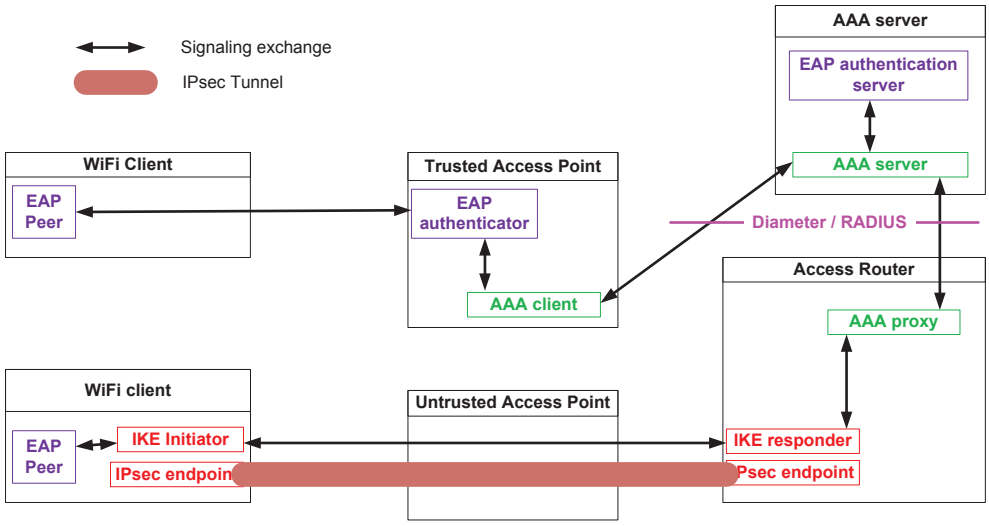


Figure 6.2: Test bed functional architecture

The test bed is comprised of a set of functional blocks as shown in Figure 6.2:

- WiFi Client: represents the User Equipment (UE). It includes the EAP peer IKE initiator, and IPsec endpoint components. In 802.1X, the authentication mechanism is based on the EAP protocol. Therefore, the EAP peer component is required in this block. The IKE initiator and IPsec components are required for the IPsec tunnel establishment in untrusted accesses.

- Trusted Access Point (T-AP): is the first point of contact of the subscriber with trusted access networks. As the authentication mechanism is based on the EAP protocol, an EAP authenticator component is required. The AAA client component is responsible for the communication between the T-AP and the AAA server.

- Untrusted Access Point (U-AP): represents an open access point. No special components are required in this AP.
- AAA server: is responsible for the subscriber authentication. Therefore, it includes the EAP authentication server and the AAA server components.
- Access Router (AR): is considered as the main functional block in our test bed. The IKE responder and IPsec endpoint components are required in the AR to host subscribers from untrusted accesses. The AAA proxy relays the EAP authentication exchanges to the AAA server.

The communication between the WiFi Client and the T-AP or the U-AP is based on the WiFi technology.

6.3.2 Adaptive DTP service in trusted access

In the trusted WiFi access, we target to activate the encryption and integrity protection mechanisms (i.e. DTP mechanisms) on the WiFi link according to the subscriber group.

During the subscriber authentication procedure, the T-AP retrieves the subscriber profile from the AAA server to determine the subscriber group. If the subscriber belongs to *Gr 1*, the T-AP activates the DTP mechanisms on all subscriber data traffic. If the subscriber belongs to *Gr 2*, the T-AP deactivates these mechanisms for the whole subscriber data traffic.

For subscribers belonging to *Gr 3*, the DTP mechanisms are initially activated (i.e. the eap_ciphered is the default mode). When the T-AP detects that the traffic is already secured, it deactivates these mechanisms on the WiFi link (i.e. move to eap_clear mode). Similarly, when the T-AP detects that the traffic need protection and the DTP mechanisms are deactivated, it activates them. To detect the traffic security status, a traffic inspection mechanism is required in the AR. In addition, to maintain the DTP service status (i.e. activated or deactivated) in the current network connectivity, a local database should be added to the AR. This database is targeted to maintain the DTP service status for the current connectivity and, thus, acts as the SSD database in the CACM model.

First, we give details about the network connectivity establishment in trusted accesses. Then, we propose a DTP service adjustment mechanism that is required for subscribers of *Gr 3*.

6.3.2.1 Network connectivity establishment

In trusted accesses, there are two alternatives for relaying the authentication exchanges between the T-AP and the AAA server:

- The AAA client in the T-AP communicates with the AAA server directly.

- The AAA proxy in the AR relays the authentication exchanges between the AAA client in the T-AP and the AAA server.

In our case, the second alternative is more appropriate as the AAA server should notify the AR about the subscriber profile (*Gr 1*, *Gr 2* or *Gr 3*). In addition, the AR is responsible for supervising the subscriber traffic and deciding the activation/deactivation of the DTP mechanisms. Therefore, the AAA proxy in the AR uses the authentication exchanges to relay the subscriber profile and the DTP service status to the T-AP. Thus, the T-AP just determines and applies the corresponding connectivity mode.

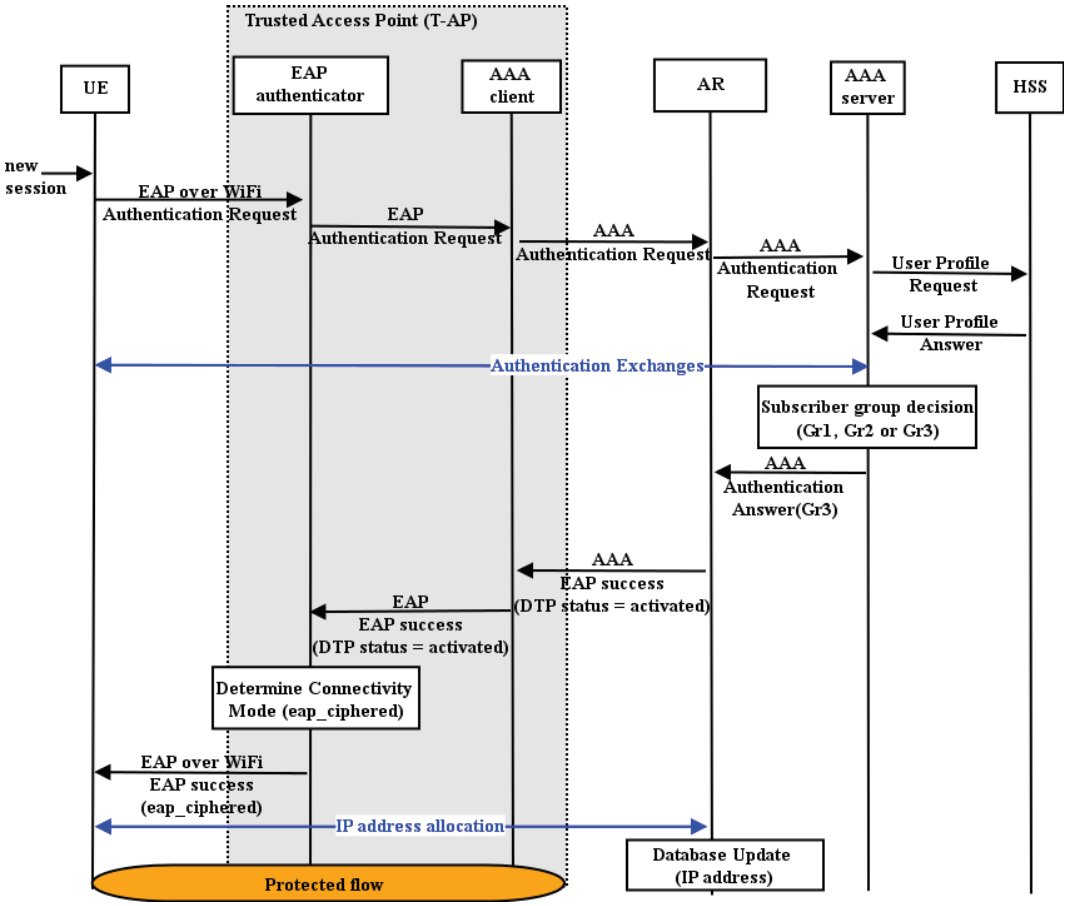


Figure 6.3: Connectivity Establishment flow chart

The flow chart related to the network connectivity establishment for a new session in trusted accesses is shown in Figure 6.3. The network connectivity is established as follows:

1. The User Equipment (UE) initiates the network connectivity establishment by sending the EAP request message to the T-AP as shown in Figure 6.3.

2. Inside the T-AP, the EAP authenticator relays the EAP request to the AAA client. This latter relays the EAP request to the AAA proxy in the AR.

3. After successful authentication, the AAA server decides the subscriber group (i.e. *Gr1*, *Gr2* or *Gr3*) to which the UE belongs. Then, it provides the AAA proxy in the AR with the decided subscriber group. This enables the AR to decide whether the DTP service should be activated or deactivated. For subscriber of *Gr 1*, the AR decides to always activate the DTP mechanisms. Similarly, for subscriber of *Gr 2*, the AR decides to always deactivate the DTP mechanisms. In case the subscriber belongs to *Gr 3*, the AR retrieves the DTP service status (i.e. activated or deactivated) related to this subscriber from the local database. If there is no entry for this subscriber, the AR inserts a new entry in the database and associates the default status (i.e. activated).

4. The AAA proxy in the AR relays the DTP service status to the AAA client in the T-AP. This latter relays the DTP service status to the EAP authenticator in the same T-AP. Based on this information, the EAP authenticator determines the corresponding connectivity modes (i.e. eap_clear or eap_ciphered) and setup the WiFi link security accordingly. At the end of the authentication procedure, the AR allocates the IP address. For subscribers of *Gr 3*, the AR update the local database with the allocated IP address.

6.3.2.2 DTP service adjustment mechanism

The principle of the DTP service adjustment mechanism in trusted access is shown in Figure 6.4. The AR supervises the traffic related to subscribers of *Gr 3*. Upon detecting the traffic security status changed (e.g. from encrypted (HTTPS) to non-encrypted (HTTP)), a script is triggered to take the adequate decision (e.g. activate DTP service). This script updates the local database and triggers the connectivity update procedure. The AAA proxy in the AR retrieves the DTP status from the database and relays it to the AAA client in the T-AP. After that, the AAA client triggers the EAP authenticator to update the current network connectivity with the corresponding connectivity mode (i.e. eap-ciphered).

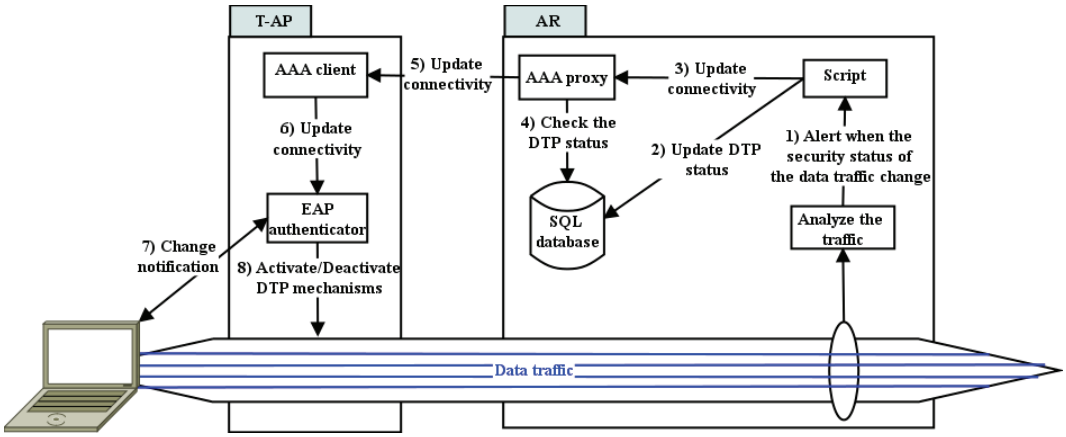


Figure 6.4: Adaptive DTP service in trusted access

6.3.3 Adaptive DTP service in untrusted access

In the untrusted WiFi access, we target to activate the DTP mechanisms on the IPsec tunnel according to the subscriber group.

During the subscriber authentication procedure, the AR should retrieve the subscriber profile from the AAA server to determine the subscriber group. If the subscriber belongs to *Gr 1*, the IPsec endpoint always applies the encryption and integrity protection mechanisms on the subscriber data traffic. If the subscriber belongs to *Gr 2*, the IPsec endpoint deactivates these mechanisms for the whole subscriber traffic.

For subscribers belonging to *Gr 3*, the encryption and integrity protection mechanisms (i.e. DTP mechanisms) are initially activated (i.e. the eap_ciphered is the default mode). When the AR detects that the traffic is already secured (i.e. using a traffic inspection mechanism), the IPsec endpoint deactivates these mechanisms (i.e. move to eap_clear mode). Similarly, when the AR detects that the traffic need protection and the DTP mechanisms are deactivated, the IPsec endpoint should activate them. Therefore, the DTP service status (i.e. activated or deactivated) in the current network connectivity should be maintained in a local database.

First, we give details about the network connectivity establishment in untrusted accesses. Then, we propose a DTP service adjustment mechanism that is required for subscribers of *Gr 3*.

6.3.3.1 Network connectivity establishment

The flow chart related to the network connectivity establishment for a new session in untrusted access is shown in Figure 6.5. The network connectivity is established as follows:

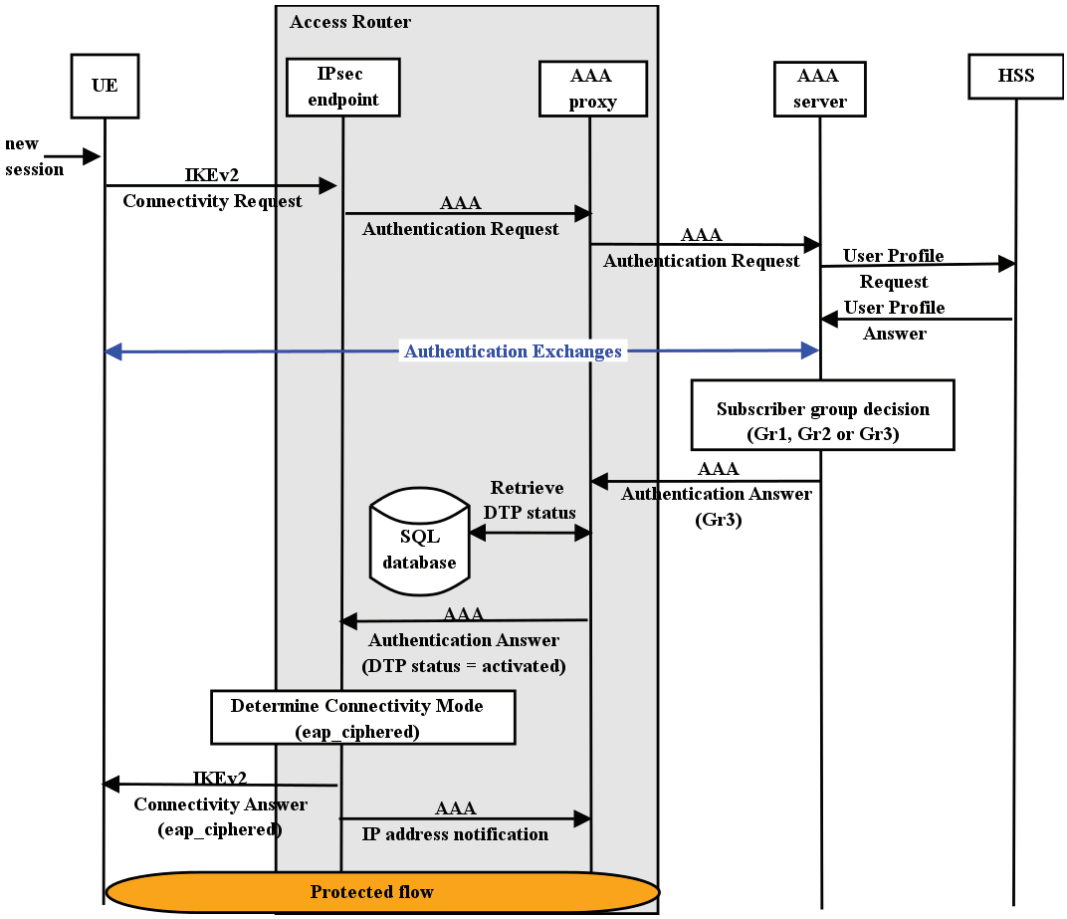


Figure 6.5: Connectivity Establishment flow chart

1. The User Equipment (UE) initiates the IPsec tunnel establishment by sending the EAP request encapsulated in the "IKE-AUTH Request" message to the AR as shown in Figure 3.5.

2. Inside the AR, the IPsec endpoint relays the authentication request to the AAA proxy. The AAA proxy encapsulates the EAP request in the AAA message and sends it to the AAA server.

3. After successful authentication, the AAA server decides the subscriber group (i.e. *Gr1*, *Gr2* or *Gr3*) to which the UE belongs. Then, it provides the AAA proxy in the AR with the decided subscriber group. In case the subscriber belongs to *Gr 3*, the proxy retrieves the current DTP service status (i.e. activated or deactivated) from the local database. If there is no entry for this subscriber, the proxy inserts a new entry in the database and associates the default status (i.e. activated) to this entry.

4. The AAA proxy relays the DTP service status to the IPsec endpoint in the AR. Based on this information, the IPsec endpoint determines the corresponding connectivity modes (i.e. eap_clear or eap_ciphered) and established the IPsec tunnel accordingly. The IP address is allocated by the IPsec endpoint during the tunnel establishment. Then, the IPsec endpoint should notify the AAA proxy about the subscriber IPv6 address. The AAA proxy updates the subscriber entry in the local database with the allocated IP address.

6.3.3.2 DTP service adjustment mechanism

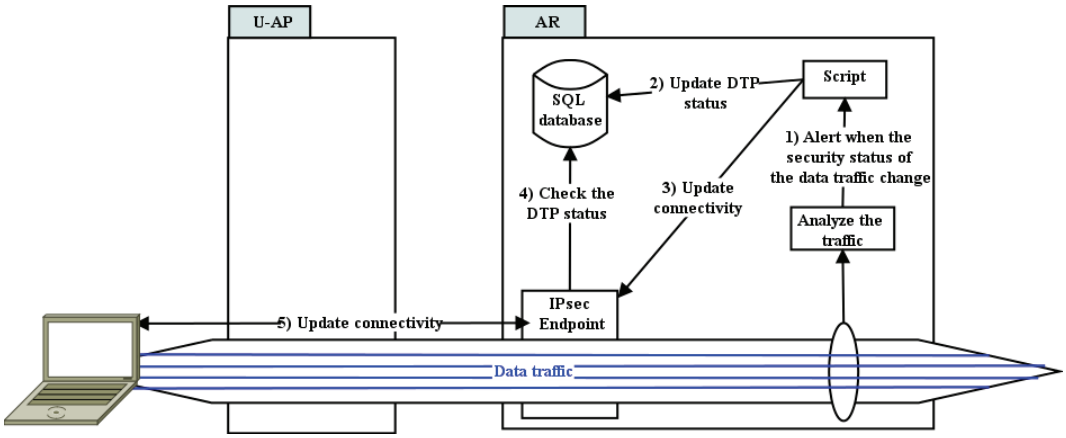


Figure 6.6: Adaptive DTP service in untrusted access

The principle of the DTP service adaptation mechanism is shown in Figure 6.6. The AR supervise the traffic related to subscribers of *Gr 3*. Upon detecting the traffic security status changed (e.g. from encrypted (HTTPS) to non-encrypted (HTTP)), a script is triggered to take the adequate decision (e.g. activate DTP service). This script triggers the IPsec endpoint to reset the IPsec tunnel with the corresponding connectivity mode (i.e. eap-ciphered).

6.4 Adaptive DTP service Validation

6.4.1 Test bed detailed description

Each functional block of the test bed is implemented in a separate Virtual Machines (VMs). All entities in our test bed use open source implementations. We used an Intel core 2 duo machine with $2.4GHz$ as processor, $4Go$ of RAM and $150Go$ of hard disc. This machine run Ubuntu linux with a 13.04 kernel. We use Oracle VM VirtualBox software [Vir] for virtualization. Regarding the IP addressing, we decided to use the IPv6 addressing in our test bed.

Although all VMs are located in the same host, we managed to setup a WiFi communication between the Client and Trusted/Untrusted access point using a Wireless USB adaptor. In the following, we give details about the different Virtual Machines composing our test bed.

6.4.1.1 Access Router

We created a Virtual Machine (VM) to act as an Access Router (AR). This VM has 1024 MB of RAM and runs the XUbuntu Linux [XUb] with a 13.04 kernel. We configured three network interfaces in the AR VM as shown in Table 6.1.

Regarding the trusted access, the AR VM should announce the IPv6 prefix. For that reason, we installed the "radvd" software [rad]. With this software, the AR VM is able to send Router Advertisement (RA) messages periodically or upon receiving Router Solicitation (RS) messages.

