Service continuity in heterogeneous wireless networks for time constrained applications

Mohammed Boutabia

Ecole Doctorale EDITE

Thèse présentée pour l'obtention du diplôme de Docteur de Télécom & Management SudParis

Doctorat conjoint TMSP-UPMC

Spécialité : informatique et réseaux

Par

Mohammed BOUTABIA

Titre

Continuité de service dans les réseaux sans fil hétérogènes pour les applications à contrainte de temps

Soutenue le 08 avril 2011 devant le jury composé de :

André-Luc BEYLOT	**ENSEEIHT, Toulouse**	Rapporteur
Dominique GAITI	**UTT, Troyes**	Rapporteur
Guy PUJOLLE	**LIP6, Paris**	Examinateur
Laurent TOUTAIN	**Telecom Bretagne**	Examinateur
Pascal LORENZ	**IUT, Colmar**	Examinateur
Marion BERBINEAU	**IFSTTAR, Lille**	Examinateur
Hossam AFIFI	**Telecom SudParis**	Directeur de thèse

Thèse n° 2011TELE0011

Abstract

Service continuity is an important component in mobile communications. With the coexistence of different access network technologies and the emergence of multi-interface mobile devices, service providers should maintain the ongoing communication when the mobile travels among heterogeneous networks. Services like IPTV, video on demand or voice over IP are widely proposed by operators for which service continuity should be guaranteed. This thesis is devoted to service continuity of real-time applications in heterogeneous networks. We tackle this problem from two perspectives: session mobility and terminal mobility. Although these two mechanisms have the same purpose which is ensuring service continuity when changing the terminal or the access network, each technique has its own challenges and constraints.

As far as session mobility is concerned, a new signaling protocol has been proposed to transfer the session between different terminals. This protocol has been implemented in video streaming scenarios and evaluated in a testbed. Moreover, we address the problem of media adaptation, especially renegotiation of QoS parameters since session might be transferred to a new terminal with different capabilities than the original one. QoS renegotiation can be extended to cover the case where some internal parameters are degraded during the session in the same terminal.

For terminal mobility, we propose a new handover mechanism using IEEE802.21 with Fast handover for Mobile IPv6. The purpose of this proposal is to reduce the handover delay and the dedicated buffer in access routers. In addition, an optimization is proposed for Fast handovers for Mobile IPv6 in order to maximize the probability of its predictive mode. In the same context, mobility in IMS is considered and an appropriate solution is proposed to answer IMS requirements. Finally, we conduct a comparison study between different mobile IP variants in the case of vertical handover. Based on this comparison, we give some guidelines that should help in choosing the most efficient protocol following specific parameters. The proposed solutions and studies have been evaluated analytically or/and using a simulation tool.

Résumé

La continuité de service est un élément important dans les communications mobiles. Avec la coexistence de différentes technologies d'accès au réseau et l'émergence de dispositifs mobiles avec plusieurs interfaces réseau, les fournisseurs de services doivent maintenir la communication en cours lorsque les mobiles voyagent entre des réseaux hétérogènes. Des services comme l'IPTV, vidéo à la demande ou la voix sur IP sont largement proposés par les opérateurs pour lesquels la continuité de service doit être garantie. Cette thèse est consacrée à la continuité du service pour des applications temps réel dans des réseaux hétérogènes. Nous abordons ce problème de deux perspectives: la mobilité de session et la mobilité de terminal. Bien que ces deux mécanismes aient le même but qui est d'assurer la continuité de service lors du changement de terminal ou de réseau d'accès, chaque technique a ses propres défis et contraintes.

En ce qui concerne la mobilité de session, un nouveau protocole de signalisation a été proposé pour transférer la session entre les différents terminaux d'un utilisateur. Ce protocole a été conçu pour les scénarios de streaming vidéo. Son implémentation a permis la validation du protocole proposé ainsi que son évaluation. En outre, nous traitons le problème de l'adaptation des flux multimédias, notamment la renégociation des paramètres de la qualité de service puisque la session pourrait être transféré à un nouveau terminal avec des capacités différentes que le terminal d'origine. Cette renégociation peut être étendue pour couvrir le cas où certains paramètres internes sont dégradés au cours de la session dans le même terminal.

Quand à la mobilité de terminal, nous proposons un mécanisme basé sur l'utilisation de la nouvelle norme IEEE802.21 et du protocole de mobilité FMIPv6. Le but de cette proposition est de réduire le délai du handover et la taille de l'espace mémoire dédiée au niveau des routeurs d'accès. En outre, une optimisation est proposée pour FMIPv6 afin de maximiser la probabilité de son mode prédictive. Dans le même contexte, la mobilité dans l'IP Multimédia Subsystem (IMS) est considéré et une solution adaptée est proposée pour répondre aux exigences de l'IMS. Enfin, nous

menons une étude comparative entre les différentes variantes de Mobile IP dans le cas de handover vertical. En se basant sur cette comparaison, nous donnons quelques directives qui devraient aider à choisir le protocole le plus efficace suivant des paramètres spécifiques. Les solutions proposées et les études ont été évaluées avec des méthodes analytiques et/ou en faisant appel à des simulations.

Acknowledgments

First of all I want to express my profound gratitude to my supervisor, Pr. Hossam AFIFI for his valuable advices, for his guidance and encouragement during my thesis.

I am grateful to the members of my thesis committee, Professors Dominique GAITI, André-Luc BEYLOT, Guy PUJOLLE, Pascal LORENZ, Laurent TOUTAIN and Marion BERBINEAU for their time and effort to evaluate this work.

I am also thankful to the head of RS2M department Pr. Djamal Zeghlache for his collaboration in funding this thesis.

A special thank to Dr. Luis Rojas CARDENAS with whom I worked in SUMO project.

Many thanks go to my friends of mobility and security team especially Abid, Emad, Teck, Aroua, Chedly, Ahmed and Amira for their support and friendship.

Finally this thesis would not have been possible without the unconditional support and encouragement of my father and the endless love of my mother.

To my father,

To my mother,

To my sisters and brother

Table of Content

List of Figures

List of Tables

INTRODUCTION

Mobile and wireless communication networks have known a tremendous progress and expansion in the last few years. The fourth generation telecommunication system intends to provide a broadband wireless access to users anytime and anywhere. 4G users have the possibility to use different access network technologies from wide range to wireless local area networks (WLAN) like WIFI. The low cost of deployment and exploitation of WLANs makes them very attractive to service providers and customers. Currently France is among the countries that have a big number of deployed public WIFI access points with more than 30000 access points in 2010 [1]. Therefore, it occupies the third place after USA and china. This number does not take into account the shared connection done over the box offered by certain Internet Service Provider to their customers to enjoy internet connection when they are away from home (ex: “Free WiFi”, “Neuf WiFi”). The total number of public WiFi access points deployed in the world reaches 310000 in 2010 according to the same study, with a growth rate of 20%. In parallel to network development, mobile devices too know a complete transformation. Mobile phones, personal data assistant (PDA), Internet tablets, laptops…etc, acquire more and more hardware capabilities in terms of processing speed, memory space, communication interfaces and storage space. These capabilities allow mobile devices not only to communicate through different network technologies, but also to choose the most convenient one in case of several available networks; this latter characteristic is known as Always Best Connected (ABC). This means that at any time the mobile should be connected to the best available network. “Best” can refer to many criterions such as cost, bit rate, user preferences…etc. The mobile should decide which network will meet requirements of its applications at a given moment. Moreover, the user can even choose the right device to use depending on his situation. Transferring the current session between different terminals gives o high degree of liberty to the user and realizes a real service ubiquity. Nevertheless, the coexistence of multiple access network technologies raises the problem of interworking and mobility management

across them. Fortunately, the wide use of IP in all access and core networks makes it possible for the users to roam between different access networks while enjoying their services. However, the growing of multimedia and real-time applications usage by internet customers imposes additional constraints to mobility management protocols. Such kind of application requires a great attention to maintain the same level of quality of service. For example, in real-time applications, handover delay should be kept as short as possible to guarantee seamless service continuity.

1.1.Problem Statement

Mobility management in IP networks is gaining more and more interest from research community and service providers. The reason for this interest is the emergence of new advanced technologies in both access networks and mobile devices. From the one hand, wireless access networks are in a constant progress in terms of offered bandwidth and quality of service provisioning. On the other hand, mobile devices have known a tremendous development in terms of hardware and software capabilities which allow them to perform more complicated tasks rather than just making a phone call as the early invented mobile phones. Nevertheless, the new applications are no more based on circuit switched networks since IP has proved to be a convincing protocol for interworking between different networks and hence, adopted by the community as the protocol of infrastructure convergence and service integration. Therefore, many service providers converted there core networks to be operable on IP, and the new services are IP based as well. Voice over IP (VoIP) and IPTV are the most successful IP based services provided by many operators as part of their quadruple play service (i.e. internet, telephony, TV and mobility). Nonetheless, mobility here is limited to the access technology that the operator has chosen to carry the service on. In other words mobility is managed in a very controlled way. This restriction tightens the liberty of the user in choosing the access network he/she prefers and excludes the possibility of coexistence with other technologies even belonging to the same operator. For example, a user who is using his mobile phone to watch a TV program using 3G network would prefer to take advantage of a WIFI hot spot connection when in airport waiting room to enjoy a better quality with lower cost. Roaming between heterogeneous networks without any

interruption in the current session is a challenging task; especially with the real time character of the ongoing communication. Such task is complicated because of two reasons: i) changing access interface requires an intervention from the user since the network can only ensure horizontal handovers if supported. ii) The second reason is the inner problem of IP address function duality; the connection shall break if the IP address changes during the session. Moreover, new services should be created for session mobility support. In other words, in order to make session transfer between different terminals, new mechanisms should be defined and supported by both terminal and network.

1.2.Motivation

Continuity of service is the most important challenge that operator should face in order to offer ubiquitous service. If network heterogeneity is beneficial from user point of view, it complicates more the task for the service provider. On the one hand, solving service continuity problem will allow the operators to diversify their access network and take advantage of low cost infrastructure while maintaining the same level of QoS. On the other hand they will grant more flexibility for users to choose their favorite network or even transport the current session to a different terminal with better hardware capabilities.

The need of supporting service continuity is particularly important in applications that have certain continuity in time. For example web browsing is less stringent to service continuity since it is a discontinuous application: the user requests a web page and waits for the answer then starts reading the displayed information. At the opposite, when the application takes place for a while such in case of file download or video streaming the continuity of the service is obligatory otherwise the service will stop. The way mobility is supported is again related to the application it self. The user will not be affected by a high handover delay in case of file download since the result is not perceptible until completion of the download. On the contrary, any excessive delay in achieving the required mobility operation will affect the quality of delay sensitive applications like video streaming. Although many works have been conducted in the area of mobility, still problems are not completely solved.

1.3.Contributions

In this thesis we address the problem of service continuity in heterogeneous networks from two perspectives: session mobility and terminal mobility. Although these mobility services have the same purpose which is assuring service continuity when changing the terminal or the access network, but each technique has its own challenges and constraints.

1.3.1. Session Mobility

A new signaling protocol has been proposed for session mobility support between different terminals. The protocol has been implemented for video streaming scenario and evaluated in a testbed. In this context, we address also the problem of media adaptation especially renegotiation of QoS parameters since session might be transferred to a new terminal with different capabilities than the original one. QoS renegotiation can take place in either session transfer or during the session in the same terminal when change in some internal parameters occurs. Change in such parameters affects the quality of experience of the user if no measures are taken to adapt the media accordingly.

1.3.2. Terminal Mobility

Contribution 1:

After studying the state of the art related to mobility techniques and protocols, we made a classification of terminal mobility protocols over IP in terms of operation steps. These steps are summed up in four major sub-operations that are not necessarily present in all mobility protocols. This classification is particularly interesting in the analysis and diagnosis of any mobility protocol and help in designing new protocols.

Contribution 2:

We propose a new handover mechanism using the new standard IEEE802.21 together with Fast handover for mobile IPv6. This scheme is different from the classical paradigms which are mobile initiated and network initiated handovers. Actually we put the mobile and the network in collaboration relationship and the result is a new paradigm: mobile initiated-network terminated handover.

Contribution 3:

We tackle the operation mode of FMIPv6 in order to maximize the probability of its predictive mode. This contribution can be considered as a continuation of the previous one. When the previous contribution is more based on layer 2 mechanisms to improve the handover performance, this one bring enhancement to the mobility protocol itself to maximize the probability of a successful proactive handover

Contribution 4:

In the context of IMS we propose a new hybrid mobility management protocol based on both network layer and application layer mobility protocols. This approach avoids redundancy of functionality and network entities. In addition it shows better results compared to the classical approaches. Actually this is one of the direct results of the mobility analysis in IMS under the methodology made in contribution 1. Therefore we bring the missing part in mobility operation without making any redundancy.

Contribution 5:

We conduct a comparison study of mobile IP variants in a vertical handover scenario. This work came from the fact that new protocols claim improving performance of handover. If this is true for horizontal handovers it is not always true in case of vertical handover. Through this study we show that multi interface users can perform seamless handovers with classical mobility protocols better than sophisticated ones.

1.4.Structure of The Thesis

Chapter 2: this chapter is devoted to the state of the art of both access network technologies and mobility mechanisms. It gives an overview of nowadays wireless access networks belonging to the different standard bodies. Afterwards, a number of mobility mechanisms and protocols are investigated. They are classified to layer 2 mechanisms, layer 3 mechanisms and upper layer mechanisms.

Chapter 3: in this chapter we tackle session mobility issue. This chapter is divided in two parts. The first part presents the proposed session mobility protocol “SESSAMO”. This lightweight peer to peer protocol is designed for session mobility in video streaming. Performance evaluation of the new signalling protocol is conducted

through a testbed. The second part treats the problem resulting from the mechanism of session transfer which is media adaptation. A mechanism of renegotiating the new QoS parameters depending on the capabilities of each terminal is proposed. It is based on SDPng and MPEG21 standards. In the same perspective we extend the use of renegotiation mechanism to adapt the media within the same terminal when a change of capabilities occurs during the session.

Chapter 4: after an overview of mobility protocols we draw some conclusions from the way most of IP mobility protocols work. Actually, any mobility management protocol operating in network layer, transport layer or application layer follows more or less the same sub-operations to achieve seamless transitions. In this chapter we enumerate these steps and give a recapitulation of most known mobility protocols following the defined steps.

Chapter 5: in this chapter a new handover mechanism is presented and evaluated. Based on collaboration between the mobile node and the network, this scheme allows a fast handover with minimum buffered packets. MIH is used in efficient manner with FMIPv6 to perform intelligent handovers for both heterogeneous and homogeneous networks. The choice of MIH comes from its independence regarding access network technology and its manageability by upper layers, particularly FMIPv6. A theoretical study is conducted in order to compare the performance of the new scheme with the classical ones. A set of simulations are executed as well in network simulator 2. Simulation results confirm the theoretical ones.

Chapter 6: here is yet another improvement to the previous handover process. But this time it concerns FMIPv6 operation itself. It seems that predictive mode is tied to the result of FBACK message which should be received by the mobile node in the old link; otherwise, the reactive mode is activated. In order to avoid this mode we propose to forward the FBACK message to the new location of mobile node through the established tunnel. Therefore the only case FMIPv6 operates in the reactive mode is when the fast binding update is not received correctly by the old access router. We show through analytical study that this small modification in FMIPv6 protocol has a good effect on maximizing the probability of predictive mode success.

Chapter 7: in this chapter mobility in IMS is tackled. As one of IMS purposes is to provide access network independent services, the question of service continuity between heterogeneous network technologies is important but still unsolved. We first analysed the existing signalling protocols within IMS and than deduced that the missing part towards a seamless handover is the handover smoothing step as we stated in the classification done in chapter 4. Then we propose to use FMIPv6 to perform the missing operation. Thus we use two mobility protocols: SIP at the application layer and FMIPv6 at the network layer. This mechanism does not only avoid redundancy in the network but improve the performance of the handover as well. Through an analytical study we show the advantages of our proposal over previous proposed solutions.

Chapter 8: in this chapter we present a comparative study of mobile IP variants in vertical handover. We demonstrate through this comparison that the choice of the best mobility protocol is not obvious as it might appear. In fact, having multiple interfaces, the mobile node can perform faster handovers with the conventional MIPv6 than FMIPv6. We show that the choice of the best variant of mobile IP depends on certain parameters. We finally give some guidelines that should help in choosing the most convenient mobility protocol.

STATE OF THE ART: WIRELESS ACCESS NETWORKS AND MOBILITY PROTOCOLS

This chapter is divided in two parts: the first part gives an overview of the most known access technologies from different standard bodies (i.e. IEEE family and ITU-T family). The second part investigates legacy mobility protocols from different perspectives. Mobility protocols are classified to micro and macro mobility protocols following their administrative range. Macro mobility protocols in their turn are classified following the layer at which they operate.

1.5.Overview of Access Network Technologies

1.5.1. IEEE Family

1.5.1.1. IEEE802.11

Wireless local area networks (WLAN) have gained a big success in the last decade. This success is due to the easy deployment of this type of networks and its low cost compared to wired solutions. Moreover, the offered bandwidth in such network is still increasing with the improvement of modulation schemes and the use of smart antennas along with advanced error correction schemes. From 802.11b to 802.11n, the bit rate passed from few mega bits per second to several hundreds. Hereafter we give an overview of the famous IEEE 802.11 releases.

1.5.1.2. IEEE802.11b

802.11b [2] is the first standardized WLAN technology operating in the 2,4GHz unlicensed band. The bandwidth is divided into channels of 22MHz. The physical layer uses Direct Sequence Spread Spectrum technique and Binary Phase Shift Keying (BPSK), Differential Quadrature Phase Shift Keying (DQPSK), and Complementary Code Keying (CCK) as modulation schemes for 1Mbps, 2Mbps and (5Mbps, 11Mbps)

bit rates respectively. The medium access control is based on carrier sense multiple access with collision avoidance (CSMA/CA).

1.5.1.3. IEEE802.11a

IEEE 802.11a [3] is an amendment to 802.11 standard which operates in the 5GHz licensed band. It was designed for higher bandwidth applications than those provided by IEEE 802.11b. The maximal offered bit rate is 54Mbps using orthogonal frequency division multiplexing (OFDM). Modulation schemes start from BPSK for bit rate of 6Mbps and ends with 64-QAM for 54Mbps bit rate.