For Untrusted access, the AR VM should include an IPsec endpoint to be able to setup IPsec tunnels with the Client. For that end, we used the StrongSwan software [Str]. Ubuntu provides a native support of IPsec suite. However, this support does not consider the dynamic authorization extensions to Radius [CDE+08] (e.g. Disconnect Messages) that we will need when we implement the adaptive DTP service. In fact, the version 4.6.3 of StrongSwan is required. Therefore, we disabled the native support and instead installed a more recent version of StrongSwan. In addition, the AR VM includes a RADIUS proxy to relay the authentication requests between the IPsec endpoint and the AAA server.

Interface Name	Definition	IP address
Eth1	connected to Trusted network	@IPv6: 2a01:cf00:79:2::1/64
Eth2	connected to Untrusted network	@IPv4: 192.168.56:101 (used by the IPsec tunnel) @IPv6: 2a01:cf00:79:1::1/64 (used inside the IPsec tunnel)
Eth3	connected to Internet	@IPv6: 2a01:cf00:79::2/64

Table 6.1: Network Interfaces in the AR VM.

The default route in the AR is defined as follows: *Default via 2a01:cf00:79::1 dev eth3*. However, as we can see from Table 6.1, the AR connects three subnets (2a01:cf00:79::, 2a01:cf00:79:1::, and 2a01:cf00:79:2::). To enable the routing between these subnets, we need a Neighbor Discovery (ND) proxy [TTP06]. So, we installed the "ndppd" daemon [ndp] which is launched automatically at VM starting.

6.4.1.2 AAA server

We created a Virtual Machine (VM) to act as a AAA server. This VM has 128 MB of RAM and runs the Ubuntu Linux [Ubu] with a 13.04 kernel. The AAA server is based on the *FreeRadius* software [Fre], which is an open source implementation of the RADIUS authentication server.

The AAA server includes an SQL database where subscriber profiles are stored. This database acts as the HSS database in the test bed.

6.4.1.3 Access Point

We created a VM to act as an Access Point (AP). This VM has 1024 MB of RAM and runs the Ubuntu Linux [Ubu] with a 13.04 kernel. The AP implementation is based on the *hostapd* software [hos]. We used this software and a wireless dual band USB adaptor [D-l] to emulate a WiFi AP. We configured this AP to act as trusted (i.e. AP based on the 802.1X protocol) and untrusted (i.e. open AP) at the same time. Therefore, this AP broadcast two SSIDs: TTLS_GS where the access is controlled with 802.1X protocol, and OPEN_GS where the access is open for all users.

6.4.1.4 Client

We created a VM to act as the subscriber device (Client). This VM has 1024 MB of RAM and runs the KUbuntu Linux [KUb] with a 13.04 kernel. Similarly to the AP VM, we associate a wireless dual band USB adaptor [D-l] to the Client VM to emulate a wireless device. Therefore, the Client VM use this wireless interface to connect to Trusted as well as Untrusted accesses.

To enable the Client to use the trusted access (i.e. based on 802.1X), we used the wpa_supplicant software [Mal] that is an open source implementation of the IEEE 802.1i supplicant working on Linux. In the wpa_supplicant configuration file, we define the SSID (i.e. TTLS_GS) to which the Client should connect and the authentication method to be used (i.e. EAP-TTLS method). In this method, the Client authenticates the AAA server using the certificate and establishes with this server a TTLS tunnel. Inside this tunnel, the client use the Login/password method (i.e. MD5) to be authenticated.

Regarding the untrusted access, the Client should simply connect to open WiFi AP (i.e. OPEN_GS). In this case, the StrongSwan software should be installed in the Client VM to be able to setup IPsec tunnel with the AR. Similarly, the Client use the EAP-TTLS as the authentication method.

6.4.2 Adaptive DTP implementation in untrusted access

We setup a mechanism to enable adaptive DTP service in the untrusted access. To simplify the implementation of our mechanism, we decided to put the Security decision-maker module (i.e. for DTP mechanisms activation/deactivation) as well as the SSD database in the AR VM instead of AAA server VM as a first step. Obviously, these modules can be transferred to the AAA server VM later and this will not impact our mechanism.

The local database in the AR is based on the MySQL software (version 5.5.32) [MyS]. The table structure in this database is shown in Table 6.2. The *UserName* attribute represents the user identifier. The *DTP-status* attribute is a boolean and refers to DTP service status (i.e. 0 = DTP is deactivated, 1 = DTP is activated). The IP address attribute includes the subscriber IPv6 address allocated by the IPsec endpoint after authentication and authorization.

UserName	DTP-status	IP address
Client1	0	2a01:cf00:79:2::3/128
Client2	1	2a01:cf00:79:2::4/128

Table 6.2: Table in MySQL database.

We installed the SNORT software [Sno] in the AR VM to inspect the subscriber traffic and detect the session security status (i.e. whether the session is already protected such as HTTPS). SNORT is an open source software and is able to inspect traffic without interrupting it. That means it has no impact on the subscriber QoE.

Upon detecting that the traffic security status changed (e.g. from encrypted (HTTPS) to non-encrypted (HTTP)), SNORT sends an alert to Syslog-ng software[Sysa]. This software reproduces the Syslog standard [Ger09] for computer message logging. We use the Syslog-ng software to transform the SNORT alerts into actions. For instance, we update the syslog-ng configuration file by adding a source for SNORT (i.e. to listen any alerts coming from SNORT), a filter to separate alerts related to clear traffic from alerts related to encrypted traffic, and a destination that performs the adequate action (i.e. the scripts that should be run).

We defined "activate_DTP.sh" and "deactivate_DTP.sh" scripts. The Syslog-ng software executes the "activate_DTP.sh" script when the traffic pass from encrypted to clear. This script resets the IPsec tunnel while updating the DTP_status in the SQL database from *deactivated* to *activated*. Similarly, the Syslog software executes the "deactivate_DTP.sh" script when the traffic pass from clear to encrypted. This scripts resets the IPsec tunnel while updating the DTP_status in the SQL database from *activated* to *deactivated*. In both cases, during the IPsec tunnel reestablishment,

the IPsec endpoint checks the SQL database to know whether the DTP service should be ensured.

The diagram in Figure 6.7 illustrates how our mechanism works in case of HTTP or HTTPS traffic.

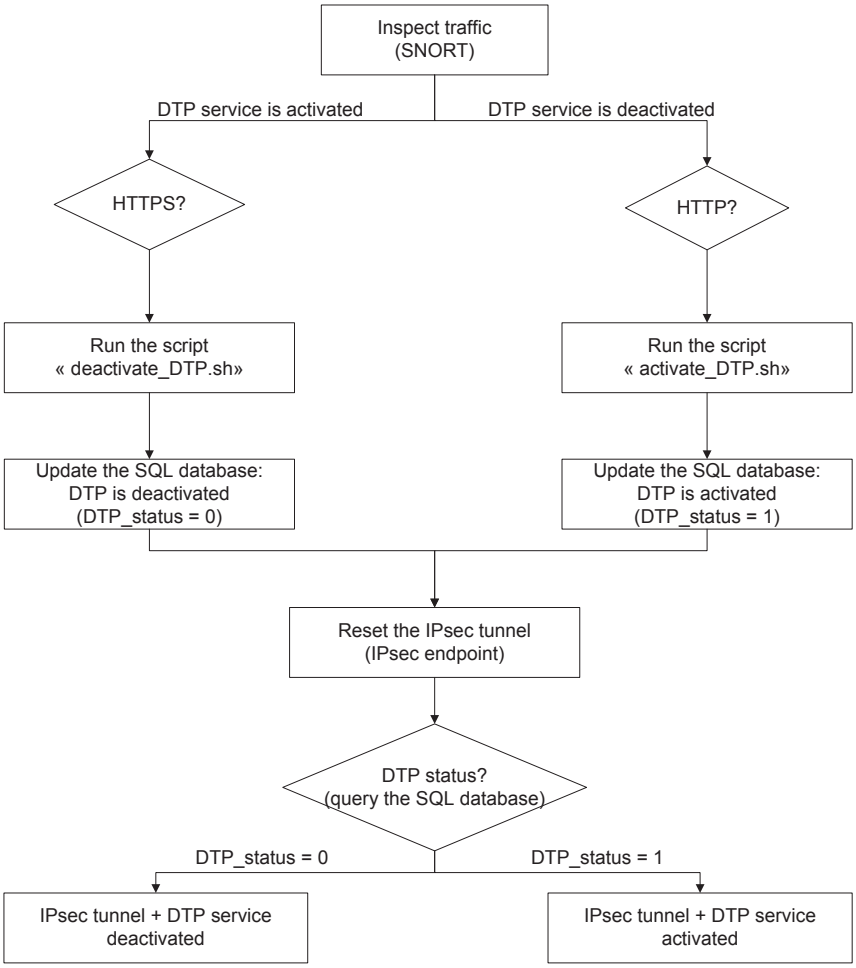


Figure 6.7: Adapting the DTP service for HTTP/HTTPS traffic

6.4.3 Evaluations

We performed several experiments to assess the impact of the DTP service activation/deactivation on the subscriber bitrate and packet loss. For this end, we used the D-ITG software [Dis]. This software is capable to produce traffic at packet level by accurately replicating appropriate stochastic processes for both IDT (Inter Departure Time) and PS (Packet Size) random variables (exponential, uniform, normal, pareto, etc.). The D-ITG comprises different components as shown in Figure

6.8. The ITGSend and ITGRecv components exchange traffic and produce log files containing detailed information about every sent and received packet. These log files are sent to the ITGLog component. The ITGDec component is in charge of analyzing the log files in order to extract performance metrics (e.g. Bitrate, packet loss, latency, etc.) related to the traffic flows.

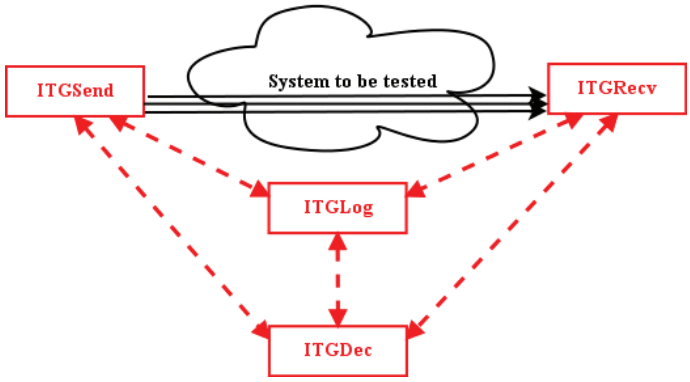


Figure 6.8: D-ITG architecture

To perform experiments on our mechanism performances, we created a VM to act as a Web Server where the ITGRecv component is running. This VM has 128 MB of RAM and runs the Ubuntu Linux [Ubu] with a 13.04 kernel. The ITGSend component is installed in the Client VM. The D-ITG integration in our test bed is depicted in Figure 6.9.

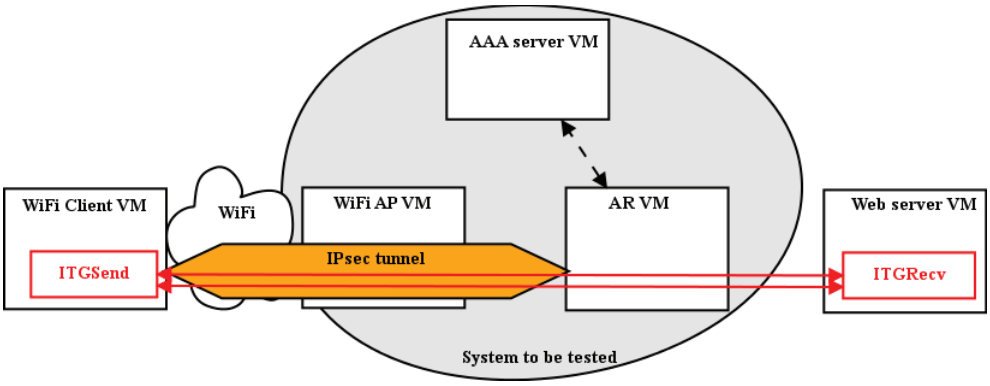


Figure 6.9: D-ITG integration in the test bed

We configure the ITGSend component to send UDP traffic (-T) to the Web Server VM on port 9501 (-rp) during $60s$ (-t). We set the packet rate mean (-O) to $40kpkt/s$. The packet inter-arrival follows the Poisson distribution. The packet size follows also the Poisson distribution with the mean 1kBytes (-o). The ITGSend produces the "sender.log" as a log file. In the Web Server VM, we just launch the

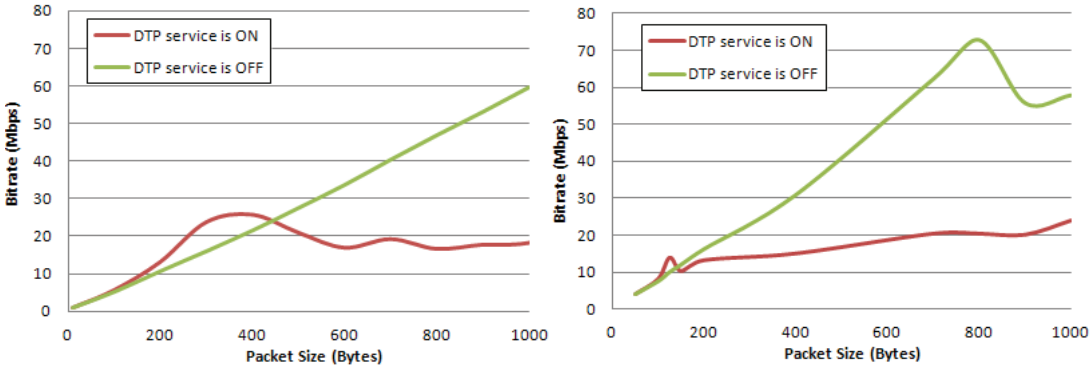
```
ITGSend -T UDP -a 2a01:cf00:79:2::80 -rp 9501 -O 40000 -o 1000 -t 60000 -l sender.log
```

ITRecv. After running the test, we analyze the log files via the ITGDec component.

```
ITGRecv -l receiver.log
ITGDec receiver.log
```

Our methodology consists in varying the packet size (-o), running the test bed, and extracting the bitrate and packet loss values each time. First, we run the test while the DTP service is activated, then, while the DTP service is deactivated. The bitrate and packet loss values are shown in Figure 6.10 and Figure 6.11, respectively.

As it was expected, a connectivity in which the DTP service is deactivated outperforms the connectivity in which the DTP service is activated. For instance, Figure 6.10b shows that, by deactivating the DTP service, the highest bitrate that the test bed could achieve is $70Mbps$ approximately. However, the highest bitrate that the test bed could realize is around $20Mbps$ when the DTP is activated. Similarly, for a connectivity with a deactivated DTP service, the packet loss starts to increase when the packet size exceeds $600Bytes$. For a connectivity with an activated DTP service, the packet loss starts to increase when the packet size exceeds $250Bytes$.

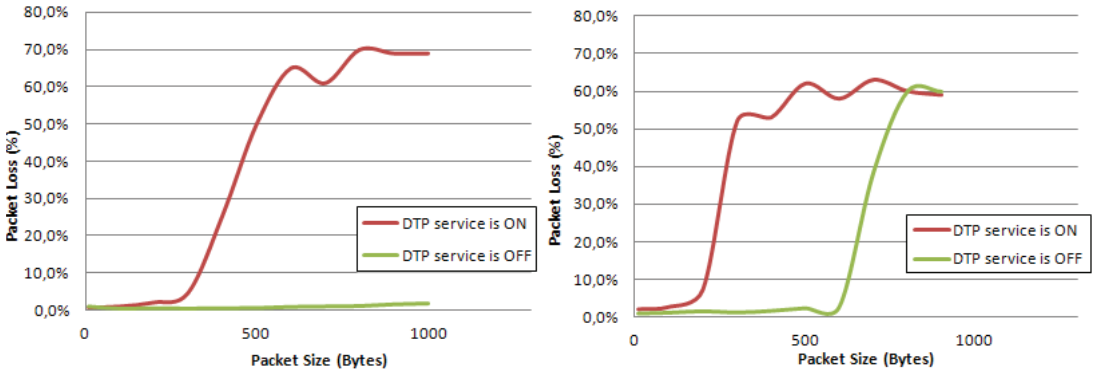


(a) 20 packets per second

(b) 40 packets per second

Figure 6.10: Bitrate in Untrusted access

As we said before, generally, the use of SNORT has no impact on the subscriber bitrate as it just takes a copy of the subscriber traffic and inspects this copy. To check this assertion, we compare the bitrate values in two cases: (i) SNORT is running, and (ii) SNORT is stopped. We configured the ITGSend component in the Client VM to send a TCP traffic to the port 80 in the Web Server VM where we



(a) 20 packets per second (b) 40 packets per second

Figure 6.11: Packet Loss in Untrusted access

launched the ITGRecv component. These packets pass through the AR VM where we have SNORT. Therefore, we can compare the bitrate in the above cases.

We set the packet size (-o) to 100 Bytes. We varies the packet rate (-O) from $0\,pkt/s$ to $300\,pkt/s$ and measure the bitrate each time. The bitrate values are shown in Figure 6.12. As we showed before, the DTP deactivation increases the subscriber bitrate. Moreover, it is clear that SNORT has no or little impact on the subscriber bitrate.

```
ITGSend -T TCP -a 2a01:cf00:79:2::80 -rp 80 -O 300 -o 100 -t 30000 -l sender.log
```

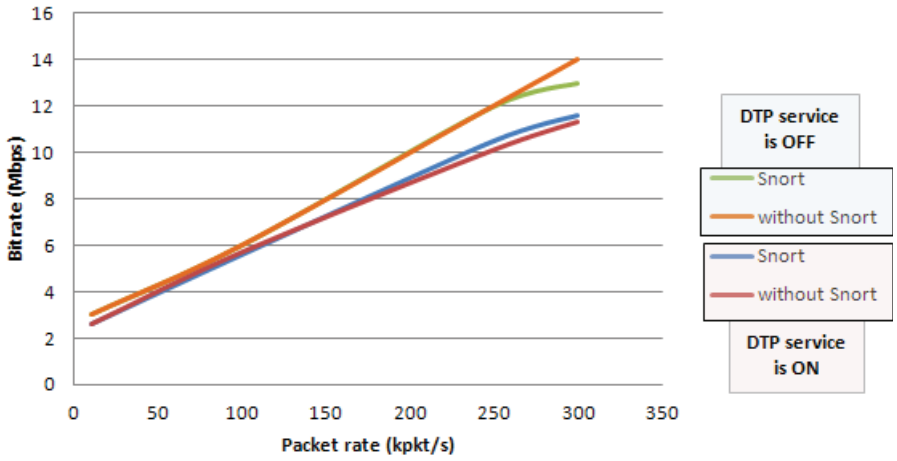


Figure 6.12: SNORT impact on subscriber Bitrate

6.5 Implementation Challenges

We met several technical challenges during implementing adaptive DTP service. Measuring the CPU usage in the AR related to each security operation was our first challenge. By having such values, we can determine how far the CPU usage can be optimized when the DTP mechanisms are used in adaptive manner. We used the *htop* command in Linux to get approximated percentages of the CPU usage by the different process. However, we were not able to get the exact CPU usage percentage related to the encryption and integrity protection mechanisms. In fact, we can only get the CPU usage percentage related the whole IPsec process.

In the current version, to activate/deactivate the DTP Service, the IPsec endpoint interrupts the tunnel IPsec, triggering, then, the Client to re-establish the IPsec tunnel. Such behavior is not desired as it interrupts the subscriber session and impacts, therefore, the subscriber QoE. Obviously, the proposed mechanism needs improvements. There are two alternatives to overcome such behavior:

- Establish and maintain two different IPsec tunnels at the same time between the Client and the AR. The first tunnel activates all the security mechanisms, including the encryption and integrity protection mechanisms (i.e. DTP service). The second tunnel does not ensure the DTP service. Having these two tunnels, the Client will just forward the traffic to the adequate tunnel based on the nature of traffic.

- Notify the Client about the DTP activation/deactivation via the IKEv2 protocol without interrupting the IPsec tunnel. In this case, we need to update the IKEv2 protocol to support additional signaling exchanges for that purpose.

The excessive use of network resources is the main drawback of the first alternative. In fact, two tunnels are maintained simultaneously for each subscriber. Therefore, the second alternative seems to be more reasonable. However, the IKEv2 implementation in StrongSwan need to be updated to support new signaling exchanges.

Regarding the Trusted access, implementing adaptive DTP service was constrained by the hostapd software. Actually, hostapd uses the driver "nl80211" to operates and controls the WiFi AP. This driver is not allowing to have unprotected wireless link. To activate/deactivate these mechanisms dynamically, we need to update the driver with new policies.

6.6 Conclusion

In this chapter, we specified mechanisms that enable adaptive Data Traffic Protection (DTP) service in non-3GPP accesses. After that, we implemented a test bed

that reproduces the trusted and untrusted non-3GPP accesses. In this test bed, we tried to implement the proposed mechanisms. Our ultimate purpose was to check the portability of the proposed mechanisms in real architectures.