1.5.1.4. IEEE802.11g

The goal of 802.11g [4] was to provide a high throughput in the 2,4GHz band while maintaining compatibility with IEEE 802.11b. The resulting standard provides optional data rates of up to 54Mbps, and backwards compatibility with 802.11b devices to protect investments in legacy WLAN installations. It uses OFDM (the same technology used in 802.11a but in the spectrum of 802.11b) and CCK.

1.5.1.5. IEEE802.11n

The goal of the IEEE802.11n standard [5] is to increase the peak throughput, making data flow as fast as possible. The bit rate is up to 600Mbps using 40 MHz channel. The IEEE802.11n standard group makes use of Multiple-Input/Multiple-Output (MIMO) and OFDM in several configurations and provides also backwards compatibility with already installed systems in the same frequency.

1.5.1.6. IEEE802.16

The IEEE 802.16 family was originally designed to provide fixed broadband wireless access for residential and enterprise use in a point-to-multipoint (PMP) architecture. Allowing high bandwidth and rapid deployment of wireless systems, the IEEE 802.16 was immediately recognized as an interesting alternative to the conventional broadband access solutions like Digital Subscriber Line (xDSL) and Fiber To The Home (FTTH), especially in rural areas and developing countries that suffer from the lack of telephony infrastructure. Simple maintenance, scalability and speed of installation will offer better revenue for service providers.

1.5.1.7. IEEE802.16d

Also known as fixed WiMAX [7], it was approved as an upgrade to the IEEE 802.16a [8] standard. The design of a Non Line Of Sight (NLOS) system drove the design of its physical layer, which operates in the 2-11 GHz range. The channel bandwidth is variable from 1.25MHz to 20MHz, which uses 256-carrier OFDM scheme and grants access to different subscriber stations (SS) using Time-Division Multiple Access (TDMA) method. The standard supports multiple modulation levels, including BPSK, Quadrature Phase Shift Keying (QPSK), 16-Quadrature Amplitude Modulation (QAM) and 64-QAM. The modulation in the same sector is adaptive and allows subscribers to adjust the channel modulation scheme according to Signal to Noise Ratio (SNR) of the radio link. When the SNR is good the system can switch to the highest throughput modulation (ex 64-QAM). If fading occurs, the system can shift to a lower throughput modulation without dropping the connection.

1.5.1.8. IEEE802.16e

IEEE 802.16e [9] is the mobile version of WIMAX, it is intended to enable a single base station to support both fixed and mobile broadband wireless access. It provides high data rate Wireless Metropolitan Area Network (WMAN). This standard employs a scalable OFDMA system with 2048-carrier, which can scale the Fast Fourier Transform (FFT) size depending on the channel conditions. As IEEE 802.16e supports mobility, it must cope with two problems not faced by the previous standards: Power Saving and Handover. Mobile WIMAX uses OFDMA. Resources are granted to SS by assigning different subsets of carriers in different time slots.

1.5.1.9. WPAN Family

Wireless personal area networks (WPANs) are focused on a very limited range that can reach several meters. It is called personal because it concerns only the space around a single person or object. The purpose of WPAN standards is to define networks with low-cost, low power, short range and very small size. The IEEE 802.15 working group has defined three classes of WPANs that are differentiated by data rate, battery drain and quality of service (QoS). The high data rate WPAN (IEEE 802.15.3) [10] is suitable for multi-media applications that require very high QoS. Medium rate WPANs

(IEEE 802.15.1/Blueetooth) will handle a variety of tasks ranging from cell phones to PDA communications and have QoS features suitable for voice communications. The low rate WPANs (IEEE 802.15.4/LR-WPAN) [11] is intended to serve a set of industrial, residential and medical applications with very low power consumption and cost requirement not considered by the other WPANs and with low needs for data rate and QoS. The low data rate enables the LR-WPAN to consume very little power. ZigBee is an example of the IEEE802.15.4 that is used in multiple sensor applications.

For these devices to interoperate and communicate over IP networks, a common packet format must be defined to encapsulate layer 3 protocols. Bluetooth Network Encapsulation Protocol (BNEP) [13] encapsulates packets from various networking protocols, which are transported directly over the Bluetooth Logical Link Control and Adaptation Layer Protocol (L2CAP) [14]

1.5.2. Mobile Network Systems

1.5.2.1. 1G

The first commercial cellular network was the Nordic mobile telephone (NMT) deployed in Scandinavian countries in 1981, followed by the advanced mobile phone service (AMPS) in USA in 1983. The European system was known as total access communication system (TACS). These analog wireless systems are referred to as first generation or 1G. They use FDMA as medium access scheme by allocating a 30KHz wide channel for each user.

1.5.2.2. 2G

The need for proposing mobile service to wide number of customers pushed the use of digital system instead of analog one. Hence the global system for mobile communication (GSM) was developed by the European Telecommunications Standard Institute (ETSI), whereas, in North America an equivalent system was developed under the name of IS-95 CDMA known also as cdmaone. 2G systems have known the introduction of the frequency plan which is the reuse pattern of the frequency used in the cells as long as they are not adjacent. GSM system is based on time division multiple access (TDMA) with carrier bands of 200KHz whereas IS-95 system use code division multiple access CDMA.

General Packet Radio Service (GPRS) is the support of data services developed by ETSI. It is referred to as 2.5G cellular system. It defines a packet transmission system that overlays GSM and interworks with internet. A GPRS terminal is assigned an IP address and charged on the basis of transferred data. GPRS can reach a bit rate of 115kpbs by allocating the eight GSM slots to transmit or receive data packets. Data traffic and voice traffic are split by the Base Station Controller (BSC). On the one hand, it sends data packets to the Serving GPRS Support Node (SGSN) and routed afterwards toward packet data network via the Gateway GPRS Support Node (GGSN). On the other hand, voice communications are routed to the circuit switched network via the Mobile service Switching Center (MSC). An enhancement of the data rate of GSM/GPRS consists of changing the modulation scheme from Gaussian Minimum Shift Keying (GMSK) to 8 Phase Shift Keying (8-PSK). The resulting product is called Enhanced Data Rates for GSM Evolution (EDGE). However, the maximum offered rate is only 384kbps which keeps EDGE in the 2.5G category [21].

1.5.2.3. 3G

The evolution towards third generation cellular system was driven by the need of high bit rate and quality of service in order to serve multimedia content like video. Characteristics and requirement of 3G systems are specified in the ITU project called International Mobile Telephony 2000 (IMT-2000).

The goal of IMT-2000 is to have one worldwide standard and a common frequency band with a maximum data rate of 2Mbps. Two standards answered the requirements of ITM-2000: UMTS and CDMA2000. Universal Mobile Telecommunications System (UMTS) is the evolution of GSM system managed by third Generation Partnership Project 3GPP. cdmaone evolution has led to CDMA2000 which is managed by the 3GPP2 standard body. Newer standards surpass the requirements of IMT-2000 and are referred to as 3.5G and 3.75G like High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA) respectively.

Long Term Evolution (LTE) and mobile WIMAX are considered as pre-4G technologies or 3.9G. In spite of the high service level provided by these technologies, they offer less than what is expected for 4G networks.

1.5.2.4. 4G

ITU has defined in 2002 a new vision for future mobile system called IMT-advanced as requirements for the fourth generation. 4G aims to provide high quality multimedia applications to mobile and fixed terminals. The targeted data rates are 100Mbps in high mobility and 1Gbps for fixed access. ITU has received two proposals that are candidates for IMT-advanced [15]: 802.16m from IEEE and LTE-advanced from 3GPP. These standards are still in progress and should be finalized by 2012.

1.5.3. QoS Provisioning in Wireless Access Networks

1.5.3.1. IEEE 802.11e

IEEE 802.11e [6] is an amendment to 802.11 standard introducing quality of service provisioning for different types of traffic in particular VoIP, video, best effort and background traffic. An enhancement is made to the basic Distributed Contention Function (DCF) by defining different behavior for backoff timer and transmission operation time in order to provide the needed resources for high priority traffic. Enhanced Distributed Channel Access (EDCA) defines four access categories (ACs) with different priorities depending on the application as listed in Table1.

Table1: traffic classification in IEEE802.11e

Priority	Access Category	designation
1-2	AC_BK	Background
0-3	AC_BE	Best effort
4-5	AC_VI	Video
6-7	AC_VO	Voice

1.5.3.2. QoS in 802.16

QoS is an inherent feature of WIMAX. Scheduling mechanisms are implemented in both base station and mobile terminal to match QoS requirement of the application to

the available time slots and sub carriers. In the uplink, the allocation of the needed resources is done accurately thanks to bandwidth request/bandwidth grant paradigm prior to data transmission. Five QoS categories are defined following the requirements of the application; these categories are listed in Table 2.

Table 2: traffic classification in WIMAX

Class of service	description	Type of application
Best Effort (BE)	Basic service with no guaranty for packet delivery	Web browsing, email,
non Real-Time Polling service (nRTPS)	Terminals are polled before allocating bandwidth with no latency constraints	File download (FTP)
Real-Time Polling Service (RTPS)	Terminals are polled before allocating requested bandwidth and packets should leave the network within certain latency	Video streaming (MPEG)
extended Real-Time Polling Service (eRTPS)	This method is in midway between UGS and RTPS that matches better requirements of VoIP with silence suppression	Voice over IP with silence suppression
Unsolicited Grant Service (UGS)	Bandwidth is granted without prior solicitation	Voice over IP, video conferencing

1.5.3.3. QoS in UMTS

3G systems introduce new IP-based services for mobile users. Some of the new services such as video streaming applications need certain level of QoS in order to provide acceptable quality of experience for the user. 3GPP has released special technical specification [16] for QoS support in UMTS. In this specification, five QoS classes have been defined (see Table 3).

Table 3: traffic classification in UMTS

Traffic class	Characteristics	Type of application
Conversational class	Preserve time relation between information entities of the stream with stringent and low delay	Voice

Streaming class	Preserve time relation between information entities of the stream	Video streaming
Interactive class	Request response pattern Preserve payload content	Web browsing
Background class	Destination is not expecting data within a certain time Preserve payload content	Emails ,background download

1.5.3.4. QoS Mapping in Heterogeneous Environment

It is very important to guaranty the same QoS level for the user when roaming between heterogeneous networks. The absence of such support will affect the quality of experience of the user. This task is not simple since each network has its own definition and categorization of service classes as it has been shown in the above paragraphs. Though, an approximation of the QoS level can be made by mapping service classes from each network technology according to the served application. Table 4 gives the proposed mapping for UMTS, WIMAX and 802.11e standards.

Table 4: mapping of QoS classes

Application type	UMTS	802.16	802.11e
Voice	Conversational class	UGS	AC_VO
Video	Streaming class	RTPS	AC_VI
Web browsing, file download	Interactive class	Best effort, nRTPS	AC_BE
Background traffic	Background class	Best effort	AC_BK

1.6.Overview of Mobility Mechanisms

1.6.1. Types of Mobility

There are several types of mobility following the action taken by the user and the use case. Some mobility types are considered as an extra service while others are indispensable. Hereafter, the four main mobility types are introduced.

1.6.1.1. Personal Mobility

We talk about personal mobility when a single user is located at different terminals using the same logical address. Two cases are possible: one address for many potential terminals and many addresses reaching one terminal. Mapping between the different terminals and the logical address may need a dedicated server. An example of such server is the SIP registrar which maps the URI to the different IP addresses. If a user would like to be reachable on a mobile phone, a PC and a wireless device, he/she may use these devices either at the same time or alternate between them.

1.6.1.2. Service Mobility

Service mobility allows users to maintain access to their services while moving or changing devices and service providers. Services like address books, call logs or presence service can be maintained in the new location of the user. Service mobility adds certain difficulties for home service provider such as media adaptation and QoS provisioning in the new network in order to maintain service delivery at an acceptable level. This kind of mobility needs prior agreement between service providers. It should also be possible to update and customize these service definitions from the new location (new terminal or network).

1.6.1.3. Terminal Mobility

This is the most considered type of mobility in research since it is a mandatory service in wireless networks. Terminal mobility allows a device to move between different networks while continuing to communicate with its corresponding peers. Terminal mobility management protocols can intervene in different levels in the protocol stack in order to ensure uninterrupted service. Several issues should be faced by the terminal mobility protocols in order to keep acceptable perceived quality of experience regarding the ongoing communication.

1.6.1.4. Session Mobility

Session mobility allows a user to maintain a media session while changing terminals. For example, a user watching a video film in a laptop may want to continue the session in his high definition screen. Here again, media adaptation should be supported by the content provider, otherwise the QoS of the media will be affected.

Another scenario of session mobility is when the user wants to split the media to be played in different devices, for example, playing the audio stream in a speaker phone and the video stream in a video projector.

1.6.2. Layer 2 Mechanisms

Layer 2 mobility management consists mainly of ensuring access to network resources in the target network.

1.6.2.1. Mobility in Mobile Systems

Handover management in GSM networks is known as mobile-assisted handover (MAHO) because it is handled entirely by the network. More specifically the BSC and MSC are in charge of this task. When the mobile moves from one BTS to another within the same BSC the handover coordination is done by the BSC. In the case this handover is done towards another BTS which is under control of a different BSC, the handover is coordinated by the MSC. Continuous signal measurements are collected by the BSC and MSC and used in deciding which mobile should perform the handover and which cell it should handover to.

1.6.2.2. IEEE 802.11r

This standard specifies fast Basic Service Set (BSS) transitions. The main target of the IEEE 802.11r standard [17] is to reduce handoff time in order to avoid connectivity failures and packet losses while users move from a serving Access Point (AP) to a target AP. This handoff time is due to the re-authentication process that occurs in the target AP. IEEE 802.11r avoids the re-authentication process. This is especially relevant for real time applications. Coordination between serving AP and target AP takes place before traffic is routed between them.

1.6.2.3. Mobile WIMAX

Handover is handled in mobile WIMAX by using MAC messages. Various handover strategies are defined [18].

- Hard handoff (HHO): in this handoff method, the mobile station breaks contact with the serving base station before making connection with the new one. This

approach is a break-before-make approach but the handover delay is kept less than 50 ms.

- Fast base station switching (FBSS): this is a soft handover method where the mobile station is in contact with all active base stations within its coverage called “active set”. When the mobile decides to change its serving base station or “anchor base station” it sends a message in a special channel called channel quality indicator. This method requires a perfect synchronization between base stations.
- Macro diversity handover (MHDO): in this method the handover is made more seamlessly than FBSS because the mobile has simultaneous communication with all base stations of the active set.

1.6.3. Layer 3 Mechanisms

Mobility is an intrinsic issue of IP protocol. This problem comes from the fact that the design of TCP/IP stack gives two roles to the IP address simultaneously. The first role is the localization. Therefore, any node in internet is reachable with its IP address which appears in the destination field of the packet. As the internet is organized in hierarchical structure and the routing protocol uses principally the network address part of the IP address to decide which route to follow, it is not acceptable for a node to have an IP address which is not homogeneous with its sub-network. Otherwise, the packet destined to this node will never be delivered. The second role of IP address is the identification of the application, especially when using TCP as transport protocol. Actually every application uses the couple (IP address, port number) as unique identifier for the end to end communication. Any change in one of these two parameters leads to a breaking in the application flow. Discussion about the duality of the IP address has been addressed during the specification of IPv6, but the IETF members did not succeed in separating the localization from the identification in IPv6. Although some proposals try to overcome this problem like Locator/ID Separation Protocol (LISP) [19] and Host Identity Protocol (HIP) [20], they are not widely deployed. Instead of treating the source of the problem, the whole research community is concerned about finding solutions for

the symptoms. In the following we give an overview of the most known mobility management schemes in the network layer.

1.6.3.1. Micro Mobility

When a terminal changes frequently its subnet within the same domain we talk about micro mobility.

1.6.3.1.1. Cellular IP

Cellular IP [22] inherits cellular principles used in cellular networks like GSM for mobility management, but implements them around the IP paradigm. Cellular IP access networks require minimal configuration, therefore, easing the deployment and management of wireless access networks. The major component of Cellular IP access networks is the base station which acts as wireless access point and router of IP packets. Base stations are built on a regular IP forwarding engine with the exception that IP routing is replaced by Cellular IP routing with location management support. Mobile hosts attached to the access network use the IP address of the gateway as their care-of address. Figure 1 illustrates the path taken by packets addressed to the mobile node. All packets destined to the MN reach first the gateway from which they are routed through the base stations to their respective IP address.

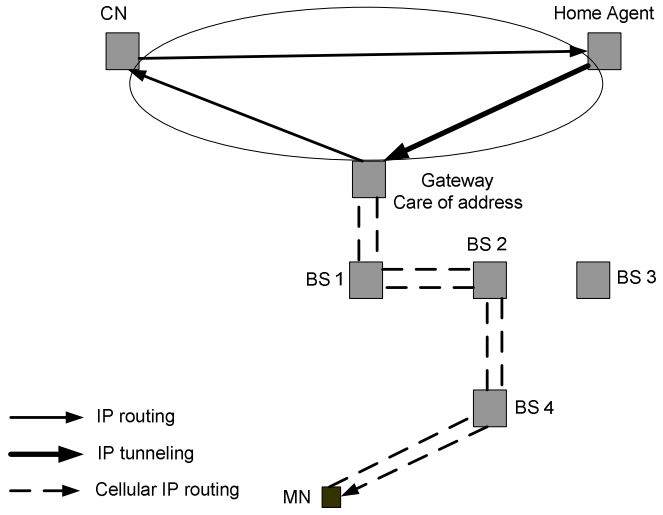

Figure 1: Cellular IP operation

In Cellular IP, location management and handoff support are integrated with routing. To minimize control messaging, regular data packets transmitted by mobile

hosts are used to refresh host location information. Uplink packets are routed from a mobile host to the gateway on a hop-by-hop basis. The path taken by these packets is cached by all intermediate base stations. To route downlink packets, the path used by packets recently transmitted from the mobile host is reversed. When the mobile host has no data to transmit, it sends small, special IP packets toward the gateway to maintain its downlink routing state.