The work done under this stage has many results. First, the test bed enabled us to validate the feasibility of our proposal - adapting the network connectivity to the contextual information. The evaluation of the proposed mechanism (i.e. activating/deactivating the DTP service according to the traffic nature) showed that this mechanism improves network performances. In fact, the obtained results showed that the bitrate increases and the packet loss decreases when the DTP service is deactivated.

As we said in the implementation challenge section, the proposed mechanism need improvement. Therefore, the first perspective of this work consists in updating the IKEv2 protocol to support additional signaling exchange. The second perspective consists in implementing new mechanisms to ensure adaptive authentication, access control and privacy services in the untrusted access.

Chapter 7

Implementing Adaptive Mobility Services in LTE/EPC Access

7.1 Introduction

In this chapter, we propose to integrate the Software Defined Networking (SDN) principles into the LTE/EPC access as a way to ensure adaptive mobility services. SDN is a recent trend in communications networking, whereby the behavior of network equipments can be specified and controlled from a single, high-level software program. This trend is reshaping the way networks are designed, managed, and secured. In fact, SDN replaces manual interface of network equipment with a programmatic interface, which enables the automation of tasks such as configuration and policy management [ONF12a]. OpenFlow (OF) is one of the main protocols that apply the SDN concepts as it enables the remote software-based control and management of network equipments with open interfaces in the data plane. Indeed, opening up interfaces to program the network equipments makes the network control, upgrade and management easier. Moreover, networks can be easily extended with new functionalities.

We first recall our motivation and the SDN and OpenFlow principles. After that, we propose a new control plane for LTE/EPC architecture, which is based on the OF protocol. We then show that the proposed control plane enables adaptive mobility services in the LTE/EPC architecture. Although our proposal can appear appealing from a theoretical standpoint, we do not hide that such shift is not a trivial task. We discuss some of the most prominent implementation issues that prevent fast deployment of SDN-based EPC in today's network. After that, we provide a preliminary signaling cost evaluation of the proposed architecture. We conclude this

chapter by presenting some related works.

7.2 Motivation

The aim of this chapter is to propose an example of implementation in LTE/EPC access of the proposed Mobility Manager (i.e. the red components in Figure 7.1). This implementation should ensure adaptive mobility services and the flexible use of network resources. Particularly, we focus on the Session Continuity service.

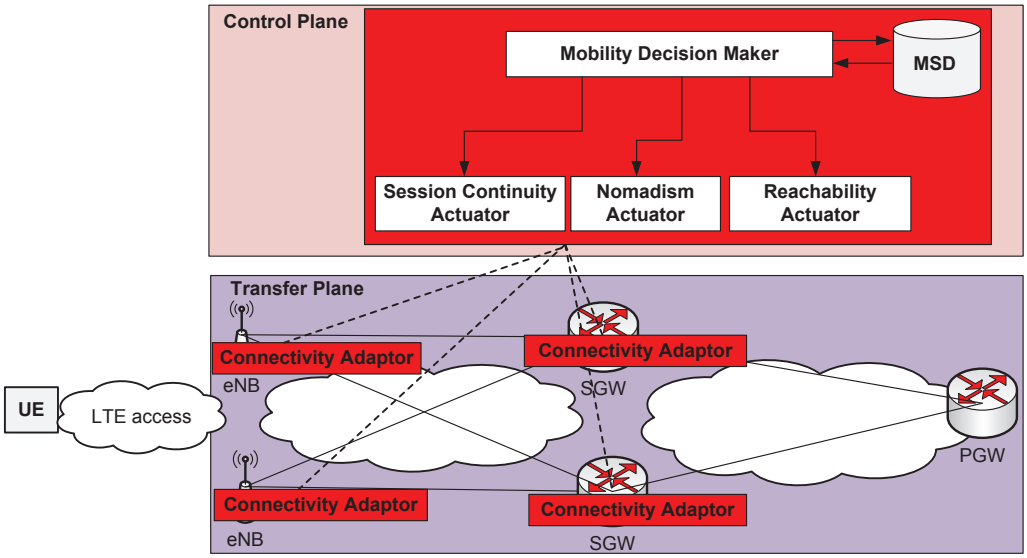


Figure 7.1: Mobility Manager in LTE/EPC access.

To ensure adaptive Session Continuity service, the Mobility Manager should be able to select the adequate mobility protocol, activate/deactivate mobility anchors in the data plane, and configure these anchors in accordance with the data flow needs. Moreover, to preserve the session continuity it should include a mechanism that *moves active sessions transparently and temporarily from one network equipment to another without causing user session interruption.* This is especially critical in situations such as network equipment failure, overload situations and during energy saving measures. To propose such mechanism in LTE/EPC access, we need to analyze the GPRS Tunneling Protocol (GTP) as it is the main mobility protocol used within this access.

The GTP tunnel is identified in each node with a Tunnel Endpoint Identifier (TEID), an IP address and a UDP port number [3GP13b]. The TEID unambiguously identifies a tunnel endpoint in the receiving GTP-U protocol entity. The receiving end side of the GTP tunnel locally assigns the TEID value that the trans-

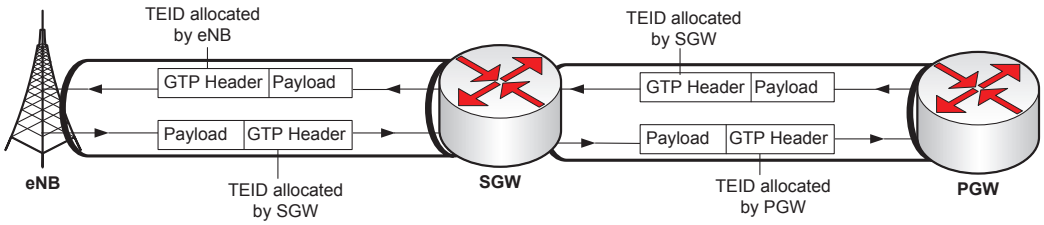


Figure 7.2: GTP tunnels in EPC.

mitting side has to use. For instance, the PGW should use the TEID value allocated by the corresponding SGW for the downlink traffic over S5 interface (see Fig.7.2). In the same way, the SGW should use the TEID value allocated by the corresponding PGW for the uplink traffic over S5 interface. Then, the TEID values are exchanged between tunnel endpoints using GTP-C and S1-AP.

To summarize, the TEID values are locally allocated by each node. Therefore, new TEID values should be exchanged for each node relocation. This challenges the network elasticity and, consequently, the flexible use of network resources. Scenario D in Chapter 3 underlined the same issue.

To address the above issue, we propose to reshape the LTE/EPC access by introducing the OpenFlow protocol.

7.3 SDN and OpenFlow

The SDN architecture is depicted in Figure 7.3. As we can see, the control and data planes are decoupled in this architecture. The network intelligence (i.e. the control plane) is logically centralized in a set of SDN controllers and the underlying network infrastructure (i.e. data plane) is simplified and abstracted from applications [ONF12a]. In fact, the network equipments in the data plane need just to accept instructions from the SDN controllers instead of implementing and processing thousands of protocol standards.

The SDN replaces the manual interfaces of the networking equipment at the data plane with a programmatic interface that enables the automation of tasks such as configuration and policy management. The SDN paradigm has two main benefits: (i) frees network operators from manually configuring and instructing the network equipments, and (ii) enables the network to dynamically respond to application requirements [Net12].

The OpenFlow (OF) protocol is one of the main components of the SDN concept. It enables a remote software-based controller to manage the connected OF switches through a well-defined "forwarding instruction set" as shown in Figure 7.4. This

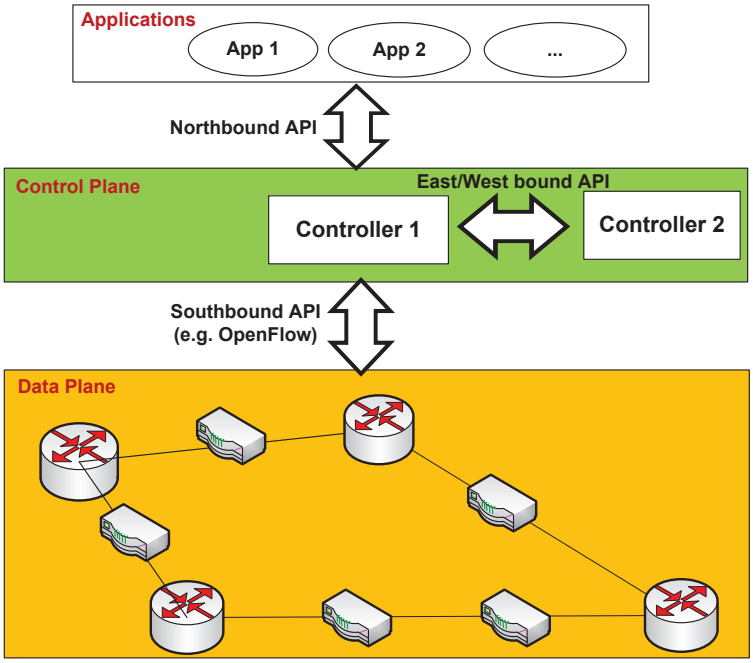


Figure 7.3: SDN architecture

figure is taken from [ONF12b].

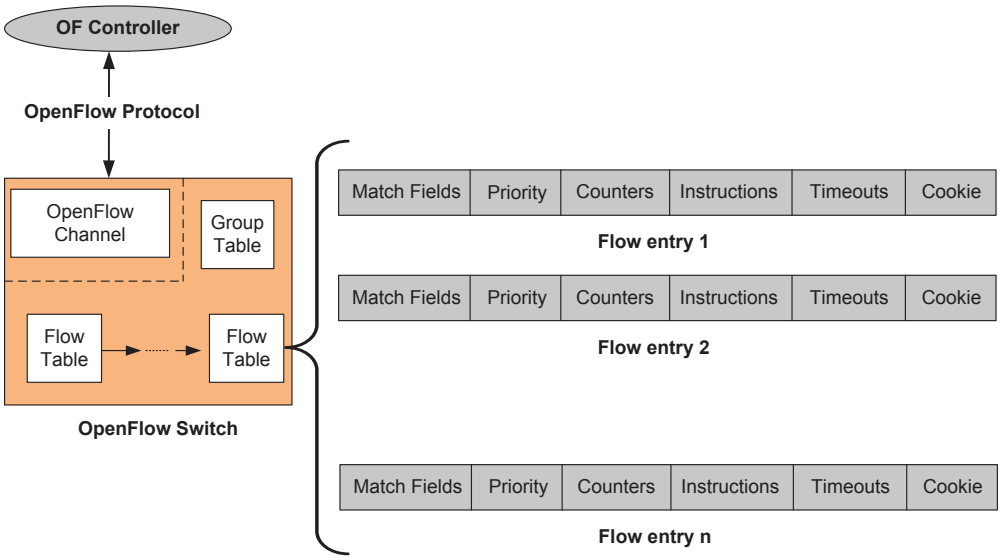


Figure 7.4: OpenFlow switch

- **OF Controller**: It consists of a network operating system, running a collection of modules and translating external attributes into the network (e.g. translates any operators policy into rules that should be enforced into OF switches). The common function for controller applications is to respond to a packet arrival by installing rules in OF switches for handling subsequent packets related to the same flow. Using the OF protocol, the controller can add, update, and delete flow entries in flow tables [MAB+08]. It can also read traffic statistics collected by the OF switch.

- **OF Switch**: It includes one or more flow tables. Each flow table contains a set of flow entries which consist of match fields, counters, and a set of instructions to apply to matching packets. Upon receiving data packet, the OF switch looks up at the flow tables for the flow entry related to this packet. If a matching entry is found, the instructions associated with this packet are executed (e.g. forwarding, dropping, or modifying the packets). If no matching entry is found, the packet is forwarded to the controller. This later creates and forward the associated flow entry to the concerned OF switches. Once set up, the flow forwarding tables remain cached on the OF switches so that this process is not repeated for subsequent packets in the same flow. The OF controller configures how long the flow table is cached: either indefinitely, after a fixed timeout, or after a period of inactivity.

7.4 OF-based LTE/EPC architecture

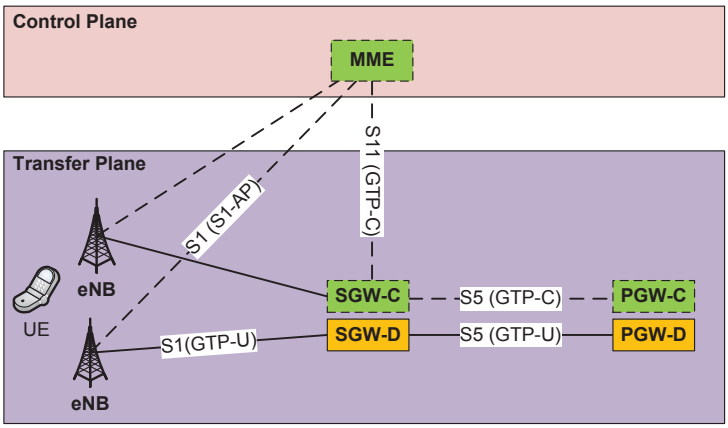
7.4.1 Architecture Description

To enable adaptive mobility services, we propose a new control plane for the LTE/EPC access (i.e. OF-based architecture). Figure 7.5 presents the LTE/EPC architecture according to the 3GPP and OF-based models. In the OF-based LTE/EPC architecture, we replace the control protocols that run on the S1-MME (between MME and eNB) and the S11 (between MME and SGW) interfaces by the OF protocol as shown in Figure 7.5b.

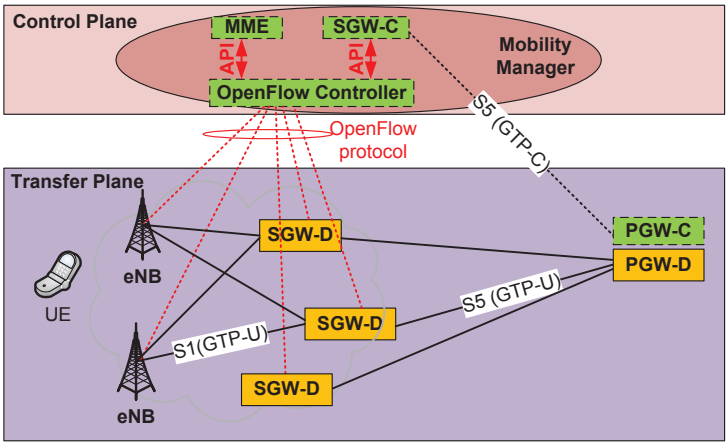
In the line with the SDN principle, we propose to separate out all control functions from the data forwarding function in SGWs of the same pool area. As a result, the whole intelligence in the SGW (SGW-C software) is centralized and runs on top of the OF Controller (OF-ctr) as an application. The data forwarding function is performed by the SGW data plane (SGW-D). Also, the MME software runs on top of the OF-ctr as an application.

The OF-based LTE/EPC architecture is composed of the following entities:

- **OpenFlow Controller (OF-ctr)**: is the main component of our architec-



(a) 3GPP model



(b) OF-based model

Figure 7.5: LTE/EPC architecture.

ture as it manages the forwarding plane of eNB and SGW-D. The OF-ctr is responsible for user session establishment and load monitoring at the data plane.

- ***MME***: is responsible for UE authentication and authorization, and intra-3GPP mobility management. In our architecture, the MME is no more responsible for the SGW and PGW selection. The MME communicates with the OF-ctr using Application Programming Interface (API). The 3GPP interface between the MME and HSS is still maintained.

- **SGW control plane (SGW-C)**: represents the SGW's intelligence part. It is responsible for GTP tunnel establishment including TEIDs allocation. The SGW-C allocates *unique TEID value per session for the uplink traffic within*

the S1-U interface. It allocates also *unique TEID value for the downlink traffic within S5-U interface*. With the OF protocol, the OF-ctr can set counters in the SGW-Ds in order to get periodic load statistics. By comparing the received load statistics with the SGW-D capability, the OF-ctr can easily get the load status of each SGW-D and therefore perform more efficient load balancing (i.e. based on the current load of SGW-Ds).

- **SGW data plane (SGW-D)**: represents an advanced OF switch that is able to encapsulate/decapsulate GTP packets. This switch applies the rules received from the OF-ctr. It is responsible for packet forwarding between the eNB and PGW.
- **eNB**: keeps the same radio functions specified by 3GPP standards (e.g. scheduling, radio resource management, etc.). Only the communication between the eNB and the core network is changed. In fact, the eNB is enabled with the OF protocol for the data forwarding through the S1 interface. Therefore, the data forwarding is based on instructions received from the OF-ctr.
- **PGW**: still has the same function as in the 3GPP standards.

The TEID values allocation in the SGW-C is performed once per session. These values remain invariant during moving the session from one SGW-D to another. Actually, when the SGW-C commands the OF-ctr to relocate a new SGW-D for a specific session, the OF-ctr will just update in the eNB the flow entry related to this session with the IP address of the new SGW-D. Also, the SGW-C updates the SGW-D IP address in the PGW.

As we can see, the OF-ctr with the SGW-C and MME applications represents our Mobility Manager. This controller is responsible for adapting the mobility services to the contextual information as it will be shown in the following sections. The GTP tunnels and bearers (i.e. mobility tunnels) in the transfer plane are programmed according to instructions received from the OF-ctr. For instance, the OF-ctr may decide the OF entry *Release Timer* in eNB. This parameter represents the OF entry lifetime (i.e. the period during which the S1 bearers is maintained). Then, the OF-ctr transfers this parameter to eNB via the OF protocol. The eNB maintains the radio and S1 bearers as long as the session is active. Upon detecting UE inactivity, the radio bearer is released and the S1 bearer is maintained (i.e. the OF entry is maintained) as long as the Release Timer is not expired. The Release timer is configured according to the traffic pattern (e.g. session type, session duration, periodic connection request, etc.).

Moreover, to not modify the UE software, the 3GPP signaling protocol used between UE and eNB (i.e. RRC protocol) is still considered in the OF-based architecture. In addition, the 3GPP signaling protocol used between UE and MME

(i.e. NAS protocol) is still considered and transported transparently between eNB and MME via the OF protocol. These protocols handle bearer establishment and release, tracking area update, paging, etc.

7.4.2 Session Management Procedures

7.4.2.1 Initial attachment procedure

The UE uses this procedure to register to the network and get an IP address. This IP address is required for the further data plane establishment. Compared to the classic LTE/EPC initial attachment procedure, the data plane is not established systematically. The initial attachment procedure is depicted in Figure 7.6.

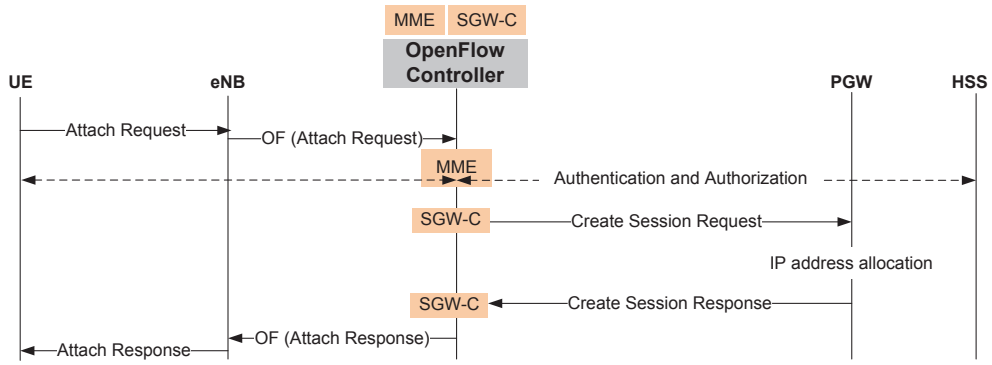


Figure 7.6: Initial attachment procedure.

The UE sends the Attach Request message to the eNB. The eNB piggybacks the Attach Request message into OFPT_PACKET_IN message[1] and sends it to OF-ctr (see Fig.7.6). The OF-ctr forwards the Attach Request message to the MME. This triggers the authentication and authorization exchanges in the software MME. The authentication exchanges between the MME and the UE go through the OF-ctr.

After successful authentication, the OF-ctr triggers the SGW-C to select the default SGW-D and PGW. The SGW-C selects a default SGW-D for this UE and generates an SGW-TEID value for the downlink traffic in S5 interface. Then, it stores these parameters in its table. Moreover, the SGW-C sends these parameters to the selected PGW via the classic GTP-C message "Create Bearer Request". The PGW stores the received parameters in its table. Therefore, whenever the PGW receives a packet destined to this UE, it knows where to send it and the GTP header to add.

[1]This message is a standard OpenFlow protocol message [ONF12b]. When a packet arrives and no flow entry matches to this packet, the OF switch sends the packet header to the OF controller via this message

The PGW allocates an IP address for this UE and generates a PGW-TEID value for the uplink traffic in the S5 interface. These parameters are sent to the SGW-C via the classic GTP-C message "Create Session Response". Then, the SGW-C updates its table with the received parameters. The SGW-C notifies the MME about the UE IP address via the OF-ctr. Consequently, the MME include this IP address in the Attach Response message and sends it to the UE. After successfully completing the initial attachment procedure, the UE is authorized to use the LTE/EPC access and has an IP address.

7.4.2.2 Data Plane establishment procedure

This procedure is required for each newly launched session and is depicted in Figure 7.7. First, the UE sends to the MME the NAS Service Request message to get the authorization to establish the radio data bearer. This is required to avoid the unauthorized use of radio resources.

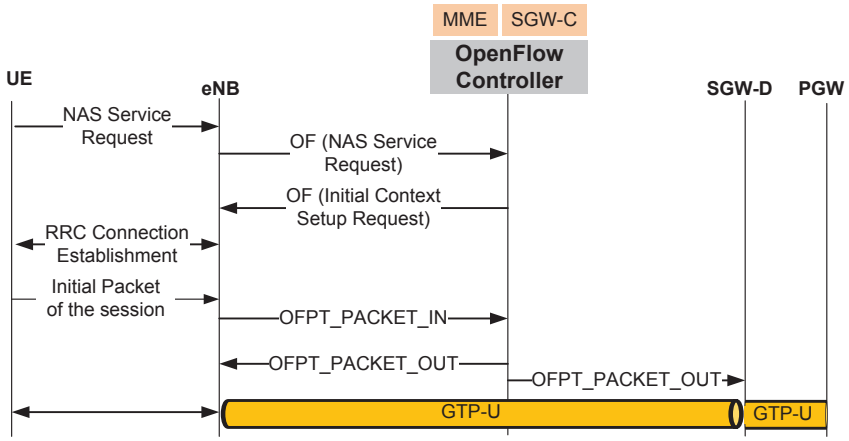


Figure 7.7: Data plane establishment.

The UE sends to the eNB the initial packet via the established radio data bearer. Then, the eNB checks its flow tables. As no flow entry exists for this initial packet, the eNB sends to the OF-ctr the packet header via the OFPT_PACKET_IN message. Also, the eNB includes in this message the eNB-TEID value for the downlink traffic in S1 interface. The OF-ctr analyzes the packet header to identify the source IP address, the destination IP address and the session type. The OF-ctr presents these information to the SGW-C.

Based on the IP addresses and the load statistics collected by the OF-ctr, the SGW-C selects the adequate SGW-D. In addition, The session type enables the SGW-C to decide the appropriate QoS. For instance, if the packet belongs to VoIP session and the already selected SGW-D is overloaded, the OF-ctr decides to allocate

another SGW-D with less load. The SGW-C sends back to the OF-ctr the SGW-D IP address, the SGW-TEID values and the QoS level.

The OF-ctr creates a flow entry for the subsequent packets related to the same session and sends it to the eNB via the OFPT_PACKET_OUT message[2]. The action field of this flow entry includes the SGW-D IP address and the SGW-TEID for the uplink traffic in the S1 interface. Similarly, the OF-ctr creates and sends to the SGW-D a flow entry related to this session via the OFPT_PACKET_OUT message. The action field of the flow entry includes the eNB IP address, the eNB-TEID, the SGW-TEID for the uplink traffic in the S1 interface, the PGW IP address, the PGW-TEID, and the SGW-TEID for the downlink traffic in S5 interface.

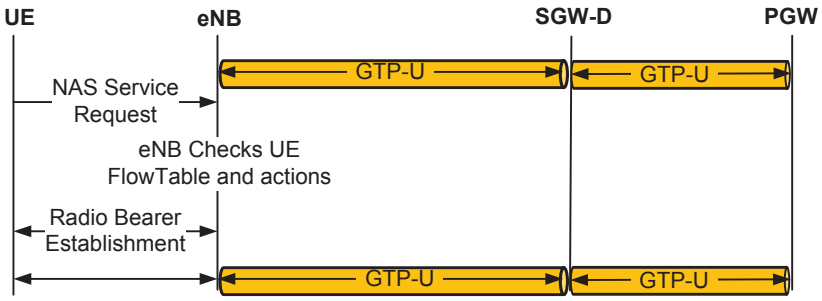


Figure 7.8: Access Bearer Setup procedure when the S1 and S5 are maintained.

One of the OF properties consists in associating a *Release Timer* for each flow entry. Therefore, the flow entry in the network equipment enabled with OF will be deleted at the timer expiration without generating any signaling load. In our proposal, we consider the same property. After performing the OF Initial Access Bearer Setup for one session, the OF-enabled eNB and SGW-D maintain the flow entry for this session as long as the related *Release timers* are not expired. Thus, when the UE reconnect to the LTE access for the same type of session before the *Release timer* expiration, it need just to re-establish the radio data bearer as shown in Figure 7.8.

This *Release Timer* corresponds to the *UE inactivity timer* in 3GPP LTE/EPC architecture. Unlike in 3GPP standards, the *Release Timer* is decided by the OF-ctr according to the session profile. If the related-application has a periodic pattern (i.e. the application connects to the network at each period T), the *Release Timer* will be equal to this period (i.e. T). If the related-application connects very rarely to the network, the network equipments will be configured to release the network resource as soon as the UE inactivity is detected.

[2]This message is a standard OpenFlow protocol message [ONF12b]. The OF controller uses this message to send to the OF switch flow entries.

By deciding Release Timer value adapted to application profile, the OF-based LTE architecture addresses the issue highlighted by Scenario D in Chapter 3. A comparison between the 3GPP and OF-based LTE architecture in terms of signaling load is presented in Section 7.7. This comparison showed that replacing the current control protocols in eNB-MME and MME-SGW interfaces by OpenFlow does not increase the signaling load in the LTE architecture. On the contrary, the signaling load is reduced in the OF-based LTE architecture because the UE S1 and S5 data bearer parameters are supposed to be kept in the network equipment during the application IDLE period. Therefore, the controller should set for each flow entry the optimal Release Timer that incurs low signaling load and less memory space usage in the network equipment, in order to avoid extra memory space usage.

7.5 Adaptive Mobility Services

In this section, we show that the proposed architecture do not only activate the Session Continity service when needed, but also adapt it to any contextual changes. Then, we show that this architecture is also able to activate/deactivate the Reachability service depending on the subscriber profile.

7.5.1 Session Continuity service

In our proposal, the data plane is only established after receiving the first data packet. In fact, the eNB relays the header of this packet to the OF-ctr. This latter analyzes the header to determine the flow type. Moreover, the OF-ctr can requests specific contextual information from the ContextMS (e.g. mobility pattern, subscriber profile, traffic pattern, device type, etc) via the East/West interface [3]. Then, it relays the collected information to the SGW-C application which runs a multi-criteria decision-maker algorithm to decide whether the Session Continuity service should be ensured for this flow.

In case the Session Continuity service should be ensured (e.g. for a highly-mobile subscriber with a VoIP session), the SGW-C application selects the adequate SGW-D and PGW and configures them to act as mobility anchors (i.e. they will establish mobility tunnels such as GTP or PMIP tunnels to transfer the flow through these tunnels). In addition, the SGW-C application selects a handover mechanism that is in accordance with the flow needs. For example, the VoIP session needs a seamless proactive handover where the perceived communication quality is not impacted (e.g. Fast Handover [Koo08]).

[3] In general, the OF controller uses the East-West interface to exchange messages with other controllers [SNK12]. We can use the same interface to communicate with a ContextMS.

In case the Session Continuity is not needed (e.g. a fairly static subscriber, short session, etc.), the SGW-C configures the SGW-D to offload the traffic (e.g. activate the Selected IP Traffic Offload (SIPTO) mechanism [3GP10d]).

In the following, we revisit Scenario D to demonstrate that the proposed architecture is also able to adapt the Session Continuity service to critical situation such as the network equipment failure and overload situation. First, we describe how the sessions restoration takes place in case of network equipment failure. Then, we demonstrate how the load balancing mechanism can be easily implemented in the proposed architecture to face the overload situations.

7.5.1.1 Resiliency in OF-based LTE/EPC architecture

In the OF-based architecture, the SGW failure can be easily handled. The restoration procedure upon SGW failure is depicted in Figure 7.9).

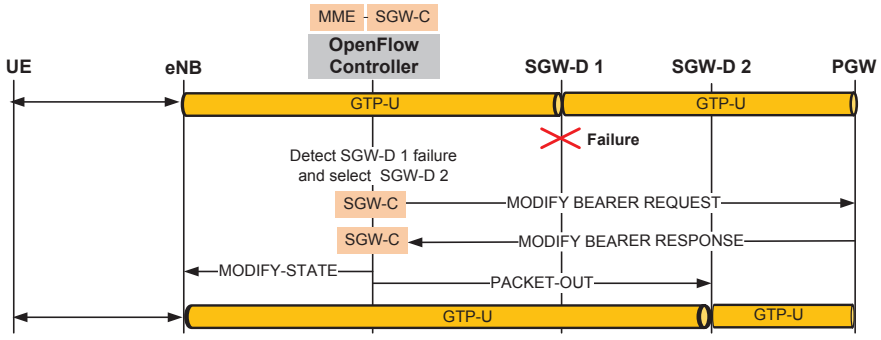


Figure 7.9: Restoration procedure in OF-based architecture.

As the OF-ctr and SGW-Ds exchange periodic Echo Request/Reply messages, the OF-ctr can detect any SGW-D failure. Upon detecting the SGW-D 1 failure, the SGW-C selects SGW-D 2 for the impacted sessions. The SGW-C updates the SGW-D IP address maintained in the PGW via the Modify Bearer Request message. As specified in our architecture, the SGW-TEID values for the downlink traffic on the S5 interface remain the same for the impacted sessions.

After that, the OF-ctr updates the SGW-D IP address in the eNB via the OFPT_MODIFY_STATE message[4].

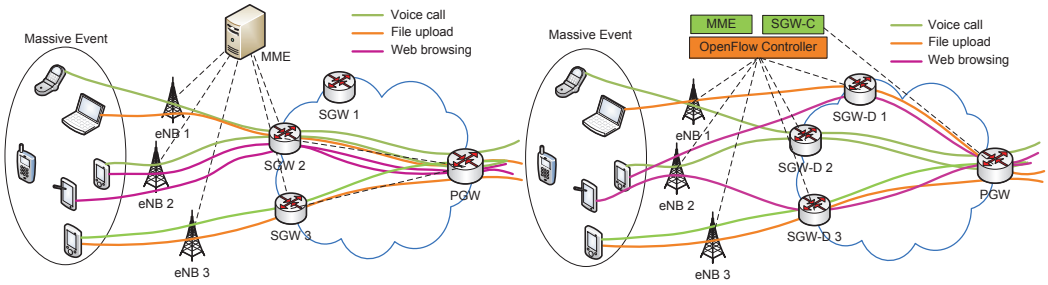
Here, we can see the advantage of centralizing the TEID allocation function related to SGWs. Indeed, the SGW-C does not create new TEID values during the restoration procedure. The OF-ctr updates just flow entries in eNBs with the new SGW-D IP address. As the eNBs remain the same for each session, the eNB-TEID

[4]this message is used by the OpenFlow controller to add, delete and modify the flow entries in the flow table [ONF12b].

values for the downlink traffic in the S1 interface does not change. The OF-ctr inserts the new flow entries in the SGW-D 2 via the OFPT_PACKET_OUT message.

7.5.1.2 Load Balancing in OF-based LTE/EPC architecture

The Load balancing mechanism is required in the LTE/EPC architecture to spread the load across multiple network SGWs and PGWs, thereby preventing bottlenecks. Figure 7.10 shows the LTE/EPC data plane with and without load balancing in case of massive events.



(a) without load balancing (b) with load balancing

Figure 7.10: LTE/EPC Data plane

Getting periodic statistics [ONF12b] about the SGW-D load is one of the advantages of implementing OpenFlow in the EPC architecture. These statistics are crucial for more efficient load balancing. For instance, based on real-time load statistics presented by the OF-ctr and the session type (i.e. determined from packet header), the SGW-C can balance the traffic between SGW-Ds leading to better traffic distribution.

Unlike the 3GPP standards, our architecture can temporarily free the overloaded SGW by moving some sessions seamlessly to another SGW in the same domain. Supposing the SGW-D 2 is overloaded. In that case, the OF-ctr notifies the SGW-C of the SGW-D 2 load status. As the SGW-D 1 and SGW-D 3 are less loaded than SGW-D 2, the SGW-C triggers the OF-ctr to move the flows related to delay-tolerant sessions such file upload and web browsing sessions from the SGW-D 2 to the SGW-D 1 and 3. Then, the OF-ctr just updates flow entries in eNBs and PGWs with the new SGW-D IP address and insert the required flow entries in the SGW-D 1 and 3, respectively.

7.5.2 Reachability service

Till now, we focused more on how providing adaptive Session Continuity service in LTE/EPC architecture. However, the proposed architecture can also ensure adap-

tive Reachability service if we introduce some changes to the proposed initial attachment procedure.

In the proposed architecture, the Reachability service is still natively ensured for all subscribers without exception. It includes the allocation of a permanent IP address and the location update mechanism (i.e. updating the context in MME and the S5 data bearer). During the initial attachment, the UE gets a permanent IP address and an entry is created for this UE in the PGW. This entry enables the PGW to relays the received packet to the corresponding SGW-D.

To enable an adaptive Reachability service (i.e. activate/deactivate Reachability mechanisms according to the contextual information), we propose to modify the initial attachment procedure. Instead of allocating a permanent IP address and creating an entry in the PGW for each subscriber systematically, the SGW-C application should decide whether the Reachability service is needed. For instance, the SGW-C can decide to deactivate the Reachability mechanisms for a sensor that just connects to the network to send data to his server and does not need to be reached by this server. Thus, whenever the sensor has data to be sent, a temporary IP address is allocated for this sensor and a temporary entry is created in the PGW to serve this sensor. These network resources are released as soon as the sensor communication is terminated.

Moreover, the SGW-C may decide to deactivate the location update mechanism as soon as it detects that subscriber is fairly static. In fact, the OF-ctr may relay the eNB identifier to the SGW-C during the first location update procedures. Upon detecting that the subscriber is always using the same eNB, the SGW-C deactivates the location update procedure and the related mechanisms.

7.6 Implementation challenges

To implement the proposed architecture, several challenges should be overcome, as listed next:

- The SGW control functions, such as the TEID allocation, should be first separated from the data forwarding plane. These control functions should run as applications on top of the OF controller. Similarly, the MME functions should be turned into applications that also run on the top of the OF controller. Likewise, the SGW and PGW selection functions should be shifted into another application.

- The OF controller should have a global vision of its domain topology and a real-time knowledge of the network equipment characteristics, such as the load. This is required for the SGW and PGW selection. We need thus to assess the

implications from an implementation point of view, e.g., the trade-off between consistency and high availability.

- Flow inter-arrival times have significant implications for the scalability of this proposal. Regarding the number of eNBs, SGWs and PGWs that could be implemented in such architecture, a centralized OF controller should be able to process a significant number of flows at the same time. Actually, a study on traffic characteristics in data centers [BAM10] shows that if OF switches are used within these data centers, the OF controller can face the challenge to process 10 million flows per second. To scale the throughput of a centralized controller while supporting a complex decision-making process, a controller with multiple CPUs or multiple controllers should be employed. For instance, [TCKS09] showed that by making a number of controllers acting in parallel for a large network topology, 20 million flows can be handled per second.

- The SGW-C is likely to need a large database to store the required information about the networking state under its domain of control (such as active flow entries, allocated TEID values, etc.). Consequently, appropriate memory, IO, and CPU capabilities are required to store such information and to calculate the adequate handling for each session (e.g., decisions for routing, mobility, and QoS treatments). Therefore, we think that the OF controller as well as the applications on top of it (MME and SGW-C) should be implemented in a cloud-like infrastructure.

- The OF protocol should be extended to transport the UE-MME exchanges transparently, e.g. the authentication exchanges. Moreover, the OF switch should be extended with the GTP encapsulation/decapsulation functions. For example, the current port data structure in the OF switch does not contain the GTP parameters, namely the destination and the source TEID values.

- The latency imposed by a controller on the new flow is another challenge in this architecture. Actually, [CFP+07] showed that OF controllers take approximately 10 ms to install flow entries for new flows when the decision is made at flow start up time. This imposes 10 % of delay overhead on flows with 100 ms. Such overhead can be acceptable for delay-tolerant applications such FTP download, but it is unacceptable for application sensitive to delay such as emergency calls.

7.7 Preliminary evaluation

In this section, we quantify and compare the signaling load generated by the data plane management procedures in both of 3GPP and OF-based LTE/EPC architec-

tures.

7.7.1 Signaling Cost Formulation

We assume that each UE is a smart phone supporting n sessions, for example browsing, email, SMS and voice calls, etc. Let λ_n be the average arrival rate of type-n sessions per UE and μ_n^{-1} be the average session duration in the network. Further, let $D_{e,c}$, $D_{c,s}$, $D_{s,p}$, and $D_{c,p}$ denotes the hop distances, i.e. the number of hops, between (eNB, MME/OF-ctr), (MME/OF-ctr, SGW/SGW-D), (SGW, PGW), and (OF-ctr, PGW) respectively. The hop distance is assumed to be symmetric ($D_{e,c} = D_{c,e}$, $D_{c,s} = D_{s,c}$, $D_{s,p} = D_{p,s}$). In our evaluation, we do not consider the authentication, attach request/response, and NAS service requests messages as we assume that their sizes stay the same in the two architectures. In addition, the signaling related to the radio bearer setup (e.g. RRC Connection establishment messages) is not considered in our evaluation.

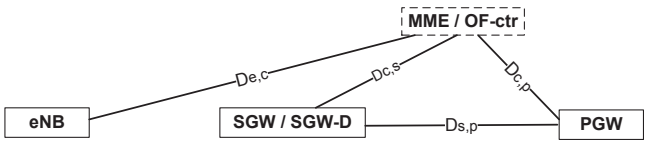


Figure 7.11: System model.

Table 7.1 provides the 3GPP LTE/EPC message sizes which are determined from 3GPP specifications [3GP11b], [3GP12c] and [3GP10c]. Table 7.2 provides the OpenFlow LTE/EPC message sizes. These messages are extensions to OpenFlow protocol [ONF12b] with respect to 3GPP specifications.

7.7.1.1 3GPP LTE/EPC architecture

We consider the initial attachment, access bearer setup and access bearer release procedures to assess the unit signaling load related to the current LTE/EPC architecture. Each time, we assume that the session is successfully established and no failure arises during any session management procedure. The unit signaling load is evaluated at four scenarios and calculated as the product of the transmitted message size and the traveled hop distance [LEC10].

Scenario 1: The UE is not registered with the network. The session arrival triggers the UE initial attachment process as shown in Figure 3.8. The unit signaling

Message	src-dst	Notation	Size (bytes)
Initial Attachment procedure			
Initial Context Setup Request	MME - eNB	M_{icsq}	145
Initial Context Setup Response	eNB - MME	M_{icsr}	86
Create Session Request	MME - SGW	M_{csq}	335
Create Session Response	SGW - MME	M_{csr}	241
Create Session Request	SGW - PGW	$M_{csq'}$	335
Create Session Response	PGW - SGW	$M_{csr'}$	224
Modify Bearer Request	MME - SGW	M_{mbq}	101
Modify Bearer Response	SGW - MME	M_{mbr}	81
Modify Bearer Request	SGW - PGW	$M_{mbq'}$	67
Modify Bearer Response	PGW - SGW	$M_{mbr'}$	81
Access Bearer Setup (Idle to connect state)			
Initial Context Setup Request	MME - eNB	MT_{icsq}	145
Initial Context Setup Response	eNB - MME	MT_{icsr}	86
Modify Bearer Request	MME - SGW	MT_{mbq}	104
Modify Bearer Response	SGW - MME	MT_{mbr}	106
Modify Bearer Request	SGW - PGW	$MT_{mbq'}$	65
Modify Bearer Response	PGW - SGW	$MT_{mbr'}$	81
Access Bearer Release (S1 Release Message)			
UE Context Release Request	eNB - MME	M_{crq}	67
UE Context Release Command	MME - eNB	M_{crd}	67
UE Context Release Complete	eNB - MME	M_{crte}	65
Release Access Bearers Request	MME - SGW	M_{rabq}	65
Release Access Bearers Response	SGW - MME	M_{rabr}	66

Table 7.1: The 3GPP LTE/EPC data plane management messages and sizes

Message	src-dst	Notation	Size (bytes)
OF Initial Attachment procedure			
Create Session Request	SGW - PGW	$M_{csq'}$	335
Create Session Response	PGW - SGW	$M_{csr'}$	224
OF Initial Access Bearer Setup			
OF Initial Context Setup Request	OF-ctr - eNB	M_{Oicsq}	78
OFPT˙PACKET˙IN˙eNB	eNB - OFctr	M_{OPin}	104
OFPT˙PACKET˙OUT˙eNB	OFctr - eNB	M_{OPout}	178
OFPT˙PACKET˙OUT˙SGW	OFctr - SGWD	$M_{OPout'}$	178
Modify Bearer Request	OF-ctr - PGW	$M_{mbq'}$	67
Modify Bearer Response	PGW - OF-ctr	$M_{mbr'}$	81

Table 7.2: The OF-based LTE/EPC data plane management messages and sizes

cost is given by

$$\begin{aligned} SC_1^{3gpp} &= (M_{icsq} + M_{icsr})D_{e,c} \\ &+ (M_{csq} + M_{csr} + M_{mbq} + M_{mbr})D_{c,s} \\ &+ ((M_{csq'} + M_{csr'} + (M_{mbq'} + M_{mbr'})P_{ho})D_{s,p} \end{aligned} \tag{7.1}$$

Where P_{ho} denotes the probability that the UE hands over from non-3GPP access to 3GPP access.