1.6.3.1.2. Intradomain Mobility Management Protocol

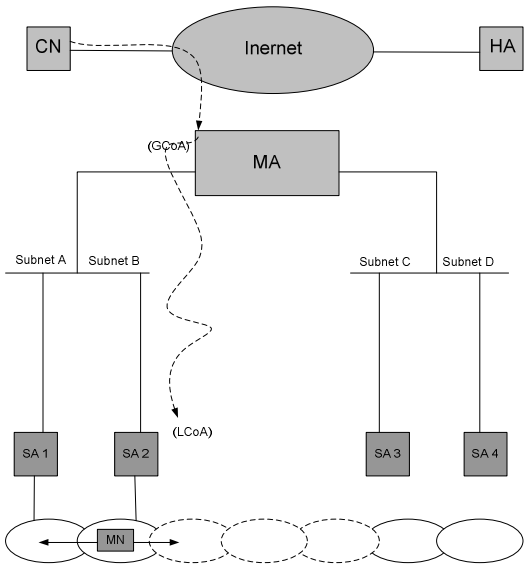

Figure 2: IDMP architecture

Intradomain mobility management protocol (IDMP) [23] is designed to work as a standalone solution. As illustrated in the Figure 2, IDMP architecture relies on two network entities: Mobility Agent (MA) and Subnet Agent (SA). MA is responsible for mobility management in the whole domain and SA is in charge of mobility management within the subnet.

In IDMP, the MN acquires two types of CoA: Local care-of address (LCoA) which identifies the MN's attachment to the subnet and Global care-of address (GCoA). The MN should inform its MA about any change in the LCoA. While MN updates its GCoA only if it changes the domain. All packets addressed to the MN are forwarded to the GCoA where they are intercepted by the MA. MA then encapsulates these packets to the MN's current LCoA.

IDMP defines also a fast handoff procedure based on layer 2 indicators. In order to minimize service disruption the MN generates a movement imminent message to the MA. Upon receiving this message, MA multicasts the packets destined to the MN towards a set of neighboring base stations. Packets are buffered in the BSs until attachment of the MN to the new subnet, and then they are immediately forwarded.

1.6.3.1.3. HAWAII

Mobile IP results in high overhead when the mobile is changing its subnet frequently because the MN should update its HA each time it acquires a new CoA. Moreover when QoS is provided by the network, the QoS reservation from HA to FA has to be reestablished even if most of the path remains unchanged. For these reasons Handoff Aware Wireless Access Internet Infrastructure (HAWAII) [24] was introduced as a complement of mobile IP to support intradomain mobility management. It uses specialized path setup schemes which install host-based forwarding entries in specific routers to support intra-domain micro-mobility. These entries reduce mobility related disruption to user applications, and at the same time reduce the number of mobility related updates. Moreover, HAWAII simplifies quality of service support since mobile hosts retain their network address while moving within the domain.

Protocol operation is as follows: each MN has an IP address and a home domain, when moving within the same domain, MN maintains its IP address. When the MN enters into a foreign domain, the MN is assigned a co-located CoA using DHCP. Packets destined to the MN are tunneled to the CoA. They reach first the domain root router based on the subnet address of the domain and then they are forwarded over special dynamically established paths using host-based routes in routers towards the MN. Mobile IP registration is split in two parts: between MN and BS and between BS and HA. This separation helps in reducing the updates of the HA.

1.6.3.2. Macro Mobility

1.6.3.2.1. Mobile IPv4

Mobile IPv4 [25] has been introduced by IETF to deal with IP address change in access networks. It allows to all corresponding nodes that are currently in communication with

the mobile node and future correspondents to keep the same identity whatever the network to which the mobile is connected. As routing protocols in today internet are based on network addresses, it is clear that when the mobile moves to a new subnet, routing its packets in the downlink is impossible since its address doesn't belong to the current subnet. The idea behind MIPv4 is to hide the modification in the network layer from the application by defining two types of addresses: i) home address which is a fixed address used as an identifier for applications and never changes wherever the MN is located, ii) Care of Address (CoA) is a variable address used temporarily by the mobile node in the visited network. When the MN quits the home network and attaches to a foreign network, it listens to the advertisement sent by the Foreign Agent (FA) and then sends a registration to the FA which relays the request to the Home agent to check if the MN is authorized to use MIPv4 service. After verification, the FA sends a reply to the MN including the CoA to be used by the MN along its sojourn in this network. CoA in MIPv4 is of two kinds: FA-CoA and Co-located CoA. In case of FA-CoA, the MN use the address of the FA as a CoA, where as any IP address that belongs to the sub-network can be used in case of Co-located CoA (i.e. MN's CoA can be acquired by DHCP server). As far as the correspondent node (CN) is concerned, it continues sending data packets to the home address as if the MN is still in the home network. These packets are intercepted by the HA and encapsulated towards the FA. Then they are decapsulated in the FA and forwarded to the MN (see Figure 3). In Co-located CoA the tunnel is established between the HA and the MN without passing through the FA.

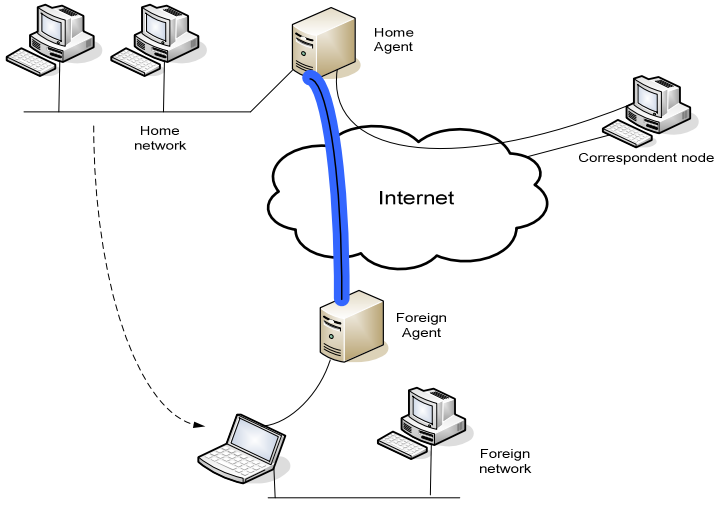

Figure 3: Mobile IPv4

Limitations of MIPv4:

The main drawback of MIPv4 is its high handover delay. This delay comes from the fact that the MN should wait for advertisements of FA, and then register to the FA before the tunnel can be established and eventually forward packets. The result of such long handover delay is the loss of some data packets. Secondly routing through the HA is inefficient in case the MN is far away from the HA and leads to a high end to end delay and high overhead, this phenomena is known as triangular routing. Finally, HA is a single point of failure on which the whole system operation depends. Moreover traffic of all MNs go through the HA which can lead to bottle neck creation in the home network.

1.6.3.2.2. Mobile IPv6

Although mobile IPv6 [26] maintains the same functioning principle of the previous version, it introduced some improvements. Taking advantage from the IPv6 structure, a mobility header was defined for MIPv6 as an extension header which entails FA suppression. In addition, the route optimization is used as a fundamental support rather than an extension, and the correspondent node is updated securely by using the new return routability procedure. Moreover, neighbor discovery is used instead of address resolution protocol (ARP), thus decoupling MIPv6 from the link layer. MIPv6 operates in two modes: reverse tunneling and route optimization. In reverse tunneling

mode, the MN receives the new CoA by means of router solicitation and router advertisement, then updates its binding with the HA which creates the bidirectional tunnel.

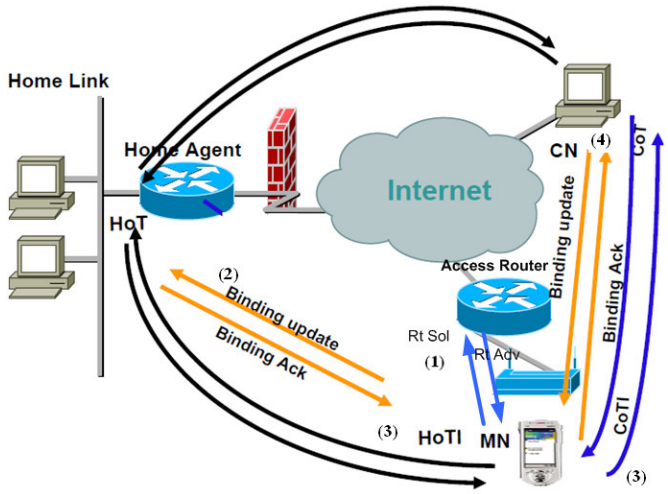

Figure 4: MIPv6 route optimization with return routability

When route optimization is used (Figure 4), the MN should also update the CN after updating the HA. A security mechanism called return routability is introduced in the MIPv6 to secure the binding update with the CN. Home Test Init and Care-of Test messages are sent simultaneously via HA and directly to the CN respectively. Afterwards, CN and MN communicate directly without going through HA. Route optimization allows avoiding triangular routing.

1.6.3.2.3. Hierarchical Mobile IPv6

Hierarchical mobile IPv6 [28] was introduced to reduce both the amount and delay of signaling messages between the MN and the HA/CN. A new entity called Mobility Anchor Point (MAP) is introduced to maintain tracking of the MN within a defined domain using a new care of address called Regional Care-of-Address (RCoA). MN should register its RCoA within the HA when moving to another network in a different domain (see Figure 5). In case of intra-MAP mobility, the only operation that the MN should perform, is binding update (BU) of its new acquired on-Link CoA with the MAP. Data packets are tunneled twice in HA and MAP before being forwarded to the MN.

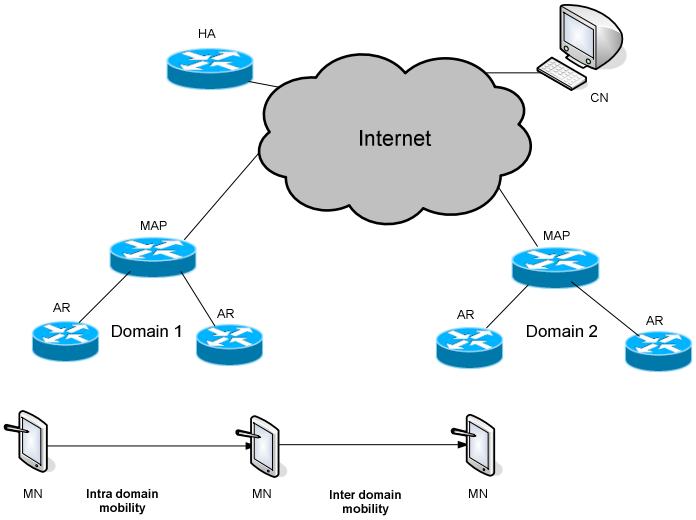

Figure 5: Hierarchical mobile IPv6

1.6.3.2.4. Proxy Mobile IPv6

MIPv6 needs the intervention of the mobile to achieve mobility operation. Network-based mobility is another approach to solve the IP mobility issue. A proxy mobility agent performs the signaling with the HA and does the mobility management on behalf of the MN. For this reason, this mobility management protocol is referred to as Proxy Mobile IPv6 (PMIPv6) [29]. PMIPv6 was recently adopted as mobility management protocol for packet data networks in 3GPP (TS 29.275) and 3GPP2 (3GPP2 X.S0057-0)

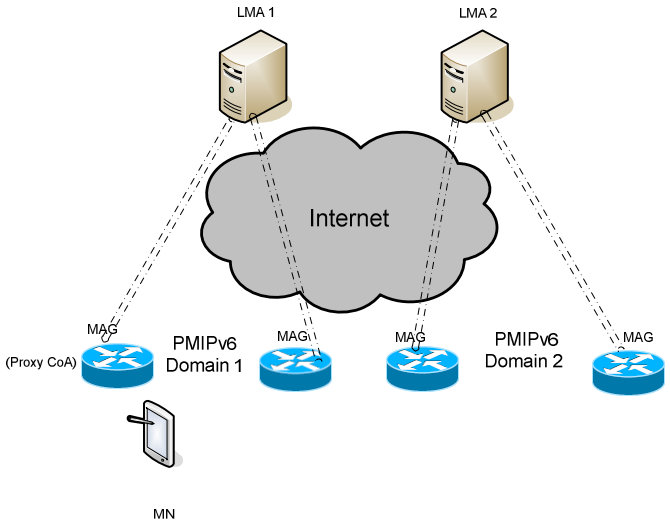

Figure 6: Proxy Mobile IPv6 architecture

Special network entities track MN's movement, initiate the mobility signaling and set up the required routing state. These tasks are performed by two network entities: Local Mobility Anchor (LMA) and Mobile Access Gateway (MAG) (see Figure 6). LMA has the functional capabilities of the HA as defined in MIPv6. MAG is a function on the access router that manages the mobility-related signaling for a MN that is attached to its access link. It is responsible for tracking the MN's movements and signaling management with LMA on behalf of the MN. LMA being the topological anchor point for the MN home network prefix(es), receives any packets that are sent to the MN by any other node and forwards them to the MAG through a bi-directional tunnel. The MAG on the other end of the tunnel removes the outer header and forwards the packet on the access link to the MN.

1.6.3.2.5. Fast Handovers for Mobile IPv6

Latency caused by MIPv6 operation is unacceptable for real-time and throughput sensitive applications. To overcome this problem a fast handover scheme [27] was proposed by the IETF. Fast handovers for Mobile IPv6 (FMIPv6) allows the MN to anticipate the IP address acquisition and forward packets from previous access router (PAR) to new access router (NAR) during the handover. FMIPv6 relies on layer 2 triggers to warn MN that the signal strength is going down, then the MN sends a router solicitation proxy (RtsolPr) and gets a proxy router advertisement (PrRtAdv) which contains information about the neighboring cells especially (AP-ID, AR-Info). This couple contains access router's MAC and IP addresses, and the valid prefix on the interface to which the Access Point (identified by AP-ID) is attached. With this information, the MN formulates a prospective New CoA (NCoA) and sends a fast binding update (FBU) to PAR. The purpose of the FBU is to authorize PAR to bind Previous CoA (PCoA) to NCoA, so that arriving packets can be tunneled to the new location of the MN. PAR sends Handover Initiate (HI) message to carry the NCoA to the NAR which determine after a Duplicate Address Detection (DAD) whether NCoA is unique on its link interface or not. In case of address conflict, NAR assigns another address and includes it in the Hack message and the PAR in return will assign it in the FBack. After attaching to the new network, The MN sends unsolicited neighbor

announcement (UNA) immediately, so that buffered packets at NAR can be forwarded to the MN right away. The tunnel created between the two routers remains active until the MN completes the binding update with its HA and/or CN.

FMIPv6 is not tied to MIPv6 or any other protocol as it might be thought. It can assist any mobility management protocol by allowing a smooth handoff wkile the mobility protocol makes the necessary updates. Usually FMIPv6 is used with MIPv6 and PMIPv6 [30] but can also be used with upper layer protocols like SIP [31].

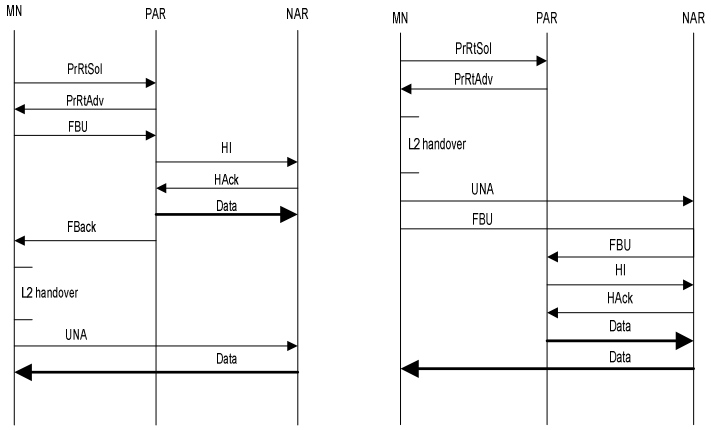

Figure 7: predictive mode (left) and reactive mode (right)

FMIPv6 is operating in two modes (see Figure 7): *predictive mode* when FBack is received in the previous link and *reactive mode* when FBack fails to attain the MN because of an unexpected link down for example. PAR starts forwarding packets through the tunnel upon receiving an acknowledgement of HI from the NAR without having any indication of FBack reception by the MN. In the case where MN does not receive the FBack because of a link down, it falls into the reactive mode and resends another FBU on the new link.

1.6.3.3. Upper Layer Mechanisms

In this section we give an overview of upper layer mobility management protocols. These protocols manage mobility in transport level or application level. At the opposite of layer 3 mechanisms, upper layers protocols get the end points more involved in the mobility operation.

1.6.3.3.1. TCP migrate

The objective of TCP migrate [32] is to handle mobility at the transport level. Where in conventional TCP, the connection shall break with any change in the couple (IP address, port number), TCP migrate deals with this change by updating the binding in the corresponding node. TCP migrate proposes to extend TCP in order to support mobility. This is done by adding a new option in the conventional SYN message that informs the other TCP peers about the support of TCP migration. The identification of the connection is done by means of token that has been negotiated at the beginning of the connection. In case of IP address change, a Migrate SYN is sent from the mobile host to the server in order to update the connection without creating another one. TCP Migrate can be used also in the server side in order to load balance long TCP sessions. TCP migrate does not need any kind of infrastructure and does not add any overhead to the packet. It adds only minimal changes to the existing TCP protocol. Nonetheless it needs a location management system that binds the new acquired IP address to the mobile host identifier. Moreover TCP migrate solves only the mobility of TCP, whereas many applications that need mobility support use UDP as transport protocol.

1.6.3.3.2. mobile Stream Control Transport Protocol

Stream control transport protocol (SCTP) [32] is a transport protocol that provides some similar services as TCP. It ensures services like reliability and congestion control. SCTP introduces the idea of multi-homing, where a single endpoint can support multiple connections with different interfaces and IP addresses simultaneously, and allows dynamic changing of addresses. To support multi-homing, SCTP endpoints exchange lists of IP addresses during the initiation of an association. Mobile SCTP [34] provides end to end communication between the endpoints without requirements to network elements. When the mobile client changes its point of attachment and after acquiring the IP address by means of DHCP, SCTP will bind the new IP address to the existing SCTP association. This is done by sending a configuration change message (ASCONF) to the fixed server. While the mobile continues to move toward the new location, it needs to change its primary IP address to the new IP address. If the old IP address gets inactive, the mobile deletes the IP address from the address list. As mSCTP

does not provide location management of the MN, it can be used in conjunction with SIP or Mobile IP.

1.6.3.3.3. Host Identity Protocol

Host identity protocol [20] aims to resolve the duality problem of IP address by isolating the identifier from the locator role. Therefore, a new layer is inserted between the network layer and the transport layer allowing the transport protocol to use the host identity (HI) as an identifier rather then IP address. Practically a HI is a public key of the end point and can be presented in a compact format by means of hash function. The resulting key is called host identity tag (HIT). HI is similar to URI in SIP and needs a new namespace to resolve the IP address given a HI. The mapping between HI/HIT and the corresponding IP address is done in a domain name system (DNS) where the HI/HIP can be communicated while trying to resolve the fully qualified domain name of a host. A second solution is to use a dedicated rendezvous server (RVS) which takes in charge the mapping between HI/HIT and IP address while DNS maintain the mapping between the FQDN, HI/HIT and the corresponding RVS which are more static entries. When a host changes its IP address without being in communication (i.e. pre-session mobility), the mobile host updates its RVS to map its HI/HIT to the new IP address. In case of mid-session mobility scenario, the mobile host sends a HIP update message to the corresponding host to update the destination of the packets. This operation is transparent for the application since the connection identifier (i.e. HI/HIT) has not been changed during the session. Security between peers in HIP is strengthened by using Diffie-Hellman key exchange for mutual authentication. It also limits man in the middle and denial of service attacks.

1.6.3.3.4. Voice Call Continuity

Voice Call Continuity (VCC) [35] is an IMS application that provides capabilities to transfer voice communications between the circuit switched domain and packet switched domain via the IP Multimedia Subsystem (IMS). VCC provides functions for voice call origination, termination and transfer between the CS domain and the IMS and vice versa. VCC application is implemented in the user's home network. Its role is to

anchor Voice calls from and to a UE to provide voice continuity for the user during transition between the CS domain and the IMS. VCC is composed of a set of functions required for a UE to establish voice calls and control the switching between CS domain and IMS whilst maintaining the active session. The corner stone function in VCC is domain transfer function (DTF) which controls and executes the transfer. It should be noted that VCC does not specify any mechanism for handover in the radio interface but maintains only the continuity of the call between different domains (i.e. CS domain and PS domain).

1.6.3.3.5. Session Initiation Protocol

Session initiation protocol [36] is an IETF standard intended to initiate, control and terminate multimedia sessions. It supports several types of mobility as it was claimed in [37]. SIP is an applicative protocol that controls the session but is independent of data transport. Usually transport parameters (i.e. IP address and port number) are negotiated at the beginning of the session using SIP messages such as INVITE. Any change of those parameters during the session can be updated by sending a re-INVITE message containing the new transport parameters. Afterwards, the RTP packets are sent from the CN with the new destination address. In the context of SIP, this operation is known as mid-call mobility since the mobile move towards a new network during the session. The other type of mobility support is called pre-call mobility. It consists simply of updating the SIP registrar server with the new IP address so that future corresponding nodes can track the mobile node and establish a call with it.

SESSION MOBILITY FOR VIDEO STREAMING

1.7.Introduction

Nowadays, telecom operators are making a remarkable progress in providing a wide offer of broadband access to answer the high demand for high bit rate applications. Nevertheless, user requirements do not stop at providing high rate connection, but exceeds it to ensuring transparent service portability among his equipments. The user would like to choose among his devices those which respond at best his needs and constraints. From small smart phones to large screen devices, the customer enjoys its entertainments or business meetings according to its current situation. Service continuity over different terminals known as session mobility is a challenging operation in terms of handover latency, context transfer and media adaptation. Moreover, this transfer requires synchronization between the involved terminals.

In this chapter we present "SESSAMO", a new lightweight session mobility protocol for streaming applications using RTSP. This solution is transparent to the network and does not require any changes in the client or server streaming application. The solution is detailed and a set of measurement results are presented. In addition, we present a new method to renegotiate session parameters following terminal capacities in order to adapt the flow accordingly. Renegotiation proposal is based on the use of SDPng and MPEG-21.

1.8.Principle of Session Mobility

The goal of session mobility is to give the users the possibility of switching from one terminal to another when enjoying the same multimedia session without any interruption. This operation proved to be useful for users who are in mobility. A typical scenario of session mobility is a user who is using his personal data assistant (PDA) to watch his favorite game when he is outside. The PDA is connected to internet via the

public land mobile network (PLMN) which provides 3G service. Once at home the user would like to take advantage of his broadband access and his high definition screen to watch the same media without any service initialization or interruption. Session transfer can be either controlled by the device from which the session is transferred (see Figure 8), this mode is called push mode, or by the device to which the session is transferred, this mode is called pull mode.

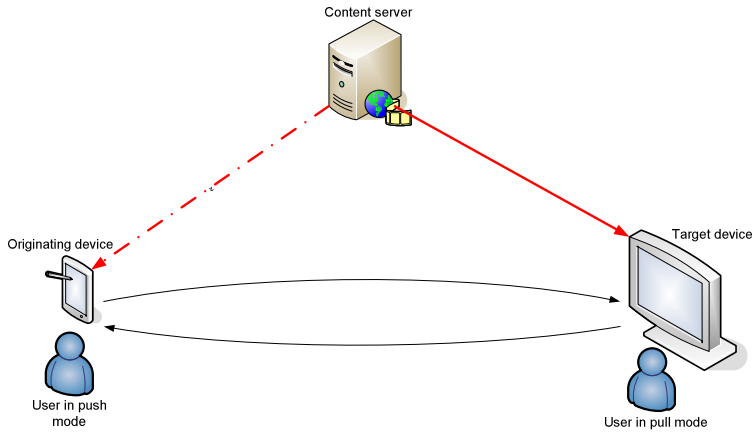

Figure 8: push mode and pull mode in session mobility

1.9.Service Continuity and its Constraints

Service continuity means that there should be no interruption when the media stream is transferred between the two devices. Therefore, session mobility adds a temporal constraint to the transfer operation. Handover delay is the period of time elapsed from the instant when the user chooses to switch to the new device by triggering the transfer (ex: button push) and the time instant when the stream starts playing in the target device. In other words, handover delay represents the reactiveness of the transfer mechanism. In the ideal case there should be no time difference between the instant when the media disappears from the first device and the instant it appears again in the target device. In practice the handover delay is not null due to the delay in transmitting signaling packets and data packets between the involved entities (i.e. media server, target device and originating device). Nevertheless, handover delay should be minimized in order not to disturb user's quality of experience. Service continuity adds another constraint related to synchronization of the media between the two devices. An

accurate session transfer requires that the stream starts in the target device from the instant it was left off in the first device. This issue is a direct consequence of the handover delay. Solving this issue is important in order to preserve the consistency of the service. This means that we must be sure that the user doesn't miss any sequence of the media and minimize any overlapping in the played material (i.e. play a sequence already displayed in the originating device).

1.10. Media Adaptation

Another challenge of session mobility is adaptation of the media being transferred. The variety of devices and their capacities makes it obligatory for the media streams played by these devices to be adapted to their capacities and connectivity rate. For instance the media stream played by the high definition screen is not suitable for a mobile device without any adaptation; otherwise the system will be overloaded and the quality will be very poor.

The difficulty of media adaptation resides in how to adapt the media streams contained on servers to the big variety of screens, CPU capacities, network rates, batteries drain…etc. This process involves a number of tasks where signaling procedures are not completely defined. To achieve this task two procedures are distinguished: negotiation and adaptation. Negotiation is initiated by the client and takes place at the beginning of the session or during the session when one of the parameters change. Adaptation is the action taken by the server after the negotiation in order to make the necessary changes on the served streams.

1.11. Related Work

1.11.1. Mobile IP

MIP is not suitable for session mobility although it can be a considered as a candidate solution. MIP can achieve session mobility since the new terminal has a new IP address and MIP can redirect data packets from the old address to the new address which is actually a new terminal. Nevertheless, not only the session will be transferred, but every thing that was destined to the origination device will be routed to the new

device. This is not the objective of session transfer in video streaming where the transfer concerns only the stream itself

1.11.2. Real Time Streaming Protocol

Real Time Streaming Protocol (RTSP) [39] is an application level protocol providing signaling service for real-time data delivery such as audio and video. Management of the session is provided by a set of methods implementing the classical control player actions such as play, record, pause and stop. Other methods are concerned about describing and negotiating session parameters. RTSP does not provide any data delivery by itself, but relies on transport protocols such as Real-time transport protocol (RTP) [40]. RTSP operation is based on the client/server approach where the client sends request to the server, and the server answers the client requests. Figure 9 illustrates RTSP session establishment and termination. As far as session mobility is concerned, RTSP does not provide any mechanism for session transfer. For this reason external mechanisms should take in charge this operation.

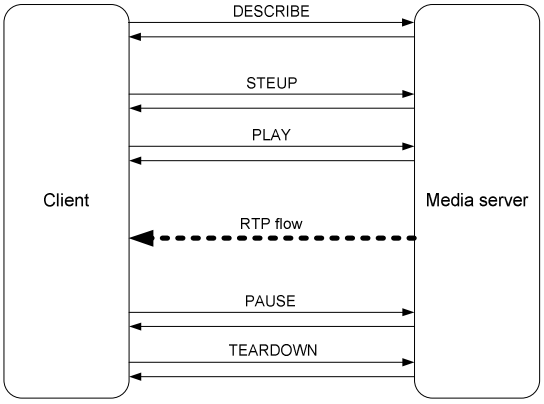

Figure 9: Session establishment and termination in RTSP

1.11.3. Session Initiation Protocol: Refer method

SIP supports many types of mobility such as terminal, session, personal and service mobility by exchanging a number of messages between the concerned entities. At the opposite of RTSP, SIP provides a built-in method for session transfer between different user agents. REFER [41] is an extension method that allows a client to transfer the session to a third party. Many other works are based on the use of SIP session mobility to provide a framework for session transfer as [42][43].

1.12. Proposed Session Mobility Mechanism

As the purpose of our study is providing service continuity for video streaming, we will choose the RTSP as basis of our solution. RTSP is more suitable for this kind of service rather than SIP which is more adequate for interactive applications like telephony and videoconference services. Therefore, a new mechanism for session mobility should be defined and integrated to the streaming application.

1.12.1. Session Mobility Operation

Session mobility allows the user to maintain its session when changing end terminals. The action is started when the user pushes the button to indicate that he/she wants to switch the flow from the first terminal to the second one. A request is immediately sent to the remote terminal. This message contains a description of the current session in an SDP like format. If the new terminal accepts the solicitation, the application sends a command to launch the viewer and the negotiation procedure between the new terminal and the server starts. After negotiation, the target device requests the flow from the server. At the end of this phase, the target terminal begins receiving the flow with requested characteristics; subsequently it sends a message to the first terminal to indicate that the process succeeded. Ending the old session can follow two approaches: make-before-break and break-before-make. Make-before-break approach consists of making sure that the stream is received by the target device before stopping it from the originating one, therefore the acknowledgement is sent only if the data packet is arriving. Whereas break-before-make approach consists simply of acknowledging the transfer request at its reception by target device therefore, the session is stopped in the originating device without any guaranty that the actual session was transferred successfully. Finally, the original RTSP entity sends a TEARDOWN message to the server to stop receiving the flow.

1.12.2. Protocol Description

SESSion And MObility (SESSAMO) protocol is a simple protocol that takes place between the two devices in a peer to peer relationship. It means that both devices implement server and client modules. SESSAMO is based on the exchange of text

messages [44] like HTTP and RTSP. Two main kinds of messages are identified: request messages and response messages. The first conveys command information, whereas the second one transports the status of the operation result. The general format of a SESSAMO message is as following:

Header 1 CRLF

Header 2 CRLF

...

Header n CRLF

A header is always composed of two elements: the header identifier and the value or attribute. Inside the message, the header can take any order. On the other hand, there is a reduced set of headers so that total length of the message does not exceed the Maximum Transfer Unit (MTU).

Hereafter the list of the headers defined for session transfer:

- "type: session ": this header determines the type of the message.
- "sequence: sequence number": this header indicates the sequence number n of the header. It is useful for controlling message loss and message duplication as well as for security.
- "time: time in seconds": this header specifies the time instant t in seconds at which a session has to start from.
- "service: URL": this header specifies the URL (Uniform Resource Locator) from where the session can be obtained. This is the location of the media server serving the current streaming.
- "status: code": this header reports the result of the requested operation. When status code is equal to 1, the operation was successful, whereas 0 means that the operation has failed.

The protocol operation is simple: for each request message there is a response message. Moreover, they have to share the same sequence number. If the sequence number is different, there is a lost or a duplicate message, therefore other measures have to be taken to deal with this inconsistency. If a response message is not acknowledged in a period of RTT seconds (100 ms if RTT is not known), it has to be retransmitted,

where the maximum number of retransmission is seven. A typical request message for session transfer is:

type session CRLF

sequence 3 CRLF

time 1040.36 CRLF

service RTSP://157.159.103.232:8554/ice_age CRLF

And its corresponding response is:

status 1 CRLF

sequence 3 CRLF

1.12.3. Implementation

We use the open source library live555 [45] under linux for RTSP protocol and VLC [46] as a video player in the end devices. Live555 is used as video server and also included as a module during the compilation of VLC player for client support of RTSP. SESSAMO coordinates between the two entities participating in session mobility operation: the originating device (OD) and the target device (TD).SESSAMO is a peer to peer protocol, for this reason any device that supports session mobility has two kinds of programs: SESSAMO server which listens in permanence to requests coming from other terminals, and client part that can transfer the session from the current terminal. Thereby, SESSAMO is operating in push mode. Client program is equipped with a graphical user interface to coordinate the operation between both entities. It includes a set of graphical resources to establish the operating parameters. On the one hand, it controls VLC to accomplish commands such as play, stop, pause…etc. On the other hand, it retrieves status reports about the current video session, such as the last received RTP timestamp. This information is used later in the session transfer operation in order to achieve the session handover accurately. The exchange of information required by the protocol is always started by OD. This occurs when the user pushes the key “transfer of current session” (see Figure 10). This action makes the SESSAMO protocol send a request message to TD. This message contains the necessary information to retrieve the video session. If the TD does not receive a response before RTT seconds, it retransmits

the same message; the operation is repeated if necessary but not more than seven times. After seven attempts, the communication with the TD is considered as impossible.

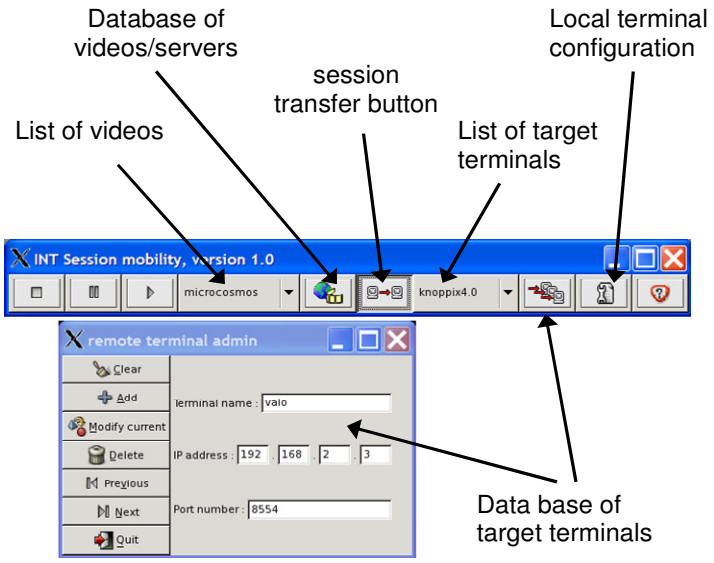

Figure 10: SESSAMO client graphical interface

As far as server program is concerned, it is a daemon running in the background which listens on a special port to the incoming requests from OD and treats them. When the TD receives a request message, it recovers the session description and launches the RTSP player under the following conditions: (1) the video session must be retaken from the point specified by the request message and (2) the video to be negotiated has to be compliant with the TD capabilities.

1.12.4. Testbed

The goal of this testbed is to validate and test the efficiency of our solution in a real video streaming scenario. The testbed is composed of a PC playing the role of the video server; the TD is represented by a laptop, and Nokia 770 internet tablet is the OD (see testbed snapshot in Figure 11). The internet tablet is connected via WIFI to the LinkSys access point which is linked to the other entities through an Ethernet Hub. The tablet is playing the role of the mobile device and the laptop is the high capability device. SESSAMO client and server are developed in C language. As Nokia internet tablet has a different environment (i.e. ARM processor) we compile SESSAMO program using scratchbox [47] (see Figure 12) which is a cross-compiler provided by

Nokia allowing the development and compilation of new applications destined to work on that tablet.

Figure 11: testbed snapshot

Table 5 summarizes the different elements of the testbed and their hardware capabilities.

Table 5: Characteristics of testbed elements

Element	Function	Hardware description
Nokia 770	Originating device	252MHz OMAP, 64 Mo RAM, 802.11b/g
Laptop Dell	Target Device	Dual core1,66GHz CPU, 2Go RAM, Fast Ethernet 100Mbps
PC Dell	Video server	Dual core 3GHz CPU, 1.9Go RAM, Fast Ethernet
LinkSys	Access Point	802.11b/g wireless interface, Fast Ethernet

Figure 12: SESSAMO running on Nokia 770

1.12.5. Performance Evaluation

In order to evaluate the performance of our session mobility solution we consider the different time instances illustrated in Figure 13. We are interested in measuring the following periods of time: session handover delay (t_{shd}), session overlap time (t_{so}) and starting time of the player (t_{stp}).