Scenario 2: The UE is successfully registered with the network but is in IDLE state. The session arrival triggers the access bearer (i.e. radio and S1 data bearers) setup procedure as shown in Figure 3.9. The S5 data bearer is already established. The unit signaling cost SC_2^{3gpp} of this scenario is given by

$$\begin{aligned} SC_2^{3gpp} &= (M_{icsq} + M_{icsr})D_{e,c} \\ &+ (M_{mbq} + M_{mbr})D_{c,s} \\ &+ ((M_{mbq'} + M_{mbr'})P_{ho})D_{s,p} \end{aligned} \tag{7.2}$$

Scenario 3: The UE is successfully registered with the network and is in CONNECTED state. The new session uses the existing bearer. Therefore, this scenario generates no signaling load ($SC_3^{3gpp} = 0$).

Scenario 4: The UE is in CONNECTED state. The termination of the ongoing sessions triggers the access bearer release procedure and moves the UE from CONNECTED to IDLE states. The unit signaling cost for access bearer release process is given by

$$\begin{aligned} SC_4^{3GPP} &= (M_{crq} + M_{crd} + M_{crte})D_{e,c} \\ &+ (M_{rabq} + M_{rabr})D_{c,s} \end{aligned} \tag{7.3}$$

We assume that each UE supports N types of application such as web browsing, SMS/MMS, email, voice call, etc. Let λ_n be the average arrival rate of type-n session at the UE and μ_n denotes the average type-n session duration. The total signaling cost per UE is calculated as follows:

$$\begin{aligned} SC_{total-ue}^{3gpp} &= \lambda_n\{(SC_1^{3gpp} + SC_4^{3gpp})P_1^{3gpp} \\ &+ (SC_2^{3gpp} + SC_4^{3gpp})P_2^{3gpp} + SC_3^{3gpp}P_3^{3gpp}\} \end{aligned} \tag{7.4}$$

Where P_1^{3gpp}, P_2^{3gpp} and P_3^{3gpp} denotes the probability that the session begins when the UE is in Scenario 1, Scenario 2 and Scenario 3 respectively. Actually, P_2^{3gpp} and P_3^{3gpp} correspond to the probability that the UE is in IDLE and CONNECTED states, respectively. To compute these two probabilities, we note that the process (X_n, Y_n) can represent CONNECTED and IDLE states related to type-n session with $E[X_n] = \mu^{-1}$ and $E[Y_n] = \lambda^{-1}$. From the theory of alternating renewal process and independence assumption of applications, we have $P_2^{3gpp} = P_{idle} = \prod_{n=1}^{N} \frac{\mu_n}{(\lambda_n+\mu_n)}$. Moreover, the UE is in CONNECTED state if it has at least one active session. Therefore, we have $P_3^{3gpp} = 1 - P_2^{3gpp}$.

If we consider N_{ue} users, the total signaling cost is given by:

$$SC_{total}^{3gpp} = N_{ue}SC_{total-ue}^{3gpp} \tag{7.5}$$

7.7.1.2 OF-based LTE/EPC architecture

In the proposed architecture, we consider the OF Initial Attachment and Initial Access Bearer Setup procedures. The units signaling cost is evaluated at three scenarios:

Scenario 1: The UE is not registered with the network. The session arrival triggers the Initial Attachment and Initial Access Bearer setup procedures as shown in Figure 7.6 and Figure 7.7. The unit signaling cost of this scenario is given by

$$\begin{aligned} SC_1^{of} &= (M_{Oicsq} + M_{OPin} + M_{OPout})D_{e,c} \\ &+ M_{OPout'}D_{c,s} \\ &+ (M_{csq'} + M_{csr'} + (M_{mbq'} + M_{mbr'})P_{ho})D_{c,p} \end{aligned} \tag{7.6}$$

Scenario 2: The UE is already registered with the network but has no data bearer maintained in the network (i.e. S1 and S5 data bearers). Therefore, the session arrival triggers the Initial Access Bearer setup procedure as shown in Figure 7.7. This scenario represents the first time the application is launched. After that, the S1 and S5 data bearers will be maintained. The unit signaling cost related to this scenario is given by

$$\begin{aligned} SC_2^{of} &= (M_{Oicsq} + M_{OPin} + M_{OPout})D_{e,c} \\ &+ M_{OPout'}D_{c,s} \\ &+ ((M_{mbq'} + M_{mbr'})P_{ho})D_{c,p} \end{aligned} \tag{7.7}$$

Scenario 3: The UE is registered with the network. It is in IDLE state but the S1 and S5 data bearers are maintained in the network. The session arrival triggers just the radio data bearer setup as shown in Figure 7.8. This scenario represents the cases where the application has used the network. The new session related to the same application uses the existing S1 and S5 data bearers and generates no signaling load at eNB-MME and MME-SGW interfaces ($SC_3^{of} = 0$).

The total signaling cost per UE is calculated as follows:

$$SC_{total-ue}^{of} = \lambda_n(SC_1^{of}P_1^{of} + SC_2^{of}P_2^{of} + SC_3^{of}P_3^{of}) \tag{7.8}$$

where P_1^{of}, P_2^{of} and P_3^{of} denotes the probability that the session starts when the UE is in Scenario 1, Scenario 2 and Scenario 3 respectively. P_2^{of} is calculated as the probability that the UE is in IDLE state (P_{idle}) by the probability that the S1 and S5 data bearers are not established in the network ($P_{no\ S1\ \&\ S5\ bearers}$). This last probability represents the first launch of the application.

$$P_2^{of} = P_{idle}P_{no\ S1\ \&\ S5\ bearers} \tag{7.9}$$

Similarly, If we consider N_{ue} users, the total signaling cost is given by

$$SC_{total}^{of} = N_{ue}SC_{total-ue}^{of} \tag{7.10}$$

7.7.2 Numerical results and discussions

In this section, we present and discuss the numerical results showing the impact of using OpenFlow in LTE/EPC architectures. The default values of the system parameters are assumed to be as follows: $D_{e,c} = 3$, $D_{c,s} = 1$, $D_{s,p} = 1$, $P_1^{3gpp} = P_1^{of} = 0.2$, $P_{no\ S1\ \&\ S5\ bearers} = 0.1$, and $P_{ho} = 0$ (i.e we assume that no handover takes place during our evaluation).

First, we investigate the impact of the number of UEs on the signaling load. We vary the number of UEs from 0 to 1000 as shown in Figure 7.12. As we expected, the proposed architecture does not increase the signaling load compared to the 3GPP LTE/EPC architecture. On the contrary, it could even decrease the signaling load.

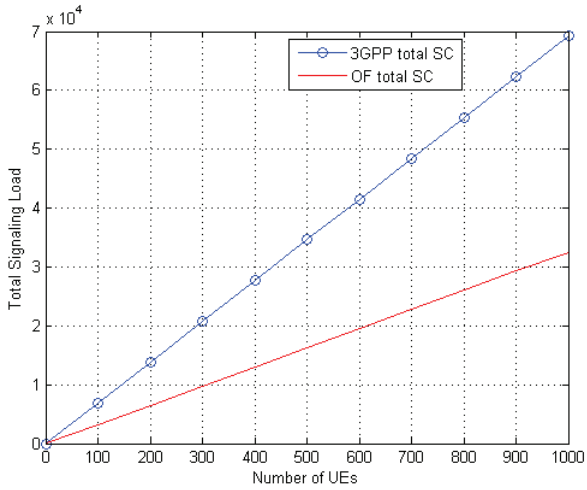


Figure 7.12: The impact of N_{ue} on SC ($\lambda_n = 0.05$).

We assume two different types of background applications namely Chat and Email applications with their associated session durations $\mu_1 = 0.01$ and $\mu_2 = 0.05$, respectively. Actually, the Chat applications sends notifications periodically to update contacts' status such as in Skype or Whatsapp. The Email application connects periodically to their server to get new emails. To examine the impact of the average session arrival rate of each of these applications on the signaling cost, we vary λ_n from 0 to 0.1 per second.

Figure 7.13 compares the signaling load for both architectures. As we expect, the signaling load increases with the increase of the average session arrival rate. We note that the signaling load in 3GPP LTE/EPC architectures varies according to the application types. This reveals the impact of the session duration on the signaling load. As we can see, signaling load increases with the decrease of the session duration. For instance, the Chat application presents more signaling load than Email

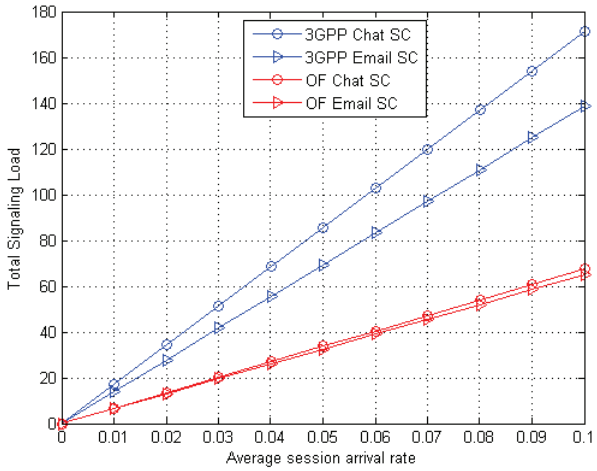


Figure 7.13: The impact of λ_n on SC.

application. Indeed, for application with longer session duration, the UE presents less alternation between CONNECTED and IDLE states (i.e. the UE more often stays connected) whereas for application with shorter session duration, the UE will more often come back to idle state before a new session arrives. Indeed, the alternation between CONNECTED and IDLE states takes place more frequently with applications presenting short session duration. Therefore, in 3GPP LTE/EPC, the access bearers are released and re-established more often for this type of applications leading to higher signaling load.

We note that our proposal presents almost the same signaling load for each type of applications (Figure 7.13). This trend is due to adapting the data plane *Release Timer* to the flow IDLE period in eNB and SGW. For example, the eNB and SGW maintain the S1 data bearer as long as the application is in IDLE state, i.e. the data plane *Release Timer* is slightly higher than the average application IDLE period. Therefore, when a new session related to the same application arrives, the UE just establishes the radio data bearer with eNB without inducing extra signaling load at the network side. Obviously, maintaining the data plane parameters (flow entries in eNB and SGW-D) for a long period drives the need for memory spaces in network equipment. Consequently, the controller should set for each flow entry the optimal value of the *Release Timer* that generates less signaling load and avoids the unnecessary usage of the memory space.

7.8 Related Work

There are several papers that call for the redesign of the LTE/EPC architecture. Particularly, [MEV13, JLLJ13, KJP+12] are the closest studies to our work.

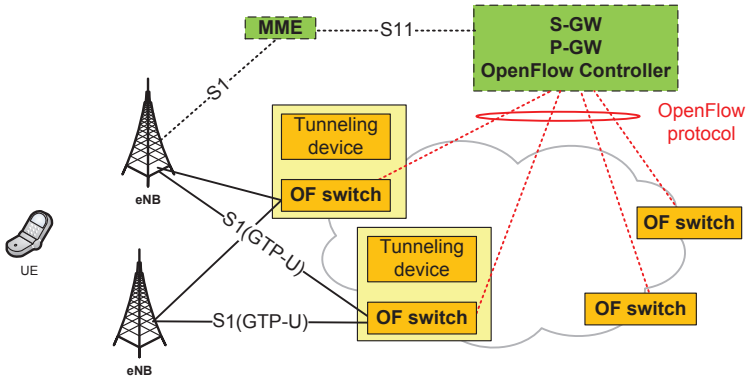


Figure 7.14: Mevico proposal

In MEVICO project [MEV13], contributors proposed to split the SGW and PGW functionalities in a new way as shown in Figure 7.14. The intelligence and decision making is centralized in the combined S/P-GW controller. The packet forwarding function is distributed among OpenFlow switches. The S/P-GW controller uses the OpenFlow protocol to enforce the forwarding policies in the OpenFlow switches. This S/PGW controller still use the standard 3GPP interfaces to communicate with other network nodes like MME. The GTP-U tunneling is still used in the S1 interface. Therefore, a tunneling device should be added to the OpenFlow switches to encapsulate/decapsulate GTP packet related to the S1 interface. In this proposal, the 3GPP restoration procedures related to SGW failure are still applicable whenever the OpenFlow switch that is right connected to the eNB fails. That means that the fast and transparent recovery cannot be ensured for the user connectivity.

A recent paper [KJP+12] proposed to move the MME functions, the control plane of the SGW (SGW-C) and the PGW (PGW-C) into separated virtual machines and lifted up to the cloud. The proposed architecture is depicted in Figure 7.15. They replaced the data plane of the EPC core by OpenFlow switches. The same standard 3GPP interfaces are still used between these control entities. The PGW-C is connected to OpenFlow controller via RPC/APIs. The OpenFlow controller is responsible for data plane establishment. The S1 interface between the eNB and the MME is kept the same as in the 3GPP specification. The OpenFlow controller cannot update the flow entries in the eNBs directly. Consequently, any SGW-D modification during the user session stays a difficult task as the OpenFlow controller has no means to promptly update the flow entry in the eNB. In fact, as the exchange

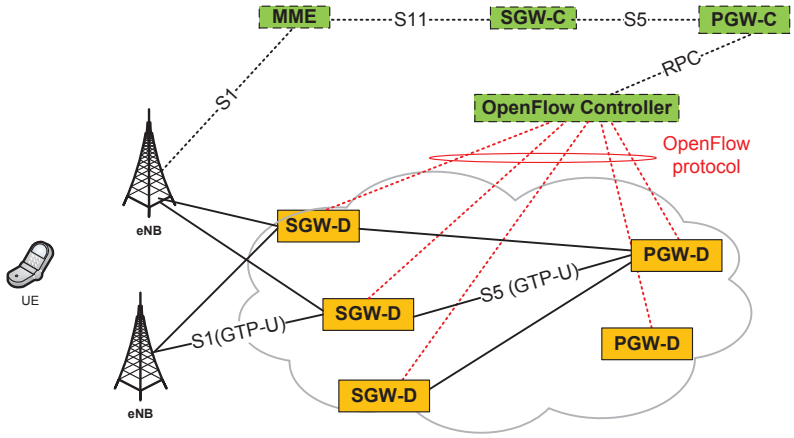


Figure 7.15: Ericsson proposal

between the control entities still relies on the 3GPP standard protocols, the on-demand connectivity service is hard to be ensured.

[JLVR13] reshapes completely the cellular networks by removing the mobility tunnels within the transfer plane and using, instead, fine-grained policies. The paper introduces the SoftCell controller to calculate the UE shortest path, select the adequate middle-boxes (e.g. firewall, transcoder, etc.) and generate the adequate policies. This controller communicates with the switches in the transfer plane via the OpenFlow protocol. It is logically centralized and ensures services by directing traffic through a sequence of middle-boxes. In case of UE mobility, the SoftCell ensures the Session Continuity service by installing the adequate forwarding rules in the new eNB.

In our approach, we are different from [JLVR13] because we propose a new control plane based on the OpenFlow protocol while maintaining the use of GTP tunnels within the transfer plane. Unlike [MEV13] and [KJP+12], our proposal enables transparent recovery procedure for any SGW-D failures. Moreover, the load balancing techniques can be easily implemented in the EPC architecture using the proposed control plane.

7.9 Conclusion

In this chapter, we proposed an OpenFlow-based control plane for LTE/EPC architecture. Particularly, this architecture splits between the control and data forwarding planes related to the Serving Gateways (SGWs). The SGW control plane is centralized and uses the OpenFlow protocol to remotely manage the SGW data forwarding plane. This feature guarantees adaptive mobility services and flexible use of network resources. We showed that the proposed architecture can easily en-

sure adaptive Continuity service even in critic situation such as network equipment failure and overload situations. The same architecture can be extended to provide adaptive Reachability service as well.

This work has many perspectives. First, the performances of the OF-based control plane may be assessed. Then, the proposal should be implemented in a test bed in order to validate the above key aspects namely adaptive Session Continuity and Reachability services.

Chapter 8

Conclusion and Perspectives

8.1 Thesis summary and contributions

In this thesis, our quest was to build a new model for the network connectivity management in multi-access architectures to address the new ecosystem challenges while using network resources in a more efficient manner. Different steps, given below, have been necessary to come out with such model.

Specifying architectural requirements to face the new ecosystem challenges.

In Chapter 2, we decomposed the network connectivity into several network services. Particularly, we were interested in security and mobility services and the related mechanisms. Then, we reviewed the current access networks and identified how these services are ensured in each access networks.

In Chapter 3, we described the challenges brought by the new ecosystem. Through network usage scenarios, we underlined several architectural requirements. These requirements are as follows:

- Requirement 1: The network connectivity in multi-access architectures need to be context-aware.

- Requirement 2: The network connectivity in multi-access architectures should be able to adjust the network mechanisms according to the real needs. To this end, the network connectivity should be modular, where the network mechanisms should be easily activated, deactivated or configured.

- Requirement 3: The network connectivity adaptation in multi-access architectures should be decided at the network side. The network mechanisms that will be activated should be selected and orchestrated by a trusted network entity.

The subscriber may assist the decision by providing the required contextual information.

- Requirement 4: A unified connectivity manager is required in multi-access architectures.

- Requirement 5: The network connectivity should be able to use network resources in a flexible and smart manner to mitigate network equipment failure and overload cases.

Chapter 4 supplemented this study by analyzing and developing an analytical model for the security costs within LTE/EPC access. This enabled us to assess the cost of the systematic activation of security mechanisms.

Part of this analysis has been published in [BHSGB13a] [BHSGB12] [BHSGB13c] [BHSGB13b]. In addition, this analysis contributed in producing an internal deliverable about the Distributed and Dynamic Mobility (DMM) impact on the network security.

Proposing a new connectivity management model for multi-access architectures.

The above requirements enabled us to develop a new approach to the connectivity management, **CACM (Context-Aware Connectivity Management)**, where the network mechanisms are activated, deactivated and configured according to the contextual information. This novel approach will enable network operators to customize network connectivity according to the real needs.

Chapter 5 described the CACM functional architecture. To enable the context-awareness features, we included a Context Management Subsystem (ContextMS) in the CACM model to collect, process and present the contextual data to the Connectivity Management Subsystem (ConnectMS). This latter is responsible for adapting the network connectivity to any change in the contextual information; thereby satisfying Requirement 1 and 2. In fact, the ConnectMS selects, orchestrates and launches security and mobility mechanisms according to the contextual information. In addition, it decides the most appropriate actions (e.g. activating/deactivating network mechanisms, move flows from an equipment to another, etc.) when a change in the context occurs. The contextual information are classified into four groups: user-related context (e.g. user profile, device type, etc.), application-related context (e.g. session type, data sensitivity, etc.), network-related context (e.g. network load statistics, the nearest AP characteristics, etc.), and environmental context (e.g. time, geographic map, etc.).

To fulfill Requirement 3, the adaptation decisions are taken in the network-

side. Indeed, the ConnectMS is located in the control plane of the access networks and includes several Decision-Maker algorithms. Regarding Requirement 4, the CACM model was designed to be access-agnostic. Therefore, it is able to manage multiple accesses reducing, therefore, the functional redundancy. Finally, the CACM is designed to use network resources in an efficient manner as it is aware of any change in the network-related context.

To check whether the CACM model fulfill the underlined architectural requirements, we carried out a qualitative evaluation through several network usage scenarios. In addition, we used this evaluation to compare between the CACM model and the current approaches in multi-access 3GPP systems.

As the validation and evaluation of the whole CACM model in real access networks is a complex task, we have provided two specific and concrete application of the CACM approach: adaptive Data Traffic Protection service in non-3GPP access (Chapter 6) and adaptive mobility services in LTE/EPC accesses (Chapter 7).

In Chapter 6, we realized an experimental test bed that reproduces the non-3GPP accesses. In this test bed, we implemented a mechanism that enables adaptive Data Traffic Protection service in untrusted non-3GPP access. This mechanism inspects the subscriber sessions and detects their security status (i.e. protected or non-protected). It activates/deactivates the encryption and integrity protection mechanisms depending on the session security status.

As a second application of the CACM model, Chapter 7 proposed a Mobility Manager module for the LTE/EPC access. As the trend in access network architectures goes forward on separation of network functionalities in network equipments into control and data forwarding planes, the proposed module includes the MME and SGW control functionalities and is located in the LTE/EPC control plane. It uses the OpenFlow protocol to remotely manage the data forwarding plane (i.e. eNB and SGW data plane). The proposed Mobility Manager is able to retrieve real time data path information from OpenFlow-enabled equipments and is therefore able to precisely identify equipment failure or overload situations. Moreover, due to the use of OpenFlow, this module is able to move different flows from one equipment to another in a flexible manner.