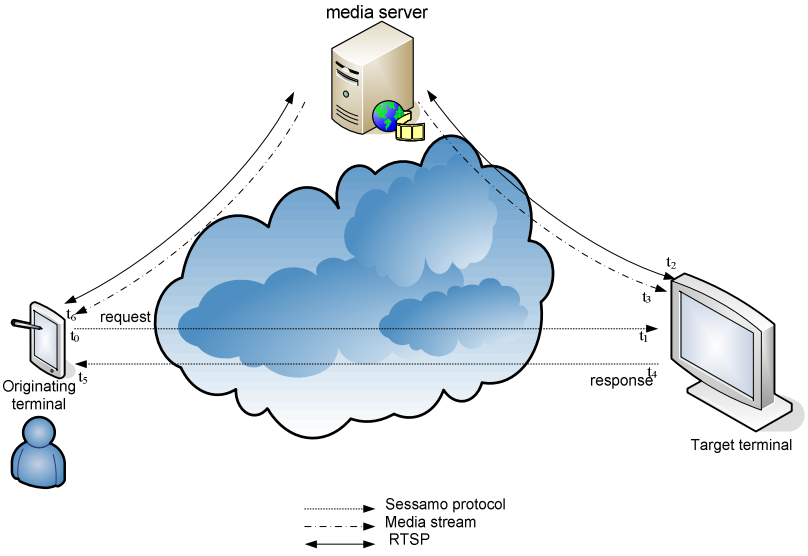

Figure 13: SESSAMO timing

Hereafter the list of the most important time instances used in measuring the performance metrics stated above, these instances are picked up using a network analyzer (wireshark [48]):

t_0: the instant of time at which the user pushes the "transfer session" button.

t_1: the time instant when SESSAMO request message arrives to TD.

t_2: the time instant when TD starts the service negotiation with the video server.

t_3: the time instant when RTP flow is received by TD.

t_4: the time instant when SESSAMO response is sent to OD.

t_5: the instant at which SESSAMO response is received by OD.

t_6: the time instant when OD receives the last packet of audio/video.

t_{shd} is equal to $t_0 - t_3$ and it corresponds to the period of time that goes from the instant t_0 when the user pushes the button "session transfer" until the time instant t_3 when the video packets start arriving to TD. In practice, this period of time is difficult to measure because it involves the clocks of different terminals. Nevertheless, we propose the following approximation: $t_{shd} = t_3 - t_1 + RTT/2$.

As far as the video session overlap time t_{os} is concerned, it indicates the duration of the video sequence that will be played in both devices prior to finish the original session. As the starting time of the session on the TD is equal to the instant of pushing the transfer button, t_{so} corresponds to $t_6 - t_0$. Finally, the starting time of the RTSP player t_{stp} is interesting because it gives as an indication about the effect of this operation on session transfer performance. It should be noted that player launching is not related to the mechanism of the mobility itself, but depends on the operating system and hardware capabilities of the terminal. Indeed, the starting time of each RTSP player varies considerably. It goes from some fractions of second to several seconds. t_{stp} can be obtained by means of $t_{stp} = t_2 - t_1$.

The experimentation was conducted ten times for each scenario and the mean value of each metric is calculated.

1.12.5.1. Make Before Break Scenario

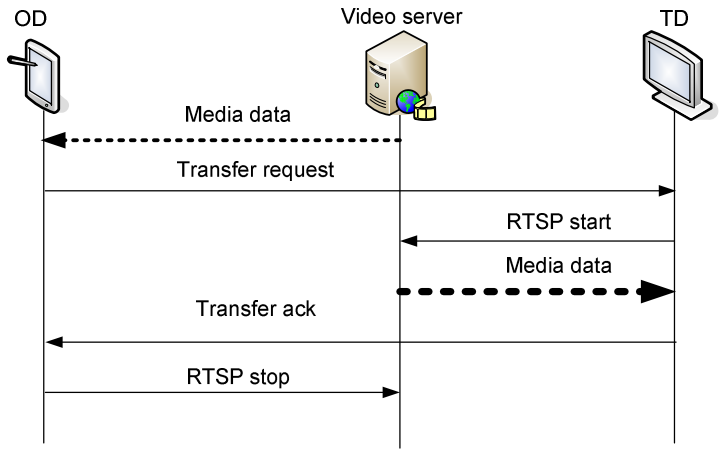

Figure 14: make before break scenario

Make before break approach as illustrated in Figure 14 consists of establishing the session on the TD before the acknowledgement is sent back to the OD. Acknowledgement here means that data traffic is arriving and the session was transferred with success. The advantage of this scheme is the guarantee that the stream is really received by the OD, if this latter can not establish the session with the video server, the acknowledgment will not be sent back and the session will not be terminated on the OD.

Table 6: make before break results

Delay	value in ms
Session handover delay	453
Video session overlap	455
VLC starting time	442

Table 6 summarizes results of session transfer. The handover delay is short and the transfer is almost instantaneous for the user. The disadvantage of this scheme is that video session overlap is high. To be noted that the starting time t_{stp} of the RTSP player represents 97% of the total session handover delay.

1.12.5.2. Break Before Make Scenario

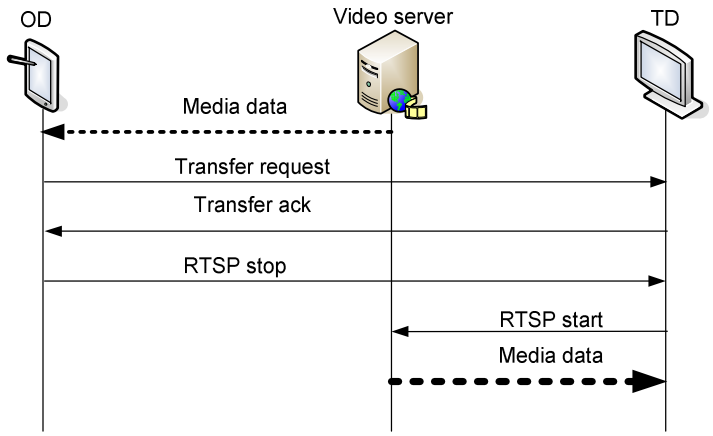

Figure 15: break before make scenario

In break before make scenario (see Figure 15) the acknowledgment is sent immediately after reception of transfer request, and subsequently the session is terminated on the OD. The advantage of this approach is that video overlap is reduced compared to the previous scheme as shown is Table 7. As for session handover delay is almost equal to the make before break handover delay since the principal cause of this delay is VLC starting time with more than 97%. The drawback of this scenario is that a silence time occurs during the transfer operation, especially when the handover delay high.

Table 7: break before make results

Delay	value in ms
Session handover delay	499
Video session overlap	9
VLC starting time	487

1.12.5.3. Synchronization Issue

An ideal session mobility mechanism should provide perfect synchronization between the two devices. In other words TD should resume the session exactly from the same instance the user has triggered transfer button. This problem is not noticeable in small networks as in the case of our testbed where the delay between the different entities is negligible. But when it comes to large networks, congested networks or busy

servers, the delay can be significant. Therefore, the overlap will be important in case of make before break and silence time will be high in case of break before make. In such conditions, make before break approach seems to be more adequate, because at least the user can continue watching the media on the old device until the service is established on the new one. As for video session overlap, it can be compensated by seeking the video in a farther point compared to the button push point. This difference in time should be equivalent to the estimated handover delay.

1.13. Renegotiation of QoS Parameters

The challenge now is to cope with the variety of mobile devices. Indeed, we can find in the market different devices that have different screen sizes, CPU powers, operating systems, network interfaces, supported codecs, battery powers…etc. A server offering a given service has to be capable of satisfying each device according to its own capacities. Signaling protocols such as RTSP and SIP transport in their message payload the QoS constraints imposed by clients and their capacities. Media servers transcode or re-quantize data content accordingly. It is obvious that the media adaptation is a must in case of session mobility as we transfer the flow from one terminal to another. Nevertheless, we tackle the problem of QoS management in session mobility in different phases. The first phase is the negotiation that takes place at the beginning of the session; the second one is the adaptation during the session when some parameters change in the same terminal such as battery level and connection rate. The third phase is the re-negotiation when transferring the session. The particularity of our work resides in the manner these tasks are achieved by means of SDPng/MPEG-21 and RTSP. In particular, QoS service adaptation is not based on a classical approach where server adapts the flow without any participation from the client. Indeed, in our approach the client drives the server to obtain the most suitable QoS.

It should be noted that adaptation to network conditions is not the subject of our study. We focus on the parameters that are related to the terminal itself. If the quality of the video is degrading because of congestion in the route down to the terminal, other end to end techniques should intervene to adapt the media accordingly. Real time control protocol RTCP [40] is one of the possible solutions to control the flow using the

periodic reports. These reports are sent back to give the server an idea about the available bandwidth along the route to the terminal.

1.13.1. Related Work

In this section we give a description of the most relevant standards that treat media negotiation.

1.13.1.1. Session Description Protocol new generation (SDPng)

SDPng [49] is a description protocol for multimedia sessions. It is an application-independent framework transported by other signaling protocols such as SAP (session announcement protocol), RTSP, SIP…etc. the main innovation of SDPng compared to the conventional SDP [50] is its extensibility by using XML, which gives the possibility to describe the different terminal characteristics and user preferences. SDPng description is an XML document divided into 5 parts: capabilities, definitions, configuration, constraints and session information. We are interested in constraints parts, in order to accomplish multimedia adaptation. The constraints section allows expressing constraints on combination of terminal configuration. This feature is intended for specialized devices with strict limitations. To be noted that SDPng base specification is only a container for constraints and does not define them.

1.13.1.2. MPEG-21

Digital Items are defined as structured digital objects, including standard representation, identification and metadata. They constitute the fundamental unit of distribution and transaction within the MPEG-21 [51] framework.

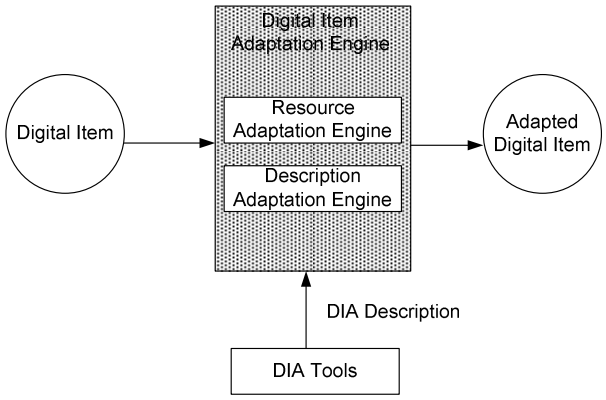

Figure 16: Concept of MPEG-21 DIA

In its seventh part, MPEG-21 defines a number of descriptors and tools to assist the adaptation of Digital Items (see the concept of DIA in Figure 16). Descriptors try to express on the one hand terminal capabilities, network descriptors, user characteristics and natural environment characteristics referred to as Usage Environment Descriptions (UED). On the other hand, it describes the high-level structure of a bitstream referred to as Bitstream Syntax Descriptions (BSD). It is important to underline here that only tools used to guide the adaptation engine are specified by the standard, as for the adaptation engines themselves are left open to various implementations. In order to be independent and open for novel developments, MPEG-21 does not specify any relationship with existing transport mechanisms.

1.13.1.3. Integration of MPEG-21 DIA in SDPng

Guenkova et al [52] propose a harmonization between MPEG21 DIA and SDPng by embedding MPEG21 DIA into SDPng. The proposed mechanism allows the definition of system configurations, performance constraints and adaptation information within the scope of SDPng using the format of MPEG-21 DIA. Furthermore, the converged format enables the integration of session management and negotiation protocols that use such enhanced SDPng descriptions within an MPEG-21 compliant environment. Since the two standards are XML-based, this combination can take place easily by integrating MPEG-21 DIA namespaces to SDPng document. This idea came from the fact that SDPng currently specifies only a container of terminal characteristics and MPEG-21 is not matched to any transport mechanism.

1.13.2. QoS Management

We propose in this work a scheme using SDPng/MPEG-21 in the context of mobile multimedia applications that works under the client-server paradigm. Here, SDPng and MPEG-21 are employed to specify the QoS requirements exposed by the client application at different stages of the session's life, specifically, QoS negotiation and renegotiation stages. Negotiation is the first operation that takes place before streaming starts. The purpose of this operation is to inform the server about the characteristics of a given device in order to serve an adequate coded stream. As far as renegotiation is concerned, we consider two cases where it can take place. The first one

occurs when a session moves from one terminal to another (i.e. session mobility) which has different capabilities than the first one. The second one arises when a "stable" session taking place over a given terminal suffers from an unexpected deficiency (battery level is low, system resources overload…etc.). In both cases, the application has to adapt its behavior according to the new circumstances in order to maintain its operation at an acceptable QoS level. In classical approaches, the server decides about the adaptation procedures to be used, relying on the information reported by the client. This information essentially describes the instantaneous network state in terms of the end-to-end delay and the packet loss ratio. One should notice that the client does not directly participate with the server to take decisions about the adaptation process because it only reports its perception about the network QoS. Indeed, the server adapts its behavior under a "best effort" scheme hoping that adaptation will be the best choice for the client. Our proposition contrasts with classical approaches because it gives a more active role to the client.

1.13.3. Specification of Qos Aspects for Session Mobility

In distributed multimedia applications, QoS management is implemented to provide the final user with acceptable service. A very important aspect of QoS management is QoS specification, which should be conveyed to the server. This specification is composed of a set of parameters designated to describe accurately the QoS requirements of a distributed application. As we stated above, the tools that we use in our work are SDPng and MPEG-21. These standards take into account the new generation of applications that operate in a highly heterogeneous mobile context, but at this time they are not completely defined. Here are the main interesting parameters of which degradation can seriously damage the running session:

- Display properties: it includes colors properties and resolution of the frame. Resolution can be determined by the size of the window displaying the video not necessary the full screen resolution.
- Battery level: when the battery reaches critical levels, power consumption has to be reduced, i.e. the server can eliminate the video stream and keep only the audio.

- Memory space: in case of overload in the terminal, memory space becomes insufficient to run all applications.
- CPU utilization: in case the terminal is running other applications simultaneously with the media session, the CPU utilization is increased as it is shared among all applications and consequently causes a degradation of the perceived quality of the stream.
- Network interface bit rate: changing access network in ubiquitous mobile environment is a potential operation. With cross layer protocols, the application can be informed about an imminent handover as well as the new access network characteristics.

Figure 17 gives an example of MPEG21-DIA integration in the SDPng document with display and network constraints of the terminal.

```
<constraints>
<constraint name="session" xsi:type="sdpng-dia:mpeg21DIA-constraint">
   <sdpng-dia:MPEG21-DIA xsi:type="m21-dia:TerminalsType">
      <m21-dia:Terminal>
       <m21-dia:TerminalCapability xsi:type="m21-dia:DisplaysType">
         <m21-dia:Display>
           <m21-dia:DisplayCapability
            xsi:type="m21-dia:DisplayCapabilityType"
            colorCapable="true"
            contrastRatio="700" refreshRate="30">
              <m21-dia:Mode>
                <m21-dia:Resolution
                  horizontal="176" vertical="144"/>
              </m21-dia:Mode>
              <m21-dia:ColorBitDepth blue="8" green="8" red="8"/>
              <m21-dia:CharacterSetCode>
              US-ASCII
             </m21-dia:CharacterSetCode>
           </m21-dia:DisplayCapability>
         </m21-dia:Display>
       </m21-dia:TerminalCapability>
      </m21-dia:Terminal>
   </sdpng-dia:MPEG21-DIA>
</constraint>
<constraint name="component" ref="videocomponent001"
   xsi:type="sdpng-dia:mpeg21DIA-constraint">
  <sdpng-dia:MPEG21-DIA xsi:type="m21-dia:NetworksType">
   <m21-dia:Network>
    <m21-dia:NetworkCharacteristic maxCapacity="384000"
      minGuaranteed="32000"
      xsi:type="m21-dia:NetworkCapabilityType"/>
    <m21-dia:NetworkCharacteristic
      xsi:type="m21-dia:NetworkConditionType">
        <m21-dia:AvailableBandwidth average="80000"
          maximum="256000" minimum="330"/>
        <m21-dia:Delay delayVariation="66" packetTwoWay="330"/>
        <m21-dia:Error packetLossRate="0.05"/>
    </m21-dia:NetworkCharacteristic>
```

```
  </m21-dia:Network>
 </sdpng-dia:MPEG21-DIA>
</constraint>
</connstraints>
```

Figure 17: Example of SDPng file [52]

1.13.4. Negotiation

Negotiation takes place at the beginning of the session. During this phase the client communicates with the server to exchange necessary information about the required service and its characteristics. This negotiation is supporter by the RTSP protocol, which actually conveys the SDPng information embedded inside the RTSP messages.

1.13.4.1. Session Establishment

This stage of the session is achieved by exchanging RTSP messages in particular DESCRIBE and SETUP messages. Here, the client asks for a given service by means of DESCRIBE method. The server answers with a message containing the description of the service as well as the resources required to obtain it. When the client decides to take the service, it sends a SETUP message containing an SDPng payload which describes the client's characteristics to be considered during the session adaptation, this description is expressed in MPEG-21 syntax and contained in constraints part of SDPng file.

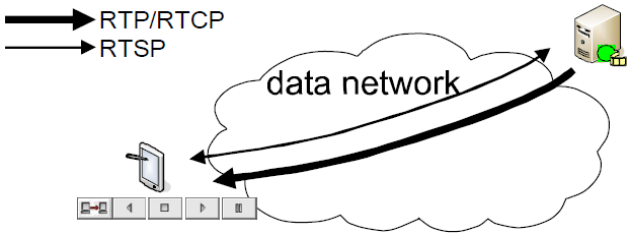

Figure 18: Classical QoS negotiation procedure

1.13.4.2. Session Remote Control

Session remote control is another case where a session needs to be negotiated. To clarify this concept, we consider a client who is watching a video of his favorite game in a PDA. Afterwards, he desires to record it with a better quality on his hard disk placed in his home to watch it later (see Figure 19).

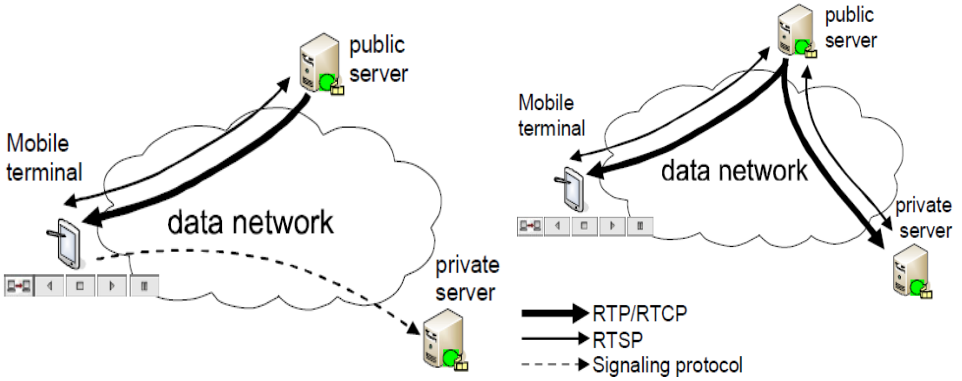

Figure 19: Remote control QoS negotiation

SESSAMO protocol can also be used in this case even if this is not a real session mobility scenario. The characteristics are specified from the PDA and sent via the SESSAMO request. These parameters are retrieved by the set-top-box and used to negotiate the session with the server. Moreover, the set-top-box uses RECORD method in establishing the new session with the server offering the media stream. In this case, the QoS negotiation can be started with better quality, via a wired network and without real-time presentation constraints, because it will be played later locally.