A first description of the Security Manager was presented in [BHSGB13b]. Regarding the application of the Mobility Manager in LTE/EPC architecture, we filed a patent with the INPI [SBHSGS13]. Likewise, we described in details this Mobility Manager in [BHSSG+13] and provided a preliminary evaluation in [SBHSGS14].

8.2 Perspectives

Although we endeavored in this thesis to address various design and optimization aspects of the network connectivity management in multi-access architectures, future

research is still required at the architecture design, performance evaluation, and implementation levels. Relevant to the first, further work is still needed to specify a complete CACM model (i.e. specifying interfaces, protocols and algorithms), flexible, scalable and powerful enough to encompass current and new network mechanisms. Moreover, we reckon that the modular approach in the CACM model, not only enables customized network connectivity, but it also prepares us for the upcoming Software Defined Networking (SDN) and Network Function Virtualization (NFV) world where most of the control functions and mechanisms can be seen as modules and implemented mainly in software (e.g., see [BHSSG+13]).

Second, at the performance evaluation level, further work is still needed to develop an analytical model for the proposed CACM in the context of multi-access architectures. This model should consider the appropriate mobility model and the connectivity request arrival distribution. However, the first entails complexity in characterizing the appropriate distribution model for the cell residence time while considering different mobility profiles (e.g. highly mobile subscribers, fairly static subscribers, etc.) and different access technologies. Characterizing connectivity request arrival distribution is also complex as it entails the consideration of several applications profiles (e.g. connections with short duration, connections with small packet, infrequent connections, etc.). Moreover, this analytical model should take into consideration the probability of obtaining the accurate contextual information that lead to the adequate decisions. In addition, with respects to security considerations, the security threats should be analyzed for the CACM model.

Third, from an implementation perspective, further work is required to extend the experimental test bed that was presented in Chapter 6 with new mechanisms to customize the remainder of security services. Moreover, future work is also needed to implement the mechanisms that were proposed in Chapter 7 to ensure adaptive mobility services. To do this, we can use the already developed test bed and the freely available open source packages such as Open vswitch [Swi] and Floodlight controller [Con]. From a testing and validation perspective, further work is required to implement connectivity request generation tool based on realistic data measurement or using analytic models. This allows testing the performance of the CACM model. The connectivity request generation tool shall rely on measurements of the user's mobility profiles as well as application characteristics.

To sum up, in this thesis, we have specified several architectural requirements, developed analytical security costs, designed a new model for the network connectivity management, and proposed two applications for the model in LTE/EPC and non-3GPP accesses. The results demonstrate that our model is efficient and flexible. Future research avenues for this thesis fall in the areas of specifying the CACM module interfaces as well as the protocols that should be employed within these interfaces, developing analytical and simulation models to evaluate the proposed

solution, and validating the work using real traffic records in representative test bed.

Appendix A

List of publications

Journal

- Aymen Belghith, Siwar Ben Hadj Said, Bernard Cousin, Samer Lahoud. Collaboration Schemes Evaluation in Multi-domain Networks. International Journal of Computer Science Issues (IJCSI), 2012 [BBHSCL12a]

- Aymen Belghith, Siwar Ben Hadj Said, Bernard Cousin, Samer Lahoud. Export Methods in Fault Detection and Localization Mechanisms. International Journal of Computer Science Issues (IJCSI), 2012 [BSCL12]

International Conferences

- Malla Reddy Sama, Siwar Ben Hadj Said, Karine Guillouard and Lucian Suciu. Enabling Network Programmability in LTE/EPC Architecture Using OpenFlow. In *12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, Hammamet, TUNISIA, May 2014 [SBHSGS14].

- Siwar Ben Hadj Said, Malla Reddy Sama, Karine Guillouard, Lucian Suciu, Gwendal Simon, Xavier Lagrange and Jean-Marie Bonnin. New Control Plane in 3GPP LTE/EPC Architecture for On-Demand Connectivity Service. In *2nd IEEE International Conference on Cloud Networking (CLOUDNET 2013)*, San Francisco, USA, November 2013 [BHSSG+13].

- Siwar Ben Hadj Said, Karine Guillouard and Jean-Marie Bonnin. On the benefit of context-awareness for security mechanisms in LTE/EPS networks. In *IEEE Personal Indoor and Mobile Radio Communications (IEEE PIMRC'13)*, London, United Kingdom, September 2013 [BHSGB13b].

- Siwar Ben Hadj Said, Karine Guillouard and Jean-Marie Bonnin. Towards Adaptive Security Mechanisms in 3GPP EPS/LTE Networks. In *IEEE Wireless Communications and Networking Conference (IEEE WCNC'13)*, Shanghai, China, April 2013 [BHSGB13c].

- Siwar Ben Hadj Said, Karine Guillouard and Jean-Marie Bonnin. On The Need for Adaptive Connectivity Management in Multi-Access Architectures. In *Network of the Future*, Tunis, Tunisia, Novembre 2012 [BHSGB12].

- Aymen Belghith, Siwar Ben Hadj Said, Bernard Cousin, Samer Lahoud. QoS fault detection and localization mechanisms (FDLM) in multi-domain networks adapted to export methods. In 14th International Conference on Advanced Communication Technology (ICACT'12), PyeongChang, Korea (South), February 2012 [BBHSCL12c].

- Aymen Belghith, Siwar Ben Hadj Said, Bernard Cousin, Samer Lahoud. Proactive and reactive collaboration schemes for multi-domain networks monitoring. In Computing, Communications and Applications Conference (ComComAp'12), Hong Kong, China, January 2012 [BBHSCL12b].

- Aymen Belghith, Bernard Cousin, Samer Lahoud, Siwar Ben Hadj Said. Proposal for the configuration of multi-domain network monitoring architecture. In International Conference on Information Networking (ICOIN'11), Barcelona, Spain, January 2011. 2011[BCLBHS11].

Book Chapter

- Siwar Ben Hadj Said, Karine Guillouard and Jean-Marie Bonnin. A Comparative Study on Security implementation in EPS/LTE and WLAN/802.11. In: *Wireless Networks and Security*, Springer, 2013 [BHSGB13a].

Patent

- Malla Reddy Sama, Siwar Ben Hadj Said, Karine Guillouard and Lucian Suciu New Control Plane in 3GPP LTE/EPC Architecture for More Flexible and Reliable Connectivity, Filed in May 2013 [SBHSGS13].

Appendix B

Acronyms

3GPP	the Third Generation Partnership Project
AAA	Authentication, Authorization, and Accounting
AC	Access Control
AH	Authentication Header
AN	Access Network
AN-CP	Access Network - Control Plane
AN-TP	Access Network - Transfer Plane
AP	Access Point
APN	Access Point Name
AS	Authentication Server
ASP	Application Service Provider
BNG	Broadband Network Gateway
CACM	Context-Aware Connectivity Management
CDN	Content Delivery Network
CN	Core Network
ConnectMS	Connectivity Management Subsystem
ContextMS	Context Management Subsystem
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name System
DSL	Digital Subscriber Line
DSLAM	Digital Subscriber Line Access Multiplexer
DTP	Data Traffic Protection

EAP	Extensible Authentication Protocol
EAP-AKA	EAP-Authentication and Key Agreement
EAP-TLS	EAP-Transport Layer Security
EAP-TTLS	EAP-Tunneled Transport Layer
eNB	evolved Node B
EPC	Evolved Packet Core
ePDG	evolved Packet Data Gateway
EPS	Evolved Packet System
EPS-AKA	EPS-Authentication and Key Agreement
ESP	IP Encapsulating Security Payload
F-AN	Fixed - Access Network
GGSN	Gateway GPRS Support Node
G-MSC	Gateway Mobile Switching Center
GPRS	General Packet Radio Service
GSM	Global System For Mobile communication
GTP	GPRS Tunneling Protocol
HLR	Home Location Register
HSS	Home Subscriber System
IKEv2	Internet Key Exchange version 2
IMS	IP Multimedia Sybsystem
IMSI	International Mobile Subscriber Identity
IMPI	IMS Private Identity
IMPU	IMS Public Identity
IPsec	IP security
ITS	Intelligent Transportation System
LTE	Long Term Evolution
MAC	Medium Access Control
M-AN	Mobile - Access Network
MCD	Mobility Context Database
MIP	Mobile IP
MME	Mobility Management Entity
MSD	Mobility Status Database
MSC	Mobile service Switching Center
MTC	Machine Type Communications

NAGw	Network Access Gateway
NAS	Non-Access Stratum
NFV	Network Function Virtualization
NSGw	Network Service Gateway
NSP	Network Service Providers
OF	OpenFlow
PCRF	Policy Control and Charging Rules Function
P-CSCF	Proxy-Call Session Control Function
PDCP	Packet Data Convergence Protocol
PDN	Packet Data Network
PGW	PDN Gateway
PKI	Public Key Infrastructure
PMIP	Proxy Mobile IP
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
RG	Residential Gateway
RNC	Radio Network Controller
RRC	Radio Resource Control
SCD	Security Context Database
S-CSCF	Serving-Call Session Control Function
SDN	Software Defined Networking
SGSN	Serving GPRS Support Node
SGW	Serving Gateway
SIP	Session Initiation Protocol
SPR	Subscriber Profile Repository
SSD	Security Status Database
SSL	Secure Sockets Layer
SM	Security Manager
TA	Tracking Area
TAU	Tracking Area Update
TEID	Tunnel Endpoint Identifier

UE	User Equipment
UMTS	Universal Mobile Telecommunication System
USIM	Universal Subscriber Identity Module
VISP	Virtual Internet Service Provider
VLR	Visitor Location Register
VM	Virtual Machine
VPN	Virtual Private Network
WAG	Wireless Access Gateway
WAP	Wi-Fi Access Point
WEP	Wired Equivalent Privacy
WF	Weight Factor
WISP	Wireless Internet Service Provider
WLAN	Wireless Local Area Network
WPA	Wi-Fi Protected Access

Appendix C

Résumé en Français

C.1 Problématique

Un rapport de Cisco prévoit une augmentation exponentielle du trafic de données d'ici 2017 [Cis13]. Cette croissance sera alimentée par de nouvelles applications (par exemple compteurs intelligents, jeux en ligne, services multimédia, etc.), la variété des appareils connectés (e.g Smart Phones, tablettes, capteurs, les voitures connectées, les villes intelligentes, etc.), et aussi l'omniprésence de réseaux d'accès à haut débit.

Une étude récente [Tel11] prévoit que le coût total de déploiement et du maintenance d'un réseaux d'accès dépassera très bientôt la recette totale du fournisseur d'accès (mobile et/ou fixe) malgré l'augmentation du nombre des abonnés. Cette constatation peut s'expliquer par le fait que les architectures de réseau existantes sont mal équipées pour faire face au double défi de la croissance du trafic des données et la baisse des revenus. En effet, les coûts de déploiement, d'exploitation et de la modernisation des équipements dans ces architectures sont en croissance alors que les revenus ramenés au nombre des clients sont en baisse [PMJ13]. À la lumière de ces prévisions, les fournisseurs d'accès sont invités à revoir la conception et les capacités de leurs architectures avec un double objectif de réduire les dépenses et introduire des nouveaux services générateurs de revenus.

La gestion de la connectivité dans les architectures multi-accès devient une question primordiale parce que ces architectures devraient être en mesure de faire fonctionner des technologies hétérogènes et de faire face aux défis imposés par le nouvel écosystème. L'organisme de normalisation 3GPP a proposé un système 3GPP multi-accès qui vise à fournir une connectivité omniprésente. Cette proposition a indubitablement des avantages, mais apporte également plusieurs défis aux fournisseurs d'accès. En effet, dans ce système, les mécanismes du réseau, telles que les mécanismes de gestion de la mobilité et de sécurité sont conçus pour être activés

d'une manière systématique augmentant ainsi le coût d'exploitation du réseau. Par exemple, le mécanisme de la mise à jour de la localisation est activé en permanence, même lorsque les terminaux ne sont pas mobiles. Cependant, les fournisseurs d'accès doivent relever le défi d'accueillir plusieurs catégories d'abonnés tels que des abonnés non-mobiles, des abonnés à forte mobilité, des abonnés qui demande un niveau de sécurité élevé, des abonnés qui sont satisfaits avec un niveau de sécurité faible, des abonnés demandant une large/faible bande passante, etc.

L'émergence de nouvelles technologies d'accès et l'apparition de divers terminaux miniaturisés et personnalisés a donné lieu à une variété de nouveaux usages. Par exemple, nos maisons peuvent être équipées par des capteurs qui surveillent en permanence l'éclairage, la climatisation et la consommation d'énergie. Les patients ont besoin de petits capteurs qui surveillent leur état de santé et qui remontent périodiquement les mesures vers un centre médical. Dans le cadre de l'applicabilité de la vidéo-surveillance dans les lieux publics, des caméras sont installées dans les rues, les centres commerciaux, et les bus. Éventuellement, ces caméras peuvent être amenées à remonter périodique ou sur demande des extraits de vidéos.

Dans un tel environnement dynamique et hétérogène, la connectivité ne peut plus rester intacte vis-à-vis la diversification des besoins des utilisateurs. En effet, l'établissement d'une connectivité dans les architectures des réseaux implique l'activation et la coordination de plusieurs mécanismes du réseau tel que les mécanismes de mobilité, de la sécurité et du contrôle de la qualité de service (QoS). Ainsi, quand l'utilisateur demande une connectivité, tous les mécanismes du réseau sont activés d'une manière systématique quels que soit le contexte de l'utilisateur (type de son terminal, besoins de l'application en exécution, ses préférences, etc.) et quel que soit l'état du réseau (état de surcharge, indicateur de congestion).

Par exemple, dans le réseau cellulaire, les services de mobilité qui incluent la continuité de session, la joignabilité et le nomadisme sont activés d'une manière systématique pour tous les utilisateurs. Ceci induit la sélection des équipements dans le réseau qui joueront le rôle des points d'ancrage et l'établissement systématique des tunnels de mobilité IP entre ces points. Un ensemble des paramètres nécessaires pour l'aiguillage du trafic au niveau de ces points d'ancrage doit être maintenu. Le maintien de ces paramètres requiert des mises à jour périodiques. La mise en oeuvre du service de mobilité pour tous les utilisateurs indépendamment de leurs vrais besoins surcharge le réseau avec la nombre des messages de signalisation, les entêtes d'encapsulation liées aux tunnels de mobilité, et les paramètres à maintenir au niveau des équipements réseau. Ceci conduit à des coûts d'exploitation et d'investissement extrêmement élevés. Par contre, il existe des cas où l'utilisateur n'a pas vraiment besoin d'un de ces services de mobilité. Par exemple, l'activation des mécanismes de mobilité est inutile pour un étudiant qui passe deux heures dans la bibliothèque et consulte parfois l'Internet avec son Smart Phone (pour chercher une définition ou

consulter sa boîte mail).

Dans les approches actuelles, la gestion de la connectivité ne prends pas en considération les informations contextuelles telles que la localisation de l'abonné, historique de mobilité, le type du terminal, l'application en cours d'exécution, l'état du réseau, etc. Par exemple, il ne prend pas en considération le fait que l'abonné ne sera probablement pas mobile pendant une session donnée (84 % des abonnés sont soit statiques ou nomades au cours d'une session et 16 % des abonnés sont des utilisateurs mobiles [TRKN09]). En outre, une étude récente [PSBD11] montre qu'une grande partie des abonnés ont une mobilité limitée. Il est établi dans la même étude que la mobilité de chaque abonné est prévisible. En se basant sur certaines statistiques, cette étude montre que les utilisateurs passent plus de 60 % de leur temps à la maison ou au travail.

Aujourd'hui, les fournisseurs d'accès devraient trouver de nouvelles solutions pour répondre à ce nouvel écosystème d'une manière plus rentable. En d'autres termes, la connectivité réseau devrait être adaptée au contexte de l'abonné, le besoins réels de l'application et l'état du réseau. Dans les architectures des réseaux actuelles, la connectivité est pré-configurée pour avoir un comportement systématique. Elle est conçu pour activer les mêmes mécanismes du réseau dans tous les cas d'usage et ne peut pas s'adapter d'une manière dynamique, par exemple, à des contraintes conjoncturelles émergentes. Cette limitation est due à l'absence de deux éléments principaux, à savoir (i) la modularité, et (ii) la sensibilité au contexte.

- *modularité*: la connectivité dans le réseau d'accès doit être décomposé en un certain nombre des services du réseau. Chaque service requiert l'activation d'un nombre des mécanismes spécifiques. Par exemple, le service de joignabilité est l'un des services du réseau et est défini comme la capacité du fournisseur d'accès maintenir ses abonnés joignables. Quand ce service est activé, le fournisseur d'accès doit mettre à jour le chemin des paquets de données à chaque fois que l'utilisateur correspondant change son point d'accès. Par conséquent, la mise à jour de la localisation, la mise á jour des tunnels de mobilité et l'allocation d'une adresse IP permanente sont des mécanismes nécessaires pour assurer ce service de joignabilité.

- *la sensibilité au contexte*: la connectivité dans les réseaux d'accès doit prendre en considération toute information contextuelle et adapte son comportement en conséquence. Les informations contextuelles comprennent tout type de données qui peuvent être utiles pour améliorer et ajuster le comportement de la connectivité. Comme exemple d'informations contextuelles, nous citons le profil de l'utilisateur, le profile de mobilité de l'abonné (statique/nomade, une mobilité faible/élevée), type du terminal, les exigences de l'application, l'état du réseau, etc.

Les exemples suivants soulignent la nécessité de la modularité et la sensibilité aux contextes dans la gestion de la connectivité:

- Les terminaux de communication qui sont prévus pour être utilisés dans le futur ont des capacités variés. Ceci comprend non seulement les différents interfaces Radio, mais aussi les capacités de traitement et de stockage variés. En effet, le rapport annuels de prévisions de Cisco prévoit que les réseaux d'accès mobiles accueilleront 8,6 billions de dispositifs portables et 1,7 milliards de connexions Machine-to-Machine (M2M) (par exemple capteurs pour des applications médicales, des systèmes de suivi dans le transport maritime, les systèmes GPS dans les voitures, etc.) [Cis13]. Cependant, un même mécanisme du réseau ne convient pas tous les terminaux. Par conséquent, le réseau d'accès doit implémenter plusieurs mécanismes du réseau et la connectivité doit être capable à sélectionner le mécanisme adéquat selon le cas d'usage.

- Avec la variété des applications qui s'exécutent sur le terminal de l'abonné, les exigences en termes des services du réseau (par exemple la mobilité, le contrôle de la QoS, et la sécurité) diffèrent fortement. Par exemple, les caméra de surveillance statiques dans les rues envoient des séquences de vidéo périodiquement n'a pas besoin des services de mobilité. En outre, des sessions qui sont déjà sécurisées (par exemple sessions SSL établies entre les capteurs médicaux et la plateforme médicale, sessions VPN établies entre les ordinateurs portables des employés à distance et l'Intranet de leur entreprise, etc.) n'ont besoin ni de la confidentialité ni de la protection de l'intégrité au niveau du réseau d'accès. Ne pas assurer la sécurité de la session au niveau du réseau d'accès peut faire baisser la charge de traitement des paquets des données au niveau des équipements du réseau; conduisant ainsi à une utilisation des ressources plus efficaces.

Par conséquent, la modularité et la sensibilité aux informations contextuelles de la connectivité vaut bien à être étudié en plus de profondeur.

C.2 Objective et contributions

Notre objectif est de concevoir une architecture de réseau qui inter-connecte les accès hétérogènes et qui s'adapte automatiquement suivant les besoins réels de chaque cas d'usage. Cette adaptation comprend l'activation ou la désactivation des mécanismes du réseau tels que les mécanisme de la gestion de mobilité, de contrôle de la QoS, et de la sécurité. En d'autres termes, un mécanisme du réseau donné n'est activé que lorsque l'utilisateur et / ou le fournisseur d'accès a vraiment besoin de son présence. Cette architecture doit permettre une utilisation souple des ressources du réseau.

Toutefois, elle doit faciliter l'inter-fonctionnement entre les accès hétérogènes tout en évitant la redondance fonctionnelle, .

Le principal défi de ce travail est de concevoir un modèle de gestion de connectivité dans une architecture multi-accès qui est conscient des informations contextuelles lié à chaque cas d'usage et qui est capable d'adapter son comportement en conséquence. Par exemple, sachant que le flux de données est déjà sécurisé au niveau applicatif (par exemple les flux sécurisés avec le mécanisme IPsec), les mécanismes de chiffrement et de protection de l'intégrité au niveau du réseau d'accès sont considérés redondants et doivent être désactivés pour ce type de flux.

Cette thèse contient principalement trois contributions.

Dans la première contribution, nous analysons des approches actuelles dans le système 3GPP multi-accès. Le but d'une telle étude est d'analyser si les approches actuelles de gestion de la connectivité pourraient faire face aux défis imposés par le nouvel écosystème. Nous commençons notre étude par la décomposition de la connectivité en des services du réseau. En particulier, notre travail porte sur les services de mobilité et de sécurité. Grâce au différents scénarios des cas d'usage, nous avons montré que les mécanismes de mobilité et de sécurité dans les différents réseaux d'accès peuvent être contournés dans certains cas. À partir de ces scénarios, nous avons spécifié cinq exigences pour la gestion de la connectivité dans les architectures multi-accès.

Comme première exigence, la connectivité doit prendre en considération les informations contextuelles lors de son établissement et activer les services du réseau en conséquence. Dans la deuxième exigence, la connectivité doit être adaptative (c.-à-d. si une information contextuelle change au cours du temps, la connectivité doit adapter son comportement en fonction de ce changement). Une gestion de connectivité unifiée pour les accès hétérogènes représente la troisième exigence. Vu que les adaptations concernent la manipulation des équipements du réseau, la décision de l'adaptation nécessaire doit être prise par un équipement dans le réseau du fournisseur d'accès. Ceci représente la quatrième exigence. Toutefois, cette prise de décision peut être assister par le terminal de l'utilisateur. Enfin, comme cinquième exigence, la gestion de la connectivité doit utilisé les ressources du réseau d'une manière flexible pour atténuer les effets des situations de panne ou de surcharge des équipements du réseau.

Cette analyse qualitative du comportement de la connectivité dans les architectures actuelles est renforcée par une étude analytique qui consiste à évalué les coûts des mécanismes de la sécurité dans le réseau d'accès LTE/EPC. Ces études ont montré que le fournisseur d'accès LTE peut faire des économies importantes en termes de la charge de signalisation et de traitement au niveau des équipements, et de la surcharge liée à la transmission lorsque les mécanismes de la sécurité sont désactivés pour un nombre donné d'abonnés.

Dans la deuxième contribution, nous proposons un modèle de gestion de la connectivité (CACM) dédié pour les architectures multi-accès. Ce modèle respecte les exigences que nous avons établies auparavant. Il est capable de récupérer des informations contextuelles et de décider les adaptations nécessaire suivant ces informations. En plus, ce modèle est capable de gérer la connectivité dans les accès hétérogènes d'une manière unifié parce qu'il ne dépend pas d'une technologie d'accès spécifique. Ceci facilite l'inter-fonctionnement des accès hétérogènes. Il sélectionne et active les mécanismes du réseau en conformité avec les informations contextuelles assurant ainsi l'adaptabilité de la connectivité aux besoins des abonnés. Dans ce modèle, la prise de décision d'adapter le comportement de la connectivité se fait bel et bien du au niveau du réseau. En se basant sur des scénarios de cas d'usage, nous évaluons et comparons le modèle proposé avec le système 3GPP multi-accès.

Dans la troisième contribution, nous proposons deux applications concrètes du modèle proposé : (i) l'adaptation du service de protection du trafic des données et (ii) l'adaptation du service de continuité de session.

Pour la première application, nous proposons un mécanisme qui adapte le service de protection des données dans les accès non-3GPP suivant les besoins des flux applicatifs. Pour valider et tester la faisabilité de ce mécanisme, nous mettons en place un banc d'essai qui reproduit l'accès non-3GPP dans lequel nous avons implémenté un mécanisme qui active/désactive le chiffrement et de protection de l'intégrité dans le tunnel IPsec selon la nature du trafic (par exemple HTTP ou HTTPS).

Pour la deuxième application, nous avons proposé un nouveau plan de contrôle basé sur le protocole OpenFlow pour l'architecture LTE/EPC afin d'assurer un service de continuité de session adaptatif. Dans l'architecture LTE/EPC, le service de continuité de session se repose sur l'utilisation des tunnels de mobilité IP pour transférer les flux de données lorsque l'abonné change son point d'accès. Une situation de panne ou de surcharge temporaire peut impacté ces tunnels de mobilité. Le nouveau plan de contrôle proposé permet d'adapte le service de continuité de session non seulement aux exigences du flux applicatifs mais aussi à l'état du réseau. Par exemple, lorsque le passerelle SGW en cours d'utilisation est surchargée, les flux de données qui tolère les délais tels que des flux FTP peuvent être transférés temporairement vers un autre SGW. Cela permet de délester la passerelle SGW surchargée et continue à assurer la continuité de session pour tous les flux en cours. Bien évidemment, ce plan de contrôle peut évoluer pour assurer aussi un service de joignabilité adaptatif.

Bibliography

[3GP10a] 3GPP. Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks (Release 10). TS 24.302, 3rd Generation Partnership Project (3GPP), 2010.

[3GP10b] 3GPP. Generic Authentication Architecture (GAA); Generic Bootstrapping Architecture (GBA). TS 33.220, 3rd Generation Partnership Project (3GPP), 2010.

[3GP10c] 3GPP. GPRS Tunnelling Protocol for User Plane (GTPv1-U). TS 29.281, 3rd Generation Partnership Project (3GPP), 2010.

[3GP10d] 3GPP. Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO). TR 23.829, 3rd Generation Partnership Project (3GPP), 2010.

[3GP11a] 3GPP. 3GPP System Architecture Evolution (SAE); Security architecture. TS 33.401, 3rd Generation Partnership Project (3GPP), 2011.

[3GP11b] 3GPP. General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access. TS 23.401, 3rd Generation Partnership Project (3GPP), 2011.

[3GP11c] 3GPP. Mobility Management Entity (MME) and Serving GPRS Support Node (SGSN) related interfaces based on Diameter protocol. TS 29.272, 3rd Generation Partnership Project (3GPP), 2011.

[3GP11d] 3GPP. Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS). TS 24.301, 3rd Generation Partnership Project (3GPP), 2011.

[3GP12a] 3GPP. Architecture enhancements for non-3GPP accesses (Release 11). TS 23.402, 3rd Generation Partnership Project (3GPP), 2012.

[3GP12b] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC);Protocol specification. TS 36.331, 3rd Generation Partnership Project (3GPP), 2012.

[3GP12c] 3GPP. General Packet Radio Service (GPRS); Evolved GPRS Tunnelling Protocol (eGTP) for EPS. TS 29.274, 3rd Generation Partnership Project (3GPP), 2012.

[3GP12d] 3GPP. Network Domain Security (NDS); IP network layer security. TS 33.210, 3rd Generation Partnership Project (3GPP), 2012.

[3GP12e] 3GPP. Study of Evolved Packet Core (EPC) nodes restoration. TR 23.857, 3rd Generation Partnership Project (3GPP), December 2012.

[3GP12f] 3GPP. Technical Specification Group Core Network and Terminals; Restoration procedures. TS 23.007, 3rd Generation Partnership Project (3GPP), 2012.

[3GP13a] 3GPP. 3G Security; Wireless Local Area Network (WLAN) interworking security. TS 33.234, 3rd Generation Partnership Project (3GPP), 2013.

[3GP13b] 3GPP. General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface. TS 29.060, 3rd Generation Partnership Project (3GPP), 2013.

[3GP13c] 3GPP. Support of Short Message Service (SMS) over generic 3GPP Internet Protocol (IP) access. TS 23.204, 3rd Generation Partnership Project (3GPP), 2013.

[3GP13d] 3GPP. Support of SMS over IP networks. TS 24.341, 3rd Generation Partnership Project (3GPP), 2013.

[3GP 9] 3GPP. General Universal Mobile Telecommunications System (UMTS) architecture. TS 23.101, 3rd Generation Partnership Project (3GPP), Release 9.

[AADB10] Tansir Ahmed, Stephane Antoine, Song Dong, and Delphin Barankanira. Multi Access Data Network Connectivity and IP Flow Mobility in Evolved Packet System (EPS). In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC'10)*, pages 1–6. IEEE, 2010.

[AAOBL13] Hassan Ali Ahmad, Meryem Ouzzif, Philippe Bertin, and Xavier Lagrange. Distributed Mobility Management: Approaches and analysis. In *Proceedings of IEEE International Conference on Communications Workshops (ICC'13)*, pages 1297–1302, 2013.

[ABV+04] Bernard Aboba, Larry Blunk, John Vollbrecht, James Carlson, and Henrik Levkowetz. Extensible authentication protocol (EAP). Technical report, RFC 3748, 2004.

[AC10] Pierre Abi-Char. *A dynamic trust-based context-aware secure authentication framework for pervasive computing environments*. PhD thesis, university Pierre and Marie Curie Paris 6, 2010.

[AL12] Alcatel-Lucent. Alcatel-Lucent lightRadio Wi-Fi WLAN Gateway. Technical report, Alcatel-Lucent, 2012.

[AM05] Jalal F. Al-Muhtadi. *An Intelligent Authentication Infrastructure for Ubiquitous Computing Environments*. PhD thesis, University of Illinois at Urbana-Champaign, 2005.

[BAM10] Theophilus Benson, Aditya Akella, and David A Maltz. Network traffic characteristics of data centers in the wild. In *Proceedings of 10th ACM SIGCOMM conference on Internet measurement*, pages 267–280, 2010.

[BBB09a] Philippe Bertin, Servane Bonjour, and Jean-Marie Bonnin. Distributed or centralized mobility? In *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'09)*, pages 1–6, 2009.

[BBB09b] Philippe Bertin, Servane Bonjour, and Jean-Marie Bonnin. An evaluation of dynamic mobility anchoring. In *Proceedings of 70th IEEE Vehicular Technology Conference Fall (VTC'09-Fall)*, pages 1–5, 2009.

[BBF03] BBF. DSL Evolution - Architecture Requirements for the Support of QoS-Enabled IP Services. TR 059, Broadband Forum (BBF), 2003.

[BBF06] BBF. Migration to Ethernet-based DSL Aggregation. TR 101, Broadband Forum (BBF), 2006.

[BBF12a] BBF. Broadband Policy Control Framework (BPCF). TR , number = 134, Broadband Forum (BBF), 2012.

[BBF12b] BBF. Interworking Between Next Generation Fixed and 3GPP Wireless Networks. TR 203, Broadband Forum (BBF), 2012.

[BBH+10] Claudio Bettini, Oliver Brdiczka, Karen Henricksen, Jadwiga Indulska, Daniela Nicklas, Anand Ranganathan, and Daniele Riboni. A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, 6(2):161–180, 2010.

[BBHSCL12a] Aymen Belghith, Siwar Ben Hadj Said, Bernard Cousin, and Samer Lahoud. Collaboration Schemes Evaluation in Multi-domain Networks. *International Journal of Computer Science Issues (IJCSI)*, 9(4), 2012.

[BBHSCL12b] Aymen Belghith, Siwar Ben Hadj Said, Bernard Cousin, and Samer Lahoud. Proactive and reactive collaboration schemes for multi-domain networks monitoring. In *Proceedings of Computing, Communications and Applications Conference (ComComAp'12)*, pages 150–157, 2012.

[BBHSCL12c] Aymen Belghith, Siwar Ben Hadj Said, Bernard Cousin, and Samer Lahoud. QoS fault detection and localization mechanisms (FDLM) in multi-domain networks adapted to export methods. In *Proceedings of 14th International Conference on Advanced Communication Technology (ICACT'12)*, pages 848–853, 2012.

[BBO08] Halil Ibrahim Bulbul, Ihsan Batmaz, and Mesut Ozel. Wireless network security: Comparison of WEP (wired equivalent privacy) mechanism, WPA (wi-fi protected access) and RSN (robust security network) security protocols. In *Proceedings of 1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop*, page 9, 2008.

[BCG11] Paolo Bellavista, Antonio Corradi, and Carlo Giannelli. A Unifying Perspective on Context-Aware Evaluation and Management of Heterogeneous Wireless Connectivity. *IEEE Communications Surveys & Tutorials*, 13(3):337–357, 2011.

[BCLBHS11] Aymen Belghith, Bernard Cousin, Samer Lahoud, and Siwar Ben Hadj Said. Proceedings of proposal for the configuration of multi-domain network monitoring architecture. In *International Conference on Information Networking (ICOIN'11)*, pages 7–12, 2011.

[BHSGB12] Siwar Ben Hadj Said, Karine Guillouard, and Jean-Marie Bonnin. On the need for adaptive connectivity management in multi-access

architectures. In *Proceedings of Third International Conference on the Network of the Future (NoF'12)*, pages 1–5, 2012.

[BHSGB13a] Siwar Ben Hadj Said, Karine Guillouard, and Jean-Marie Bonnin. A comparative study on security implementation in EPS/LTE and WLAN/802.11. In Shafiullah Khan and Al-Sakib Khan Pathan, editors, *Wireless Networks and Security*, Signals and Communication Technology, pages 457–489. Springer Berlin Heidelberg, 2013.

[BHSGB13b] Siwar Ben Hadj Said, Karine Guillouard, and Jean-Marie Bonnin. On the benefit of context-awareness for security mechanisms in LTE/EPS networks. In IEEE, editor, *Proceedings of 24th IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC'13)*, pages 2414–2428, 2013.

[BHSGB13c] Siwar Ben Hadj Said, Karine Guillouard, and Jean-Marie Bonnin. Towards adaptive security mechanisms in 3GPP EPS/LTE networks. In IEEE, editor, *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC'13)*, pages 1876–1881, 2013.

[BHSSG+13] Siwar Ben Hadj Said, Malla Reddy Sama, Karine Guillouard, Lucian Suciu, Gwendal Simon, Xavier Lagrange, and Jean-Marie Bonnin. New control plane in 3GPP LTE/EPC architecture for on-demand connectivity service. In *Proceedings of second IEEE International Conference on Cloud Networking (CLOUDNET 2013)*, 2013.

[BSCL12] Aymen Belghith, Siwar Ben Hadj Said, Bernard Cousin, and Samer Lahoud. Export Methods in Fault Detection and Localization Mechanisms. *International Journal of Computer Science Issues (IJCSI)*, 9(4), 2012.

[BTSB11] Ahmed Bouabdallah, Franois Toutain, Michal Szczerbak, and Jean-Marie Bonnin. On the benefits of a network-centric implementation for context-aware telecom services. In *Proceedings of 15th International Conference on Intelligence in Next Generation Networks (ICIN'11)*, pages 236–240, 2011.

[BZ09] Erik-Oliver Blass and Martina Zitterbart. Tailored security and safety for pervasive computing. In *Proceedings of Open Research Problems in Network Security, IFIP Advances in Information and Communication Technology (INETSEC'09)*, volume 309, pages 85–92, 2009.

[BZR12] Carlos J Bernardos, JC Zunniga, and Alex Reznik. Towards flat and distributed mobility management: A 3GPP evolved network design. In *Proceedings of IEEE International Conference on Communications (ICC'12)*, pages 6855–6861, 2012.

[CDE+08] Murtaza Chiba, Gopal Dommety, Mark Eklund, David Mitton, and Bernard Aboba. Dynamic Authorization Extensions to Remote Authentication Dial In User Service (RADIUS). RFC 5176, Internet Engineering Task Force (IETF), 2008.

[CFP+07] Martin Casado, Michael J Freedman, Justin Pettit, Jianying Luo, Nick McKeown, and Scott Shenker. Ethane: Taking control of the enterprise. *ACM SIGCOMM Computer Communication Review*, 37(4):1–12, 2007.

[Cis10] Cisco. Cisco 3800 series integrated services routers. Technical report, Cisco, 2010.

[Cis13] Cisco. Ciscovisual networking index: Global mobile data traffic forecast update, 2012-2017. White paper, Cisco, 2013.

[CK+00] Guanling Chen, David Kotz, et al. A survey of context-aware mobile computing research. Technical report, Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College, 2000.

[CM] Cortex-M. http://www.arm.com/products/processors/cortex-m/cortex-m3.php.

[CMVW10] Marius Corici, Thomas Magedanz, Dragos Vingarzan, and Peter Weik. Enabling ambient aware service delivery in heterogeneous wireless environments. In *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'10)*, pages 1–6, 2010.

[Con] Floodlight Controller. http://www.projectfloodlight.org/floodlight/.

[CPRW03] David D. Clark, Craig Partridge, J. Christopher Ramming, and John T. Wroclawski. A knowledge plane for the internet. In *Proceedings of conference on Applications, technologies, architectures, and protocols for computer communications*, pages 3–10. ACM, 2003.

[Cro13] Peter Crocker. Converged-mobile-messaging analysis and forecasts. Technical report, 2013.

[CVVM13] Marius Corici, Dragos Vingarzan, Valentin Vlad, and Thomas Magedanz. Self-adaptable IP connectivity control in carrier grade mobile operator networks. In *Proceedings of Mobile Wireless Middleware, Operating Systems, and Applications*, pages 150–163. Springer, 2013.

[CYX+11] H. Anthony Chan, Hidetoshi Yokota, Jiang Xie, Pierrick Seite, and Dapeng Liu. Distributed and dynamic mobility management in mobile internet: Current approaches and issues. *Journal of Communications*, 6(1), 2011.

[D-l] D-link. http://www.dlink.com/us/en/home-solutions/connect/adapters/dwa-160-xtreme-n-dual-band-usb-adapter.

[DA00] Anind K. Dey and Gregory D. Abowd. The context toolkit: Aiding the development of context-aware applications. In *Proceedings of Workshop on Software Engineering for wearable and pervasive computing*, pages 431–441, 2000.

[Dey01] Anind K. Dey. Understanding and using context. *Personal and Ubiquitous Computing*, 5(1):4–7, 2001.

[DGP12] Trinh Minh Tri Do and Daniel Gatica-Perez. Contextual conditional models for smartphone-based human mobility prediction. In *Proceedings of ACM Conference on Ubiquitous Computing*, pages 163–172, 2012.

[Dis] Distributed Internet Traffic Generator (D-ITG). http://traffic.comics.unina.it/software/ITG/download.php.

[DlOBC+11] Antonio De la Oliva, Carlos Jesus Bernardos, Maria Calderon, Telemaco Melia, and Juan Carlos Zuniga. IP flow mobility: smart traffic offload for future wireless networks. *IEEE Communications Magazine*, 49(10):124–132, 2011.

[Dro97] Ralph Droms. Dynamic Host Configuration Protocol. RFC 2131, Internet Engineering Task Force (IETF), 1997.

[eEP] Harmonised eCall European Pilot. http://www.heero-pilot.eu/view/en/ecall.html.

[Eng] Product Engineering. http://campm.techmahindra.com/LTE/content/lte/ProductEngineering.pdf.

[Eri11] Ericsson. More than 50 billion connected devices. Technical report, 2011.

[Ero06] Pasi Eronen. IKEv2 Mobility and Multihoming Protocol (MOBIKE). RFC 4555, Internet Engineering Task Force (IETF), 2006.

[ETS98] ETSI. Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS. TR 101.112, 1998.

[ETS97] ETSI. Digital cellular telecommunications system (Phase 2+); General description of a GSM Public Land Mobile Network (PLMN). TS GSM 01.02, European Telecommunications Standards Institution (ETSI), R97.

[FALZ12] Victor Fajardo, Jari Arkko, John Loughney, and Glen Zorn. Diameter Base Protocol. RFC 6733, Internet Engineering Task Force (IETF), 2012.

[FKC+05] Hanane Fathi, Kazukuni Kobara, Shyam S. Chakraborty, H. Imai, and R. Prasad. On the impact of security on latency in WLAN 802.11b. In *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'05)*, volume 3, page 5, 2005.

[Fre] FreeRadius. http://freeradius.org/.

[GAB+09] Alex Galis, Henrik Abramowicz, Marcus Brunner, Danny Raz, Prosper Chemouil, Joe Butler, Costas Polychronopoulos, Stuart Clayman, Hermann de Meer, Thierry Coupaye, Aiko Pras, Krishan Sabnani, Philippe Massonet, and Syed Naqvi. Management and service-aware networking architectures (mana) for future internet - position paper: System functions, capabilities and requirements. In *Proceedings of Fourth International Conference on Communications and Networking in China (ChinaCOM 2009)*, pages 1–13, 2009.

[Ger09] Rainer Gerhards. The syslog protocol. RFC 5424, Internet Engineering Task Force (IETF), 2009.

[GJ03] Eva Gustafsson and Annika Jonsson. Always best connected. *IEEE Wireless Communications*, 10(1):49–55, 2003.

[GLD+08] Sri Gundavelli, Kent Leung, Vijay Devarapall, Kuntal Chowdhur, and Basavaraj Patil. Proxy Mobile IPv6. RFC 5213, Internet Engineering Task Force (IETF), 2008.

[Hag04] Creighton T. R. Hager. *Context Aware and Adaptive Security for Wireless Networks*. PhD thesis, Faculty of the Virginia Polytechnic Institute and State University, 2004.

[HCWzL09] Chan-Kyu Han, Hyoung-Kee Choi, Jung Woo zBaek, and Ho Woo Lee. Evaluation of authentication signaling loads in 3GPP LTE/SAE networks. In *Proceedings of the IEEE 34th Conference on Local Computer Networks (LCN 2009)*, pages 37–44, 2009.

[hos] hostapd. http://hostap.epitest.fi/hostapd/.