1.13.5. Renegotiation

1.13.5.1. Session Mobility

When a session moves from one terminal to another with different characteristics, a renegotiation process is required (see Figure 20).

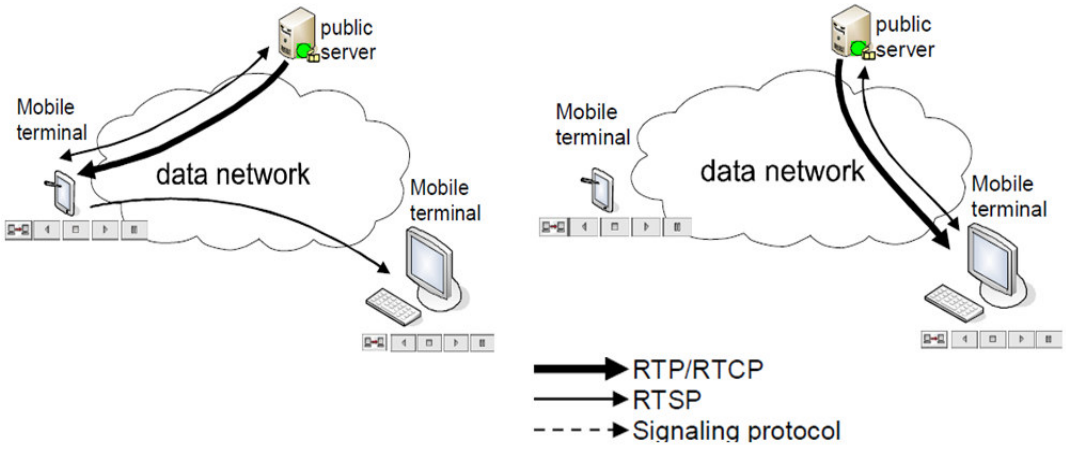

Figure 20: Session mobility with QoS renegotiation

Renegotiation of the new parameters is achieved by the new terminal itself. It specifies its constraints in the SDPng message and sends it to the server. In response, the server adapts the media flow according to the new constraints and transmits it to the new terminal.

1.13.5.2. Parameter Change

The idea of QoS renegotiation within multimedia session is based on the exchange of RTSP messages containing information about the current status of the terminal capabilities. SET_PARAMETER is the RTSP message that we choose to conclude this task, because it is used to convey parameters related to the operation of the received service. SET_PARAMETER method is used to report QoS indicators within the current session in an SDPng like message from the client to the server as showed in Figure 21. Using this information, the server can decide whether an adaptation operation is required. It should be noted that during the session only parameters that has been changed are concerned by the notification.

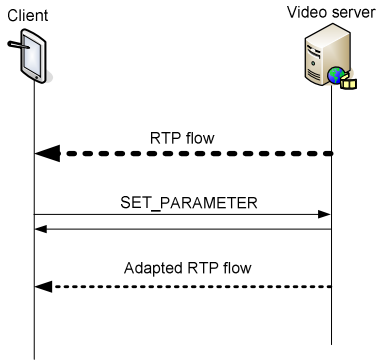

Figure 21: Message flow of renegotiation process

1.13.6. Adaptation to Network Conditions

The adaptation to network fluctuation does not need any renegotiation because the network is the cause of the problem not the terminal. As far as network communications are concerned, two approaches have been proposed to solve this problem. The first one proposes to enhance the network infrastructure by introducing resource reservation mechanisms. The second approach proposes to adapt the application to the available network resources. In the global Internet, adaptation of the flow from the application seems to be a more realistic solution. In order to implement the adaptation capability,

the application should support some additional functionality. From a network perspective, a periodic feedback containing the current reception status of the flow allows the server to adjust the bit rate accordingly. A standardized protocol named RTCP (Real Time Control Protocol) is currently used to perform this task. The reported network state information allows the server to reduce or increase the bit rate in order to alleviate the losses in the sent flow. Of course, the losses rate will be reduced but the QoS perceived by the user will also be degraded, but in a controlled manner. There are several techniques allowing a server to adapt the flow to the available network resources, such as quantization, re-quantization [53], transcoding [54] [55], frame dropping, multilayer encoding [56]…etc.

1.14. Acknowledgment

The contribution presented in this chapter is proposed to solve one of the problems posed by the European project SUMO-ITEA [57]. The primary target of SUMO (Service Ubiquity in Mobile and Wireless Realm) is providing service continuity for rich media terminals. Practical solutions have been proposed to solve service continuity for both terminal mobility and session mobility. SESSAMO solution was one of the adopted solutions for session mobility, and a demo was presented during the final review.

1.15. Conclusion

Session mobility is an optional service that can be offered by the operator or even by content provider to their customers in order to give them more flexibility and portability regarding their interaction with the served media. In this chapter, we proposed a new solution to support session mobility in video streaming services. SESSAMO is a lightweight protocol that operates between the concerned terminals and conveys the needed information for the target terminal to resume the session. SESSAMO has been implemented in a real life scenario using commercial mobile device. Moreover performance of the proposed solution has been conducted and the results show that this solution makes efficient and fast handovers. Indeed, achieving session transfer is a challenge in itself, but it has some “side effects” that should be

treated as well. In fact when transporting the session from one terminal to another it is more likely that the latter has different hardware and software capabilities. Therefore, some measures have to be taken in order to adapt the served stream to the current terminal capacities. In this context, we proposed a mechanism to negotiate and renegotiate QoS parameters of the session by using SDPng and MPEG-21. Here, SDPng and MPEG-21 are employed to specify the constraints of the client at different stages of the session's life. This description is integrated into the payload of specific RTSP messages following the required operation.

ANALYSIS OF MOBILITY MANAGEMENT PROTOCOLS OVER IP

1.16. Introduction

Many mobile devices are currently using IP based networks to access a big variety of applications including those needing insurance of holding the session when moving from one network to another. Mobility issue in IP networks comes from the fact that this protocol was not originally designed to handle mobility. Indeed, Internet protocols are not suitable for supporting mobile communications because of their principles of addressing and routing. Any host address must be derived from the network address where it is physically attached and no change in this address during the session is considered. Under such scheme, when a MN moves from its original network to a Foreign Network, it will experience at least the following problems: 1) when it reaches a new network, any communication becomes impossible. Given that its address is not valid in the foreign network, it can not be accepted neither by foreign nodes nor corresponding routers. Obtaining a new valid address from the foreign network is then necessary. 2) The ongoing communication associations are lost due to address inconsistency. 3) Mobile node disappears from the global network. Normally, hosts are found in the network by means of Location Directories (LD). It is a distributed database containing the host name and its corresponding IP address, an example of this database is the well known internet service DNS (Domain Name System). To keep in touch with the global net, MN must inform the LD each time it acquires a new IP address.

In order to cope with IP limitations in mobile communications, a number of approaches have been proposed. Although they tackle the problem from different perspectives, they agree on the way they handle it. Indeed, the main approaches rely on a number of procedures that can be classified into: **movement detection and address**

allocation, **traffic redirection**, **global location tracking update** and **handover smoothing** (see Figure 22). These four procedures and the problems they address are analysed in more detail in the following paragraphs.

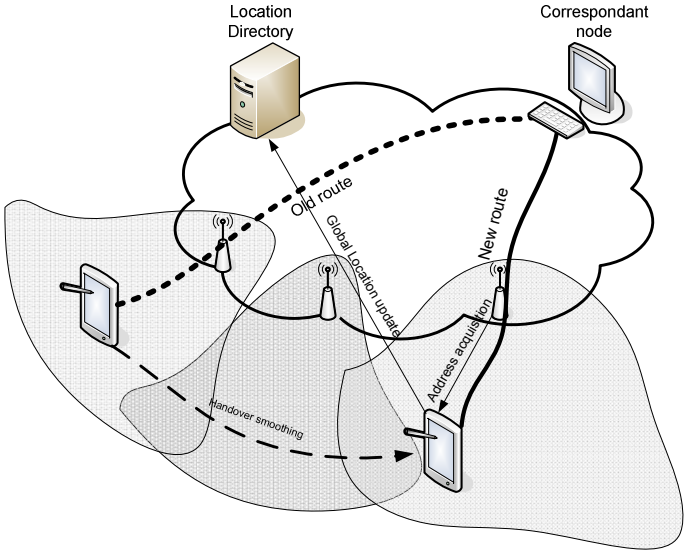

Figure 22: Mobility procedure

1.17. Movement Detection and Address Allocation

First of all, a mobile node needs to detect its movement among different networks. The discovery of a new network can be achieved by means of layer 2 or layer 3 mechanisms. The accuracy and the speed of this discovery vary from one scheme to another. Layer 2 techniques are known to be more reactive than layer 3 ones, though they add more complexity to the MN's system. For example, mobility management protocol can use quality of the signal sensed on the wireless interface as an indication of leaving one network area and entering a new one. Therefore, MN can prepare the handover as soon as it receives the movement indication. At the opposite, layer 3 techniques rely on receiving special advertisement (e.g. FA and HA advertisements in MIP) or wait for the address configuration to realize that he network to which it was attached has been changed. Address allocation can take place in many ways. The most common method is dynamic host configuration protocol (DHCP), either in IPv4 or IPv6. This method necessitates the existence of a DHCP server which distributes and manages addresses. A stateless address configuration method consists of constructing

the address without the existence of any server. After discovering the access router prefix by means of neighbour discovery protocol [58], MN concatenates the prefix with its interface address which is driven from its MAC address. MN can have several addresses under different appellations depending on its function and the authority that allocated it. In Mobile IP for example, the address allocated in a foreign network is called Care of address where the original one is called home address. Movement detection and address allocation are put together in one operation because they are tightly related. In other words, movement detection can result from the allocation of a new IP address which belongs to a different network than the original one. Inversely, discovering a new network by means of layer 2 mechanisms for example triggers the allocation of a new address.

1.18. Traffic Redirection

This phase consists of redirecting packets destined to the MN's old address to its new location. This operation is needed for maintaining the ongoing communication because of the hierarchy nature of addressing in internet and routing protocols. In general, routing tables in intermediary routers are built based on the network address not on the host address. The reason behind this choice is minimizing routing table size and consequently making routing decision as fast as possible. Traffic redirection can be guaranteed by the network or the end points; in other words, re-routing packets to the new location can be done with or without the intervention of the communicating nodes.

1.18.1. Network Based Traffic Redirection

This method is transparent for the application in MN and its CNs. Designed network entities are in charge of gathering packets addressed to the old MN's address and resend them to the new address. Usually tunnelling is used to carry out this operation without altering the original IP packets. MN should inform the redirection network entity about its current location permanently. For example, home agent plays this role in MIPv4, it intercepts packets and encapsulates them to the foreign agent.

1.18.2. End Point Based Traffic Redirection

In this scheme, end points are involved in the redirection of the traffic. At the opposite of the previous method, here the application of both MN and CN has to cope with IP address change. Therefore MN should inform all its CNs about the new acquired address and CNs in their turn, update the destination address in the sent packets. Route optimization in MIP is an illustration of this type of redirection. Indeed, end point traffic redirection is more efficient than network based one, because of the non optimized route taken by the packets (i.e. they reach the home network first before being redirected by the home agent).

1.19. Global Location Tracking Update

This operation is achieved to maintain the reachability of the MN at global network level. In fact, any node in the global network is reachable by its address. Hence any change of this address should be known to all nodes willing to communicate with this node, not only those who are already in communication with it. Practically MN can not inform all nodes in the global network about its location by itself, but instead, a designated server holds the current MN IP address. DNS server is an example of such server; it maintains mapping between the fully qualified domain name (FQDN) of a given host and its IP address. Home agent in MIP plays also this role by mapping CoA to home address.

1.20. Handover Smoothing

This is an optional operation that aims to make the disruption caused by the handover process less perceptible by the user. Indeed, during steps 1 and 2 the MN is not receiving any packet because it does not have any valid address and the traffic is not redirected until the concerned entities are updated. Those packets are lost because they are sent to the old address. An improvement of user experience during the handover consists of preserving these packets from loss and relaying them to the client in the new location. Two solutions have been proposed in the literature to perform handoff smoothing. The first one is tunnelling as it is used in low latency handoffs in mobile IPv4 [59] and FMIPv6 for example. This technique proposes to establish a tunnel between the old and new network, then forward the packet through the tunnel during a

period of time long enough to perform the needed updates. The second one is bicasting [61][62]which consists of duplicating the flow sent from the CN to both old and new addresses. Hence, at least one instance of the packet will reach the MN wherever it is located. Of course, each solution has its advantages and disadvantages in terms of overhead and resource utilization but it should be noted that this operation is limited in time and covers only the handover period.

The above operations are the most important operations in mobility management protocols over IP. We notice here that not all of them are implemented by mobility protocols nor they are necessarily executed in that order. Some of them are more critical than the others. For example, from the ongoing communication point of view, the more important operation after acquiring a new address is handover smoothing followed by traffic redirection. Whereas the global location tracking update is not critical since it is important for future communications.

1.21. Case Study

In this section, we take the example of two mobility management protocols from different perspectives and study them following the analysis done previously. We show through this case study that the most representative approaches for handling node mobility (i.e. Mobile IP and SIP) follow the steps stated earlier. Afterwards we draw a summary overview of the most relevant mobility protocols regarding the same analysis.

1.21.1. Network Layer Perspective: Mobile IP

The main goal of Mobile IP is to avoid upper layers being worried about address changing due to node mobility.

1.21.1.1. Movement Detection and Address Allocation

CoA in MIPv4 is attributed by the new FA, which periodically broadcasts a Router Advertisement message containing CoA related information. MN receives an advertisement when it joins the new network and proceeds to the registration with the FA in order to obtain a CoA. The speed of movement detection depends on the periodicity of the advertisements. In other words, the more frequent the advertisement is broadcasted, the faster movement is detected and consequently faster the handover is

performed. However, this improvement in handover performance comes with a cost which is an increasing in the signalling overhead of MIPv4. This drawback is avoided in MIPv6 by introducing router solicitation procedure. In this version of MIP, as soon as MN joins a new network it requests explicitly a solicitation rather than waiting for the advertisement. As a response to the solicitation, the access router sends a router advertisement immediately.

1.21.1.2. Traffic Redirection

In MIP, traffic can be redirected either using network based traffic redirection method or end point based traffic redirection method.

Network based method:

Home agent is the pillar entity that performs traffic redirection, since all traffic destined to MNs are routed through it. Ordinary operation mode of MIP carried out using tunnelling mechanism between HA and FA in case of FA-Coa or between HA and MN in case of co-located CoA. Indeed packets sent from the CN are destined to the home address, while MN is not in its home network, packets are intercepted by the HA using layer 2 mechanisms (i.e proxy arp). Afterwards, those packets are encapsulated in a new IP header which has HA address as source address and CoA as destination address. In case of FA-CoA the tunnel is established between the HA and FA, whereas it ends at MN in case of co-located CoA. This method has two major drawbacks: the first one is that the HA is a single point of congestion from where all traffic should be encapsulated and routed. This operation is CPU consuming and requires high performance hardware especially when the number of MN is high. The second negative point is that routing is not optimal since packets do not follow the normal route towards the destination but instead they are routed first to the HA and then from the HA to the MN. This operation is known as triangular routing.

End point based method:

In order to overcome the drawbacks of network based redirection method, IETF has proposed an optimization to MIP operation called route optimization. In this mechanism MN does not rely on the network (i.e. HA and FA) to redirect its packets from the old destination to the new one, but performs the necessary updates directly with its CNs. In other words, the tunnel is established between the MN and its CNs

without any intervention from the HA. Consequently, all packets sent from the CN will be routed by the intermediary routers directly to the new destination assigned by the CN in the outer IP packet. This optimization comes with a cost which is the involvement of CNs in the MIP protocol whereas in the previous method they were not involved.

1.21.1.3. Maintaining The Global Location Tracking

HA is responsible for maintaining the global location tracking by making a mapping between the home address and the new acquired CoA. Each time the MN attaches to a new network it should register with the HA via the FA (or directly), in this special moment the HA maps the home address to the new CoA. Therefore, any third mobile willing to communicate with the MN can always use the original home address to reach it.

1.21.1.4. Handover Smoothing

When the MN is performing layer 2 handover and movement detection along with CoA acquisition and HA registration, MN disappears from the network because it is no longer in the home network nor the HA is aware of its movement. Hence, packets addressed to MN in this period will be lost which leads to noticeable degradation of the user's quality of experience. To overcome this discontinuity in the service, fast handover scheme was introduced by Low-Latency Handoffs in Mobile IPv4 and FMIPv6. The goal of this protocol is to prevent packet loss during the aforementioned steps where the MN is not reachable. Packets are tunnelled between old FA and new FA or between old access router and new access router in IPv4 and IPv6 respectively. As soon as the MN attaches to the new network, packets are forwarded right away to the MN. The tunnel is maintained long enough so that MN can achieve its updates with HA/CN.

1.21.2. Application Layer Perspective: SIP

Handling mobility at network layer requires considerable changes in the MN kernel and it concerns all applications built on top of IP layer. Another approach consists of providing mobility support only for targeted applications like VoIP. SIP is one of the famous signalling protocols used in internet that handles terminal mobility for SIP call sessions.

1.21.2.1. Movement Detection and Address Allocation

SIP does not provide any special movement detection technique, for this purpose mobile User Agent (UA) relies on the conventional DHCP procedure to acquire a new IP address. Although this TCP/IP-based protocol was not designed to operate in mobile contexts, it is widely employed to support address allocation in access networks. DHCP satisfies most of non real-time applications but it appears to be unsuitable when it deals with real-time ones. The main problem here is related to the number of packets and the long delay that DHCP takes for address allocation. This latter is mainly caused by the address conflict checking mechanism based on ICMP Echo request/reply, this operation is commonly known as duplicate address detection (DAD). A DHCP server sends out an ICMP Echo request to the address in question before responding to Discover message. If nobody responds with an ICMP Echo reply within certain interval of time, the DHCP server will answer with Offer message. There are some proposals to reduce the number of packets from four to only two [63] and others suggest to remove DAD. Finally, Dynamic Registration and Configuration Protocol (DRCP) [64] is proposed to replace the conventional DHCP to meet requirements of real time applications.

1.21.2.2. Traffic Redirection

The procedure allowing redirecting the traffic from old to new address is known as mid-call mobility procedure in the context of SIP. The principle is the following: when the MN reaches a new network and a new address has been acquired, the MN sends a re-INVITE request to the CN. This operation is accomplished without intervention of any intermediate SIP proxies. INVITE request contains an updated session description with new transport parameters including the new IP address. The CN starts sending data to the MN's new location as soon as it gets the re-INVITE message. Thus, SIP implements the end point based traffic redirection scheme.

1.21.2.3. Global Location Tracking Update

SIP users are identified by a unique URI. Each time the UA joins a new network it should update the SIP registrar by means of register message in order to bind the new acquired IP address with the URI. When future clients want to contact the mobile UA using its URI, they check first (directly or via a SIP proxy) SIP registrar to retrieve the

current IP address mapped with URI and then initiates the communication with the mobile UA.

1.21.2.4. Handover Smoothing

SIP standard does not specify any mechanism for the purpose of handover smoothing. It should be noted that SIP mobility engenders many messages exchange before re-establishing the session which cause a significant handover delay [60]. Hence a handover smoothing mechanism is required in order to provide seamless service continuity. Bicasting the stream [61] towards both destinations (old and new address) is one of the solutions proposed to perform the handover smoothing for SIP.

1.22. Summary

Table 8: summarizing table

Mobility protocol	***SIP***	***MIPv4***	***MIPv6***	***PMIPv6***	***TCP-migrate***	***HIP***
Layer	application	Network	Network	network	transport	transport
Movement detection and Address allocation	DHCP , PDP	FA advertisement (FA-CoA) – DHCP (co-located CoA)	Neighbor discovery, autoconfiguration, DHCPv6	Home network prefix allocation and auto-configuration or DHCP	DHCP	DHCP
Traffic redirection	Re-invite method	Route Optimization, Reverse tunneling	Route Optimization with RR, Reverse tunneling	Proxy binding update	End to end	HIP update
Handoff smoothing	Bicast	FMIPv4	FMIPv6	FMIPv6	-	-
Global tracking agent	SIP registrar	Home agent	Home agent	Local mobility anchor	-	Rendez vous server

Table 8 provides a summary of the most known mobility protocols over IP and classifies their operation in accordance with the above study.

1.23. Conclusion

In this chapter we analyzed the procedure of mobility management protocols over IP and we deduced that most of them follow the same logic in tackling mobility problem. Four main sub operations are identified: movement detection and address allocation, traffic redirection, global location tracking update, and handover smoothing. These sub-operations are not necessarily implemented in every mobility protocol but some of them are mandatory and others are optional. Moreover, cooperation between different protocols even belonging to different layers is possible in order to achieve seamless handovers. We believe that this classification is very useful in analyzing and designing new mobility protocols. To demonstrate the validity of our classification, we took the example of MIP and SIP and we showed from their thorough analysis that they are compliant with the proposed classification. Finally, a summary of other mobility protocols and their compliance with the proposed classification is given.

COLLABORATIVE HANDOVER MECHANISM FOR REAL-TIME SERVICES

1.24. Introduction

Next generation networks intend to offer pervasive and ubiquitous services to the customers wherever they are and whatever the application they are using. To achieve this goal, service continuity across homogenous and heterogeneous networks must be guaranteed. Thanks to mobile IP and its variants, continuity of service in all IP networks can be ensured. MIPv6 was designed by the IETF to provide mobile nodes (MN) with the possibility of maintaining the ongoing communication when changing the access network and MN's IP address. Although MIPv6 presents some improvements compared to MIPv4, its long handover delay makes it unsuitable for real time applications. FMIPv6 tries to reduce this delay by using link layer triggers to perform address acquisition before layer 2 handover. Moreover, FMIPv6 prevents packet loss by creating a tunnel between the Previous Access Router (PAR) and the New Access Router (NAR) to forward packets until the correspondent node and the Home Agent update the Care of Address (CoA). However, FMIPv6 shows some limitations in fast-moving terminals. On the one hand, the terminal has not enough time to exchange all messages with PAR in the initiation phase. On the other hand, access routers require a big buffer size to buffer packets sent to the MN between the moment of FBack transmission and the moment of unsolicited neighbour advertisement (UNA) reception respectively by PAR and NAR.

In this work we use MIH services more efficiently in different steps of FMIPv6, not only in the discovery step. MIH services are solicited by AR each time it cannot ascertain the transmission of a message and learns the MN situation by an event subscription. This scheme increases the probability of FMIPv6 predictive mode by reducing the handover initiation delay. Moreover, by using link event indications the mobile will be served until its physical disconnection, and as soon as it connects to the

new network, packets will be forwarded from the new access router without waiting for attachment announcement from FMIPv6. Therefore buffer size dedicated to buffer tunneled packets to the new MN's location is reduced. Another advantage of this proposal is ping pong effect resiliency. Indeed, handover is initiated by the MN but finalized by the network allowing cancellation of the already initiated handover which was triggered by a "false alarm".

We show through analytic analysis and simulation, that our scheme reduces handover delay as well as packet loss in both predictive and reactive handovers. In addition, buffer size needed to buffer packets in access routers during the handover is smaller. Finally, the proposed mechanism avoids excessive ping pong events among wireless networks.

1.25. FMIPv6 Limitations

The benefit of FMIPv6 resides in its predictive mode. Any failure in performing this mode leads to reactive mode which adds more delay in handover. Therefore, in case of mobiles in fast movement, It is likely that MN will not be connected to the previous network long enough to send and receive all FMIPv6 messages. Thereby, the terminal should accomplish the initiation step in short period of time in order not to loose any FMIPv6 message. Actually solicitation message is sent only after layer 2 trigger which results in a long initiation time. Moreover, if the connection is broken before the reception of the FBack message, all packets will be lost until connecting to the new network and creating the tunnel between NAR and PAR. The other important delay which prolongs service breaking period is between the moment MN receives FBack and the moment it disassociate from the old network, all packets destined to MN during this period are forwarded to NAR and buffered, whereas MN can still receive these packets on the old link. In fact FMIPv6 proposes also to start buffering packets and forward them (i.e. buffer and forward) at the same time upon receiving FBU even before creating the tunnel. This solution is good to ensure that all packets addressed to MN during FMIPv6 operation will be received either in the old or the new location. In fact this solution can be considered as a bicast solution addressed in chapter 4 where the traffic is sent to both old and new addresses. Nevertheless this solution has at least two

drawbacks: it increases the overhead and needs a mechanism in MN operating system to avoid packet duplication in case it takes place.

1.26. Related Work

Some works proposed enhancements for FMIPv6 to cope with real time applications for fast-moving terminals. In [65] and [66], MIH is used to reduce access router discovery by using media independent information server to retrieve necessary information of neighbouring networks without using RtSolPr and PrRtadv messages. But the utilization of MIH is limited to discovery step. [67] and [68] propose schemes to reduce the effect of duplicate address detection (DAD) performed by NAR before sending Hack to PAR. As for [69], it proposes to use a disassociation message sent to PAR so that it will start buffering from this moment. However, if the signal becomes too weak, this message has no chance to reach the PAR.

Ping pong effect:

Usually Link_Going_Down threshold is higher than Link_Down level of the received signal. Due to unpredictable mobility of the user and variability of the signal strength because of multipath and fast fading [70], an erroneous movement detection can occur; hence probability of having a ping pong event increases. Erroneous movement detection or ping pong has negative consequences on the quality of the perceived stream since many handovers take place in close periods of time. Some works try to reduce ping pong effect by acting on the handover triggering mechanism itself. For example, authors of [71] try to avoid unnecessary vertical handover by determining the appropriate time at which handover should be initiated.

In our proposition ping pong is avoided in a different way. Our scheme does not rely on signal strength at certain period of time to decide whether to make or not the handover, but the handover is finalized by the network only if the signal is down. Whenever the signal comes up again after a link going down notification, handover operation is cancelled. Our scheme is independent from the triggering algorithm and allows avoiding ping pong events.

1.27. Media Independent Handover Overview

IEEE 802.21 [72] or Media independent handover is a new standard that has been approved by IEEE in early 2009. It provides link-layer intelligence and other network information to upper layers in order to perform optimized handovers between heterogeneous networks. Network technologies covered by the standard are both IEEE and non IEEE networks like 3GPP networks. The purpose of this standard is to enhance the experience of mobile users by facilitating handovers between heterogeneous networks.

1.27.1. MIH Services

MIH function is the corner stone of the MIH framework. It is an intermediate logical entity between link layers and upper layers (see Figure 23). On the one hand, MIHF is connected to upper layers such as mobility management protocols via service access points called MIH_SAP. On the other hand, MIHF communicates with low layers using media specific service access points called MIH_LINK_SAP. MIHF reports network interface events to upper layers and conveys the resulting commands down to the proper interface. Therefore, upper layer does not deal with the heterogeneity of network interfaces when performing vertical handovers; in stead, they have a unified interface that handles the different network interfaces. Communication between the different components is performed via media independent services.

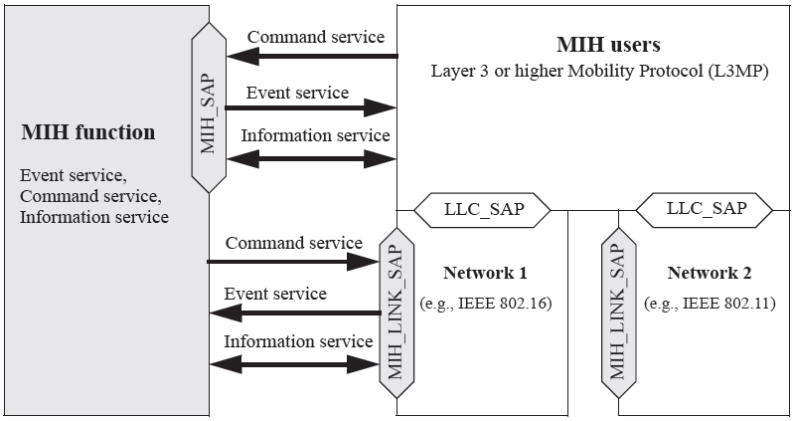

Figure 23: MIH function

Media independent event service (MIES) indicates any change in link characteristics, link status or link quality on local or remote entity. Link-layer related events are reported to MIHF which relays them to upper layers (MIH user) in order to take the adequate decision. Events such as LINK_UP, LINK_DOWN, LINK_GOING_DOWN…etc, play an important role in achieving seamless handover. To be noted that amendments to the underlying physical layer are required in order to define primitives of such events. The complete list of link events is given in Table 9.

Table 9: list of media independent link events

Link event name	**Link event type**	**description**
Link_Detected	State change	Detection of a new access network
Link_Up	State change	L2 connection is established and link is available for use.
Link_Down	State change	L2 connection is broken and link is not available for use.
Link_Going_Down	Predictive	Link conditions are degrading and connection loss is imminent
Link_Parameters_Report	Link parameters	Link parameter has crossed the pre-specified threshold
Link_Handover_Imminent	Link handover	L2 handover is imminent
Link_Handover_Complete	Link handover	L2 handover to a new PoA is completed

Media independent command service (MICS) enables higher layers to control lower layers (physical and data link layers). This service carries the decisions taken by upper layers to lower layers on local or remote entity. For example command service can be used by the decision engine to switch the connection from one network interface to another. The list of link commands is given in Table 10.

Table 10**:** list of media independent link commands

Link command	**Description**
Link_Capability_Discover	Query and discover the supported events and commands on the link layer
Link_Event_Subscribe	Subscribe to one or more events form a link
Link_Event_Unsubscribe	Unsubscribe from a set of link layer events
Link_Get_Parameters	Get parameters measured by active link such as SNR, BER, and RSSI
Link_Configure_Threshold	Configure threshold for link report event
Link_Action	Request an action on a link layer connection (ex: connect, disconnect)

Media independent information (MIIS) provides a framework by which MIHF can discover and obtain information about the networks in specific geographical area to facilitate handovers. The information element is provided following a query/response mechanism from information server or can be learned locally. MIIS typically provides static link-layer parameters such as channel information and MAC address of the PoA. This information is used later by MIH users to optimize handover to neighbouring networks.

1.27.2. Transport Protocol

MIHF communicates with its peers for various purposes. MIHF in any network entity becomes a point of service (PoS) when it communicates with local MIHF. This communication may concern information related to different MIHF services (MIIS, MIES, or MICS). MIH messages are transported between remote entities by means of L2 or L3 mechanisms. Nevertheless, the standard does not specify any transport protocol and leaves this specification to implementation. In [73], the authors propose to use UDP as transport protocol for MIH messages because of its speed and simplicity. Reliability can be provided by the application using retransmission algorithms.

1.28. Triggering Mechanisms

1.28.1. Horizontal Handover

Most of predictive algorithms are based on signal strength for handover triggering. There are different algorithms that use the received signal strength information in order to predict whether the link is going down or not. The most known methods are listed below.

- Threshold: in this method RSSI (received signal strength information) is monitored continually, when this parameter goes below a predefined value the handover is triggered.
- Hysteresis: this method uses two kinds of thresholds: entry threshold and exit threshold. Entry threshold is the minimum value of RSSI to enter a particular network, and exit threshold is the value at which the handover is triggered. This scheme reduces the oscillations between the same networks or ping pong events.

- Trend and Least mean square methods [74]: theses mechanisms are based on the history of the RSSI and try to predict the trend of this latter in the near future.

1.28.2. Vertical Handover

The above algorithms can not be applied in case of heterogeneous networks since RSSI value follows different standards depending on outdoor or indoor networks. Moreover, other criterions might intervene in order to determine the best network to connect to, such as offered bandwidth, cost and user preferences. From a list of candidate networks, MN should select the best network suitable for the expected QoS and user preferences. Selection algorithm or decision algorithms use different techniques such as Multiple Attribute Decision Making (MADM) and fuzzy logic multi-criteria functions.

1.29. Collaborative Handover Procedure

FMIPv6 intends to protect data packets from loss. Nevertheless, sending FBU message after link going down event will prevent MN from receiving any packet because they are encapsulated to NAR even if MN is still able to receive packets. Therefore, we need to trigger the tunnel creation by another mechanism which is separated from FBU transmission. In this work, we propose a collaborative scheme that involves both the terminal and the network. AR and PoA use MIH services to make it possible for the terminal to continue its communication in the previous network even after link going down notification. The actual link down is the only event that should trigger packet forwarding via the tunnel. Such event is also detectable at the PoA which will notify the PAR to start forwarding packets through the tunnel. Preserving packets of the ongoing session is the first priority that we intend to achieve before any other procedures.

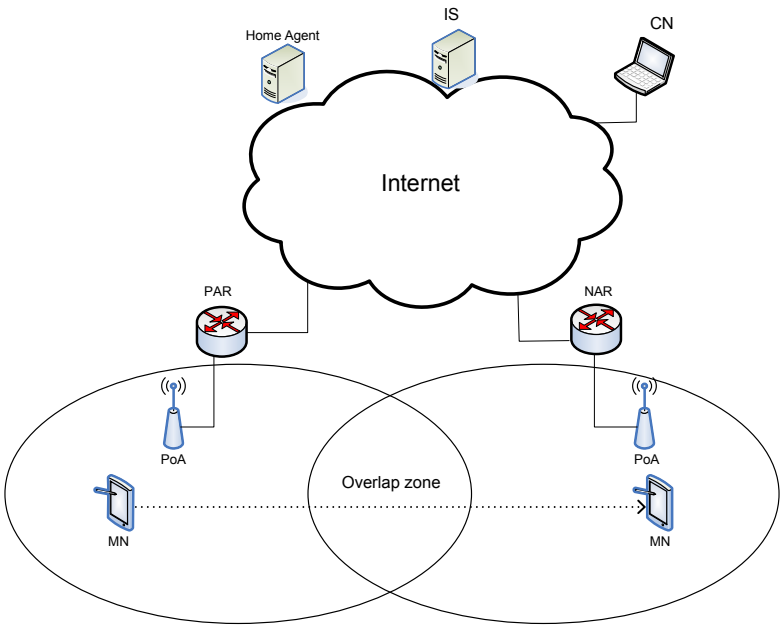

Figure 24: Mobility scenario

We consider the mobility scenario shown in Figure 24. MN is connected to the serving network and has connection to the information server. Information server location can be provided by a DHCP server when responding to MN's request at its first connection to the network.

MIH function starts with initialization operation to discover network entities that provide MIH services and its capabilities, then it registers with other MIHF and subscribes to particular set of events either locally or remotely. MN asks information server about neighboring networks by sending MIH_Get_Information request (see Figure 25). The response contains (AP-ID, AR-Info) couples of neighboring access routers. Thus MN knows about its neighboring networks a long time before the handover, not after handover trigger as it is the case in classical FMIPv6. Once the signal becomes weak, MIH layer is notified by Link_going_down event, then proceeds to the execution of the decision algorithm to decide the network to handover to, based on the already available information server response.