[IEE97] IEEE. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Specific Requirements. IEEE 802.11, 1997.

[IEE04] IEEE. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Medium Access Control (MAC) Security Enhancements. IEEE 802.11i, 2004.

[IOMS10] Masugi Inoue, Masaaki Ohnishi, Hiroaki Morino, and Tohru Sanefuji. A future access network architecture for providing personalized context-aware services with sensors. In XiaoJun Hei and Lawrence Cheung, editors, *Access Networks*, volume 37 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 121–132. Springer Berlin Heidelberg, 2010.

[Iri13] Srini Irigi. Service provider wi-fi: Architectures, use cases and deployments, 2013.

[JIFW06] H. Johnson, L. Isaksson, M. Fiedler, and S.F. Wu. A decision system for adequate authentication. In *Proceedings of the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies (ICN/ICONS/MCL 2006).*, pages 185–185, 2006.

[JLLJ13] Xin Jiny, Li-Erran Li, Vanbevery Laurent, and Rexford Jennifer. CellSDN: Software-Defined Cellular Core Networks. In *Open Networking Summit SDN Event*, 2013.

[JLVR13] Xin Jin, Li Erran Li, Laurent Vanbever, and Jennifer Rexford. SoftCell: scalable and flexible cellular core network architecture. In

Proceedings of the ninth ACM conference on Emerging networking experiments and technologies, pages 163–174, 2013.

[Kar11] Rich Karpinski. The LTE signaling challenge, 2011.

[Ken05a] Stephen Kent. IP Authentication Header. RFC 4302, Internet Engineering Task Force (IETF), 2005.

[Ken05b] Stephen Kent. IP Encapsulating Security Payload (ESP). RFC 4303, Internet Engineering Task Force (IETF), 2005.

[KJP+12] James Kempf, Bengt Johansson, Sten Pettersson, Harald Luning, and Tord Nilsson. Moving the mobile Evolved Packet Core to the cloud. In *Proceeding of the IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'12)*, 2012.

[KK06] Phongsak Keeratiwintakorn and Prashant Krishnamurthy. Energy efficient security services for limited wireless devices. In *Proceedings of the first International Symposium on Wireless Pervasive Computing*, pages 1–6, 2006.

[KMH12] Jinho Kim, Yasufumi Morioka, and Junichiro Hagiwara. An optimized seamless ip flow mobility management architecture for traffic offloading. In *Proceedings of the IEEE Network Operations and Management Symposium (NOMS'12)*, pages 229–236, 2012.

[Koo08] Rajeev Koodli. Mobile IPv6 Fast Handovers. RFC 5568, Internet Engineering Task Force (IETF), 2008.

[KS05] Stephen Kent and Karen Seo. Security Architecture for the Internet Protocol. RFC 4301, Internet Engineering Task Force (IETF), 2005.

[KTS10] Andreas Kunz, Tarik Taleb, and Stefan Schmid. On minimizing serving GW/MME relocations in LTE. In *Proceedings of the 6th ACM International Wireless Communications and Mobile Computing Conference*, pages 960–965, 2010.

[KUb] KUbuntu. http://www.kubuntu.org/.

[LEC10] Jong-Hyouk Lee, Thierry Ernst, and Tai-Myoung Chung. Cost analysis of IP mobility management protocols for consumer mobile devices. *IEEE Transactions on Consumer Electronics*, 56(2):1010–1017, 2010.

[LNPK05] John Loughney, Madjid F. Nakhjiri, Charles E. Perkins, and Rajeev Koodli. Context Transfer Protocol (CXTP). RFC 4067, Internet Engineering Task Force (IETF), 2005.

[MAB+08] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. Openflow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review*, 38(2):69–74, 2008.

[Mal] Jouni Malinen. http://hostap.epitest.fi/wpa_supplicant/.

[MCN] Mobile Cloud Networking MCN. http://www.mobile-cloud-networking.eu/site/.

[MEV13] MEVICO. D2.2 Architectural EPC extensions for supporting heterogeneous mobility schemes. Technical report, CELTIC, 2013.

[mfeOWDnC] Connectivity management for eneRgy Optimized Wireless Dense networks (CROWD). http://www.ict-crowd.eu/.

[Mob12] Mobile Networks Evolution for Individual Communications Experience - Mevico. Architecture design release 3. Technical report, CELTIC, 2012.

[MP08] Christian Makaya and Samuel Pierre. An analytical framework for performance evaluation of ipv6-based mobility management protocols. *IEEE Transactions on Wireless Communications*, 7(3):972–983, 2008.

[MSP07] Mamoru Mitsuishi, Naohiko Sugita, and Phongsaen Pitakwatchara. Force-feedback augmentation modes in the laparoscopic minimally invasive telesurgical system. *IEEE/ASME Transactions on Mechatronics*, 12(4):447–454, 2007.

[MyS] MySQL. http://dev.mysql.com/.

[ndp] ndppd. http://priv.nu/projects/ndppd/.

[Net12] NetworkWorld. What is software defined networking (SDN)? http://www.networkworld.com/news/2012/082912-insider-sdn-262010.html, 2012.

[NM13] David Nowoswiat and Gordon Milliken. Managing LTE core network signaling traffic, 2013.

[NVAGD08] Quoc-Thinh Nguyen-Vuong, Nazim Agoulmine, and Yacine Ghamri-Doudane. A user-centric and context-aware solution to interface management and access network selection in heterogeneous wireless environments. *Computer Networks*, 52(18):3358–3372, 2008.

[NXS10] Christoforos Ntantogian, Christos Xenakis, and Ioannis Stavrakakis. A generic mechanism for efficient authentication in B3G networks. *Elsevier Science Computers & Security*, 29(4):460–475, 2010.

[ONF12a] Open Networking Foundation ONF. Software-defined networking: The new norm for networks. White paper, Open Networking Foundation (ONF), 2012.

[ONF12b] Open Networking Foundation. OpenFlow Switch Specification, version 1.3.1, September 6 2012.

[Ope] OpenFlow. http://www.openflow.org/.

[OS12] Barbara Orlandi and Frank Scahill. WiFi Roaming - Building on ANDSF and HOTSPOT 2.0. Technical report, Alcatel-Lucent, 2012.

[OWZ10] Yoshihiro Ohba, Qin Wu, and Glen Zorn. Extensible Authentication Protocol (EAP) Early Authentication Problem Statement. RFC 5836, Internet Engineering Task Force (IETF), 2010.

[Per10] Charles E. Perkins. IP Mobility Support for IPv4, Revised. RFC 5944, Internet Engineering Task Force (IETF), 2010.

[PMJ13] Mark Page, Maria Molina, and Gordon Jones. GSMA: The mobile economy 2013, 2013.

[Pop12] Dana Popovici. *Gestion du contexte pour des applications mobiles dédiées aux transports*. PhD thesis, Université de Valenciennes et du Hainaut-Cambresis, 2012.

[PSBD11] Utpal Paul, Anand Prabhu Subramanian, Milind M. Buddhikot, and Samir R. Das. Understanding traffic dynamics in cellular data networks. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM'2011)*, 2011.

[rad] radvd. http://www.litech.org/radvd/.

[RAJC11] Jarno Rajahalme, Shane Amante, Sheng Jiang, and Brian Carpenter. IPv6 flow label specification. RFC 6437, Internet Engineering Task Force (IETF), 2011.

[RWRS00] Carl Rigney, Steve Willens, Allan C. Rubens, and William A. Simpson. Remote Authentication Dial In User Service (RADIUS). RFC 2865, Internet Engineering Task Force (IETF), 2000.

[SBHSGS13] Malla Reddy Sama, Siwar Ben Hadj Said, Karine Guillouard, and Lucian Suciu. New control plane in 3GPP LTE/EPC architecture for more flexible and reliable connectivity, 2013.

[SBHSGS14] Malla Reddy Sama, Siwar Ben Hadj Said, Karine Guillouard, and Lucian Suciu. Enabling network programmability in LTE/EPC architecture using OpenFlow. In *Proceedings in the 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt'14)*, 2014.

[SBL11] Izuru Sato, Ahmed Bouabdallah, and Xavier Lagrange. Improving lte/epc signaling for sporadic data with a control-plane based transmission procedure. In *Proceedings of the 14th International Symposium on Wireless Personal Multimedia Communications (WPMC'11)*, pages 1–5, 2011.

[Sca] Scalable and Adaptive Internet Solutions (SAIL). http://www.sail-project.eu/.

[SDK+06] Libo Song, Udayan Deshpande, Ulas C Kozat, David Kotz, and Ravi Jain. Predictability of WLAN mobility and its effects on bandwidth provisioning. In *Proceedings of the International Conference on Computer Communications (INFOCOM'06)*, 2006.

[Sha11] Clifford A. Shaffer. *Data Structures and Algorithm Analysis (C++ Version)*. 3.2 edition, 2011.

[Shi07] Robert W. Shirey. Internet Security Glossary, Version 2. RFC 4949, Internet Engineering Task Force (IETF), 2007.

[SK05] Nancy Samaan and Ahmed Karmouch. A mobility prediction architecture based on contextual knowledge and spatial conceptual maps. *IEEE Transactions on Mobile Computing*, 4(6):537–551, 2005.

[SM06] Muhammad Sher and Thomas Magedanz. Secure access to ip multimedia services using generic bootstrapping architecture (gba) for 3g & beyond mobile networks. In *Proceedings of the 2nd ACM international workshop on Quality of service & security for wireless and mobile networks*, pages 17–24, 2006.

[SNK12] Myung-Ki Shin, Ki-Hyuk Nam, and Hyoung-Jun Kim. Software-defined networking (SDN): A reference architecture and open APIs. In *Proceedings of the International Conference on ICT Convergence (ICTC'12)*, pages 360–361, 2012.

[Sno] Snort. http://www.snort.org/.

[Spe11] Speed cameras: how they work and what effect they have. https://www.swov.nl/rapport/Factsheets/UK/FS_Speed_cameras.pdf, 2011.

[SQBB10] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[SSGC05] Craig Shue, Youngsang Shin, Minaxi Gupta, and Jong Youl Choi. Analysis of IPSec overheads for VPN servers. In *Proceedings of the first IEEE ICNP Workshop on Secure Network Protocols (NPSec'05)*, pages 25–30. IEEE, 2005.

[Str] StrongSwan. http://www.strongswan.org/.

[SW00] Henning Schulzrinne and Elin Wedlund. Application-layer mobility using SIP. In *IEEE Service Portability and Virtual Customer Environments*, pages 29–36, 2000.

[Swi] Open Virtual Switch. http://openvswitch.org/.

[Sysa] Syslog-ng. http://doc.ubuntu-fr.org/syslog-ng.

[Sysb] Evolving Systems. http://www.evolving.com/pdfs/Ovum_Driving_Efficiencies_in_SIM_Card_Provisioning_WP_Jun10.pdf.

[TCKS09] Arsalan Tavakoli, Martin Casado, Teemu Koponen, and Scott Shenker. Applying NOX to the Datacenter. In *Proceedings of the SIGCOMM Workshop on Hot Topics in Networks (HotNets'09)*, 2009.

[Tel11] Tellabs. Tellabs "End of profit" study executive summary, January 2011.

[TK12] Tarik Taleb and Andreas Kunz. Machine type communications in 3gpp networks: potential, challenges, and solutions. *IEEE Communications Magazine*, 50(3):178–184, 2012.

[TL11] Peyman TalebiFard and Victor CM Leung. Context-aware mobility management in heterogeneous network environments. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 2(2):19–32, 2011.

[TNO10] TNO Knowledge for business. Mobile network improvements for machine type communications. http://docbox.etsi.org/workshop/2010/201010_m2mworkshop/06_m2mglobalcollaboration/norp_tno_mobilentwimprovements.pdf, 2010.

[TRKN09] Ionut Trestian, Supranamaya Ranjan, Aleksandar Kuzmanovic, and Antonio Nucci. Measuring serendipity: connecting people, locations and interests in a mobile 3G network. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet measurement*, IMC'09, New York, USA, 2009. ACM.

[TS11] Tarek Taleb and Konstantinos Samdanis. Ensuring service resilience in the EPS: MME failure restoration case. In *IEEE Global Communications Conference (GLOBECOM'11)*, pages 1–5, 2011.

[TSF12] Tarik Taleb, Konstantinos Samdanis, and Fethi Filali. Towards supporting highly mobile nodes in decentralized mobile operator networks. In *IEEE International Conference on Communications (ICC'12)*, pages 5398–5402, 2012.

[TSS11] Tarik Taleb, Konstantinos Samdanis, and Stefan Schmid. DNS-based solution for operator control of selected ip traffic offload. In *IEEE International Conference on Communications (ICC'11)*, pages 1–5, 2011.

[TST10] Dario S Tonesi, Luca Salgarelli, and Alessandro Tortelli. Securing the signaling plane in beyond 3g networks: analysis of performance overheads. *Security and Communication Networks*, 3(2-3):217–232, 2010.

[TTP06] Dave Thaler, Mohit Talwar, and Chirayu Patel. Neighbor Discovery Proxies (ND Proxy). RFC 4389, Internet Engineering Task Force (IETF), 2006.

[Ubu] Ubuntu. http://www.ubuntu-fr.org/.

[UFFGPC+11] A Urbano Fullana, Josep Lluís Ferrer Gomila, Magdalena Payeras Capella, M Hinarejos Campos, and L Huguet Rotger. Cross-Layer Secrecy Design on TCP/IP and 802.11 for Energy Saving. In

Proceedings of the 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS'11), pages 1–5. IEEE, 2011.

[Uni06] International Telecommunication Union. ITU-T Recommandation Q.1706/Y.2801, Mobility Management requirements for NGN, 11/2006. Technical report, ITU-T, 2006.

[VCP04] Pablo Vidales, Rajiv Chakravorty, and Calicrates Policroniades. Proton: a policy-based solution for future 4g devices. In *Proceedings of the 5th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'04)*, pages 219–222, 2004.

[Vir] Oracle VM VirtualBox. https://www.virtualbox.org/.

[WBLR09] Indra Widjaja, Peter Bosch, and Humberto La Roche. Comparison of mme signaling loads for long-term-evolution architectures. In *Proceedings of the 70th IEEE Vehicular Technology Conference Fall (VTC'09-Fall)*, pages 1–5, 2009.

[XLH+09] Cheng Xue, Jijun Luo, Ruediger Halfmann, Egon Schulz, and Christian Hartmann. Inter GW load balancing for next generation mobile networks with flat architecture. In *Proceedings of the 69th IEEE Vehicular Technology Conference (VTC'09)*, pages 1–5, April 2009.

[XLMS06] Christos Xenakis, Nikolaos Laoutaris, Lazaros Merakos, and Ioannis Stavrakakis. A generic characterization of the overheads imposed by ipsec and associated cryptographic algorithms. *Computer Networks*, 50(17):3225–3241, 2006.

[XUb] XUbuntu. http://xubuntu.org/.

[YPMM12] Suleiman Y Yerima, Gerard P Parr, Sally I McClean, and Philip J Morrow. Adaptive measurement-based policy-driven QoS management with fuzzy-rule-based resource allocation. *Future Internet*, 4(3):646–671, 2012.

[YPP+10] Suleiman Y. Yerima, Gerard P. Parr, Cathryn Peoples, Sally McClean, and Philip J. Morrow. A framework for context-driven end-to-end QoS control in converged networks. In *Proceeding of the International Conference on Network and Service Management (CNSM'10)*, pages 250–253, 2010.

[YPSM11] Suleiman Y. Yerima, Gerard P. Parr, McClean Sally, and Philip J. Morrow. Modelling and evaluation of a policy-based resource management framework for converged next generation networks. In *Proceedings of the 12th IFIP/IEEE International Symposium on Integrated Network Management (IM'11)*, 2011.

[ZJZ10] Mariem Zekri, Badii Jouaber, and Djamal Zeghlache. Context aware vertical handover decision making in heterogeneous wireless networks. In *Proceedings of the 35th IEEE Conference on Local Computer Networks (LCN'10)*, pages 764–768. IEEE, 2010.

Résumé

La gestion de la connectivité réseau dans les architectures multi-accès devient une question primordiale parce que ces architectures devraient être en mesure de faire inter-fonctionner des technologies hétérogènes et de faire face aux défis imposés par le nouvel écosystème.

L'organisme de normalisation 3GPP a proposé un système 3GPP multi-accès qui vise à fournir une connectivité réseau omniprésente. Cette proposition a indubitablement des avantages, mais apporte également plusieurs défis aux opérateurs de réseau. En effet, dans ce système, les mécanismes de réseau, telles que les mécanismes de gestion de la mobilité et de sécurité sont conçus pour être activés d'une manière systématique augmentant ainsi le coût d'exploitation du réseau. Par exemple, le mécanisme de la mise à jour de la localisation est activé en permanence, même lorsque les terminaux ne sont pas mobiles. Cependant, les opérateurs de réseaux doivent relever le défi d'accueillir plusieurs catégories d'abonnés tels que des abonnés non-mobiles ou à forte mobilité, des abonnés qui demande un niveau de sécurité élevé ou faible, des abonnés demandant une large/faible bande passante, etc.

Cette thèse contient principalement trois contributions. Dans la première contribution, nous analysons des approches actuelles dans le système 3GPP multi-accès. Le but d'une telle étude est d'analyser si les approches actuelles de gestion de la connectivité réseau pourraient faire face aux défis imposés par le nouvel écosystème. Pour faire ceci, nous décomposons de la connectivité réseau en services. En particulier, notre travail porte sur les services de mobilité et de sécurité. Grâce à des scénarios d'usage du réseau, nous montrons que les mécanismes de mobilité et de sécurité dans le réseau d'accès peuvent être contournés dans certains cas. En outre, nous spécifions un certain nombre d'exigences pour la gestion de la connectivité dans les architectures de future.

Dans la deuxième contribution, nous proposons un module de gestion de la connectivité (CACM) pour les architectures multi-accès. Ce module est sensible au contexte des abonnés. Il est capable de gérer et de faire inter-fonctionner les accès hétérogènes d'une manière efficace. Il sélectionne et active les mécanismes du réseau en conformité avec les informations contextuelles.

Dans la troisième contribution, nous proposons deux exemples d'applications du modèle proposé : (i) l'adaptation du service de protection du trafic des donnés et (ii) l'adaptation du service de continuité de session. Pour la première, nous mettons en place un banc d'essai qui reproduit l'accès non-3GPP non sécurisé dans lequel nous avons implémenté un mécanisme qui active/désactive les mécanismes de chiffrement et de protection de l'intégrité dans le tunnel IPsec selon les exigences des flux applicatifs. Pour la deuxième application, nous avons proposé un plan de contrôle basé sur le protocole OpenFlow pour l'architecture LTE/EPC afin d'assurer un service de continuité de session adaptatif.

Mots-clés : gestion de la mobilité, sécurité, réseaux adaptatifs, convergence fixe et mobile, gestion de la connectivité réseau

Abstract

Managing the network connectivity in multi-access architectures becomes a critical issue as these architectures should be able to interwork between heterogeneous technologies and to face the new ecosystem challenges. The 3GPP standards body proposed the multi-access 3GPP system that aims at providing ubiquitous network connectivity. This proposal has many benefits but it brings also a lot of challenges for network operators. In fact, within this system, network mechanisms such as mobility management and security mechanisms are designed to be activated in a systematic manner leading to rising network operating costs. For instance, the location update mechanism is always performed even for static devices. However, network operators face the challenge to host several categories of subscribers such as static subscribers, subscribers with high mobility, subscribers requiring high security level, subscribers satisfied with just low security level, subscribers requiring high/low bandwidth, etc.

This thesis includes three main contributions. In the first contribution, we analyze the current approaches in multi-access 3GPP system. The aim of such a study is to analyze whether the current network connectivity management approaches could face the challenges imposed by the new ecosystem. To do that, we decompose the network connectivity into several network services. Particularly, our work addresses the mobility and security services. Through network usage scenarios, we showed that security and mobility mechanisms in the access network can be bypassed in certain cases. In addition, we specified a number of requirements for the connectivity management I future network architectures.

In the second contribution, we concentrate on proposing a Context-Aware Connectivity Management (CACM) module for multi-access architectures. This module is able to manage and interwork between heterogeneous accesses technologies in an efficient manner. It selects and activates network mechanisms in accordance with the contextual information.

In the third contribution, we propose two concrete applications of the proposed model: (i) adaptive data traffic protection service and (ii) adaptive session continuity service. For the first, we setup a test bed that reproduces the untrusted non-3GPP access in which we implemented a mechanism that activates/deactivates the encryption and integrity protection mechanisms within the IPsec tunnel according to the flow requirements. For the second application, we proposed an OpenFlow-based control plane for the LTE/EPC architecture to ensure adaptive session continuity service.

Keywords : mobility management, security, adaptive networks, fixed and mobile convergence, network connectivity management

TELECOM Bretagne

n° d'ordre : 2014telb0341
Télécom Bretagne
Technopôle Brest-Iroise - CS 83818 - 29238 Brest Cedex 3
Tél : + 33(0) 29 00 11 11 - Fax : + 33(0) 29 00 10 00