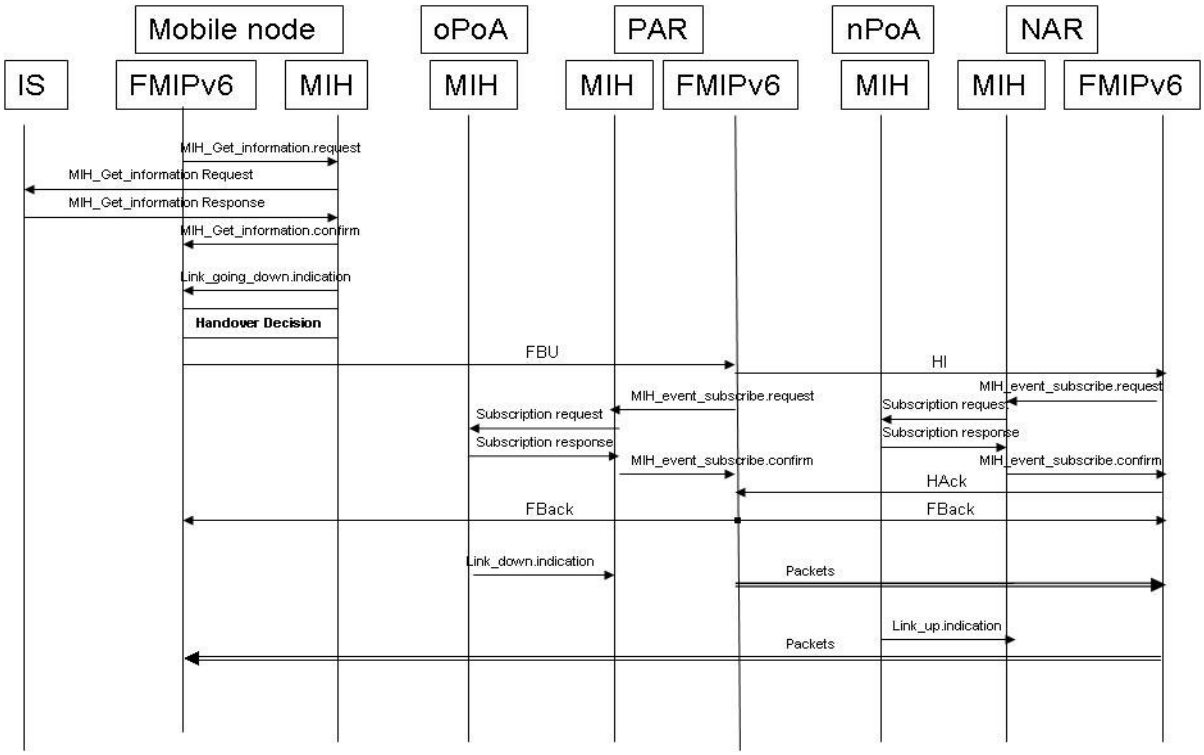

Figure 25: MIH-FMIPV6 message exchange

Decision algorithm is executed by decision engine which takes into account QoS constraints of the application and user preferences. The MIH user (FMIPv6 here) uses the result of this algorithm to send FBU message which contains the NCoA obtained by stateless address configuration using target network prefix. Upon receiving FBU message, PAR sends HI to NAR and MIH_event_subscribe to oPoA's MIHF to subscribe to Link_down_event of MN.

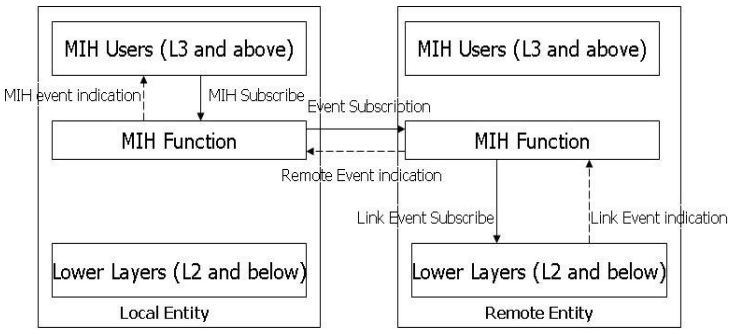

Figure 26: Remote MIH subscription and indication

Figure 26 explains subscription and event indication procedures in both local and remote entity. FMIPv6, should include the targeted MN's MAC address in the MIH subscribe message using the source address of FBU message. Thus, MIHF in oPoA can filter link down event of this specific MN. Link down event can be notified by the PoA when L2 connectivity is lost either explicitly by dissociation procedure or after successive acknowledgement time outs. After receiving HI, NAR can perform DAD for prospective CoA, and in the same time, it subscribes to link_up event of MN in nPoA. After L2 handover, MN connects to the new network by attaching to nPoA which notifies this event to NAR, consequently, NAR starts forwarding packets to MN without waiting for UNA message.

The main difference with classical FMIPv6 is that PAR starts redirecting packets only when it receives Link_down_event notification from oPoA; in the other side, NAR forwards packets as soon as MN connects to the new network.

1.30. Performance Evaluation

We assume that the processing time in the different entities is null, and message transmission delay between MIH user and MIHF is neglected when this message is locally exchanged.

1.30.1. Handover Latency

FMIPv6 handover initiation delay D_{init} is composed of router solicitation and advertisement delay D_{adv} and fast binding update message exchange delay D_{fbu} which can include DAD operation delay if it is executed by NAR before accepting NCoA. In the case of classical FMIPv6 we have:

$$D_{init} = D_{adv} + D_{fbu}$$

Whereas in MIH-FMIPv6, there is no network discovery after handover triggering since this step is done earlier through the information server, so

$$D_{init} = D_{fbu}$$

Reducing the handover initiation time increases the probability of performing predictive mode especially in case of fast-moving terminal and restricted overlap areas.

The handover latency D_{ho} is the period of time when the mobile is not receiving any packet. In case of FMIPv6 this delay starts from the transmission of FBack by the PAR until reception of forwarded packets from NAR.

$$D_{ho} = D_{pre} + D_{L2} + 2D_{MN\text{-}NAR}$$

Where $D_{MN\text{-}NAR}$ is the delay between MN and NAR, this delay is composed of the wireless link delay $D_{MN\text{-}PoA}$ and wired link delay $D_{PoA\text{-}AR}$ between the PoA and AR. We assume that:

$$D_{MN\text{-}AR} = D_{MN\text{-}PoA} + D_{PoA\text{-}AR}$$

D_{pre} is the delay preceding L2 handover during which MN is not receiving any packets. In MIH-FMIPv6, D_{pre} is equal to $D_{oPoA\text{-}PAR}$ the delay that link_down.indication takes between oPoA and PAR. Moreover, the packets are forwarded to MN upon reception of link_up.indication by NAR. Hence the handover latency is equal to:

$$D_{ho} = D_{oPoA\text{-}PAR} + D_{L2} + D_{nPoA\text{-}NAR} + D_{MN\text{-}NAR}$$

To give a concrete example of the proposed mechanism effect, we use the following numerical values:

$D_{oPoA\text{-}PAR}$=$D_{nPoA\text{-}NAR}$=5ms, D_{L2}=20ms, $D_{MN\text{-}AR}$=$D_{MN\text{-}NAR}$=$D_{MN\text{-}PAR}$.

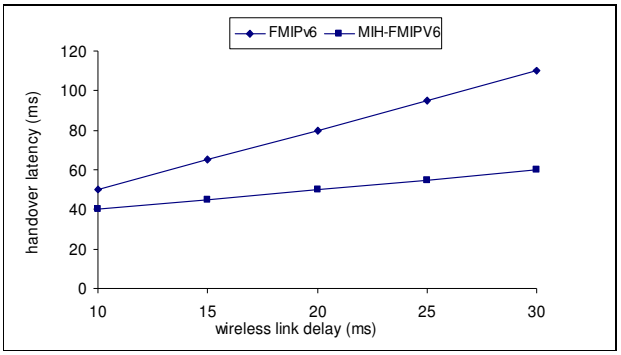

Figure 27: Handover delay vs wireless link delay

We suppose in case of FMIPv6 that at least D_{pre} is equal to $D_{MN\text{-}AR}$ which corresponds to FBack transmission delay. Figure 27 shows that using MIH services

within FMIPv6 reduces to the half the handover latency regarding the wireless link delay.

1.30.2. Buffer Size

In FMIPv6 packet loss is avoided by forwarding packets from PAR to NAR. In the case where MN is about to attach to the new PoA when packets already reached PAR they should be buffered until MN announces its attachment by sending UNA message. Buffering is needed when the forwarding delay is less than L2 handover delay plus attachment announce delay:

$$D_{PAR\text{-}NAR} < D_{L2HO} + D_{UNA}$$

With $D_{PAR\text{-}NAR}$ is the delay between PAR and NAR, D_{L2HO} is the L2 handover delay and D_{UNA} is the delay to transmit UNA from MN to NAR. Therefore, when PAR and NAR are far away from each other topologically or L2 handover is fast, there is a chance to accomplish a seamless handover without any buffering. Moreover, as buffering is a very expensive operation for routers, the buffer size dedicated to MN application will be limited to a maximal value. Therefore packets will be lost if this buffer is full. In order to avoid packet loss we should have

$$D_{PAR\text{-}NAR} + BS / BR < D_{L2HO} + D_{UNA}$$

Where *BR* is the bit rate of the application and *BS* is the buffer size

Figure 28 shows the amount of lost packets in case of different buffer sizes regarding the distance between PAR and NAR.

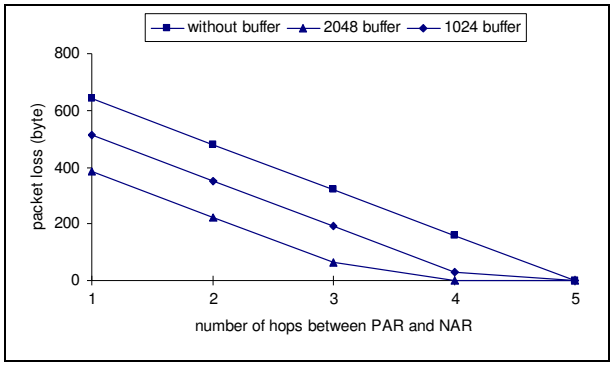

Figure 28: packet loss vs PAR-NAR distance for different buffer sizes

Assuming that all forwarded packets are buffered before receiving UNA by NAR, the necessary buffer size that should be dedicated to forwarded packets in access router is expressed as follows:

$$\mathrm{BS} = \int_0^{D_{ho}} R(T_0 + t)dt$$

Where T_0 is the moment MN stops receiving packets from PAR and R(t) is the instantaneous application bit rate. BS is proportional to handover latency, so reducing D_{ho} means also reducing BS.

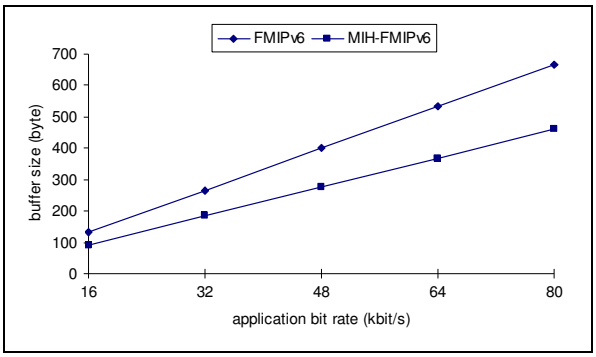

Figure 29: Buffer size vs handover delay ($D_{MN\text{-}AR}$ =15ms)

The required buffer size in the access router when the user is receiving a stream with a bit rate of 64 Kbit/s is reduced by 30% with our scheme (see Figure 29).

1.30.3. Packet Loss

We define the critical time D_{ct} as the period of time between T_{fbu} the instant of sending FBU and the instant when PAR starts buffering packets. During this delay, MN is exposed to the risk of loosing packets after an unexpected L2 disconnection. In classical FMIPv6, PAR starts buffering incoming packets only after transmission of FBack at T_{FBack}, so all packets sent between FBU and FBack can be lost. The critical time is:

$$D_{ct} = T_{fbu} - T_{FBack}$$

With utilisation of link event subscription, any unexpected disconnection will be notified by the oPoA and indicated to PAR which starts packet buffering. The critical time ends at the instant of subscription response reception T_{subs} which is prior to T_{FBack}:

$$D_{ct} = T_{fbu} - T_{subs}$$

In reactive mode, the proposed mechanism avoids packet loss when the link breaks down during exchange between the two access routers. In conventional FMIPv6 packets are buffered only when FBack is received by PAR, so if the mobile is disconnected between the instant PAR receives FBU and the instant it receives FBack, all packets will be lost during this period. The amount of lost packets depends on the delay between PAR and NAR. Even though PAR and NAR might be geographically close to each other, they can be topologically far from each other especially in heterogeneous networks.

1.31. Simulation

The proposed mechanism is evaluated using network simulator 2 (ns2.29) [92] with MIH module that was developed by the national institute of standards and technology (NIST) [93]. We developed FMIPv6 module as a new agent in the mobile node and access routers. In addition MIH is modified to support the proposed enhancement.

Table 11: simulation parameters

Parameter	value
Mobile speed	3m/s
application bit rate	49,6kbit/s
Overlap zone	2m
Propagation model	tworayground
Mobility model	linear

In order to evaluate MIH-FMIPv6 against FMIPv6 as it is described in [65] (i.e. reverse address resolution is made locally using IS information), we consider the following scenario: MN is connected with WIFI interface and moves out from WIFI

zone and enters to WIMAX sector. MN is receiving CBR (constant bit rate) traffic from a fixed node. Table 11 shows simulation parameters.

1.31.1. Handover Delay

Handover delay is the period of time separating reception of the last packet on the WIFI interface and the first received packet on the WIMAX interface. Figure 30 depicts the results of this simulation. Simulation is executed for different values of LGD coefficient with and without the proposed enhancement.

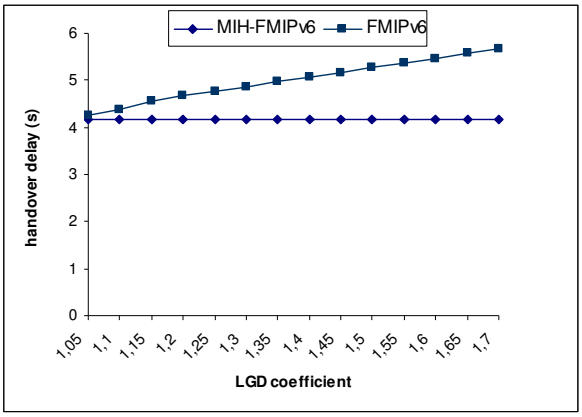

Figure 30: Handover delay versus LGD coefficient

Results of handover delay show that for FMIPv6 protocol, handover delay is increasing when LGD coefficient increases. At the opposite, MIH-FMIPv6 handover delay is not affected by the level of LGD coefficient. This result is compliant with our objectives since the link is still useable even after a link going down event. The high handover delay is due to the connection delay of WIMAX interface.

1.31.2. Buffer Size

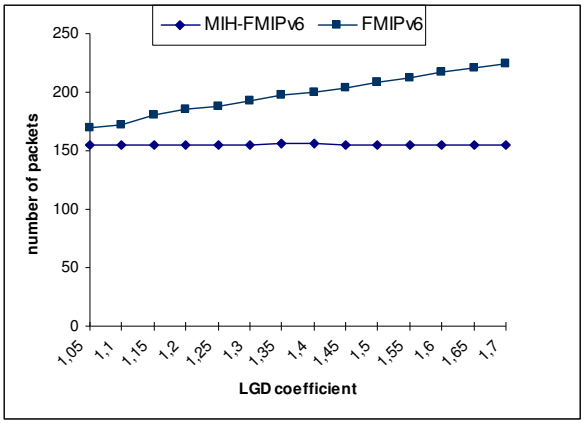

Figure 31: Buffer size versus LGD coefficient

Packets that are forwarded to the new location of MN should be buffered in the NAR since Layer 2 handover delay is too long. Buffer size dedicated to those packets depends on the aforementioned handover delay. This is approved by the pattern of the curves in Figure 31 that follow the same behavior as in handover delay. FMIPv6 needs an increasing buffer size when LGD coefficient increases since it covers more data packets. As for MIH-FMIPv6 buffer size is reduced and stays constant when LGD coefficient increases.

1.31.3. Packet Loss

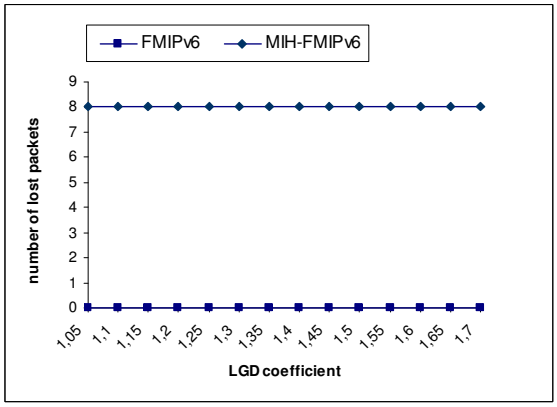

Figure 32: Packet loss versus LGD coefficient

Figure 32 shows packet loss during the handover for different values of LGD coefficient. In FMIPv6 packet loss is null because packets are forwarded through the

tunnel and buffered in the NAR. As for MIH-FMIPv6 8 packets are lost whatever the value of LGD coefficient. Packet loss in our scheme is justified by the transmission of enqueued packets in the PoA. In order to reduce packet loss an optimization can be made in the PoA in order to retrieve remaining packets in the queue and forward them to NAR.

1.32. Ping Pong Effect

In our proposition the handover is not systematic after sending FBU message, but becomes effective only after physical disconnection of the terminal. The additional time during which MN is connected to the oPoA can correct an erroneous movement detection caused by fast fluctuation of signal strength.

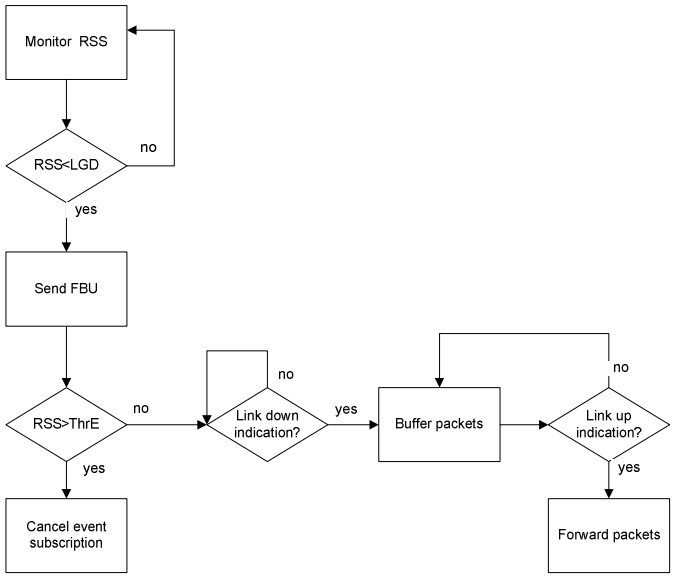

Figure 33: Ping pong avoidance Algorithm

Using the received signal strength (RSS) as criteria for choosing the best network, three thresholds are defined for each technology: i) LGD is the threshold from which the FMIPv6 process is triggered by means of FBU message; ii) ThrE is the entry threshold to the network, it corresponds as well to link_up event notification; iii) ThrO is the exit threshold and corresponds also to link_down notification. Figure 33 presents the algorithm that allows avoiding unnecessary handover due to temporary weakness in the signal strength. The advantage of using MIH event service is the control of FMIPv6

protocol. Any improvement in the signal level after a LGD indication deactivates the whole process by cancelling event subscription in both oPoA and nPoA. Cancellation mechanism can be implemented by means of timers, when expired, the event subscription is deleted.

1.33. Conclusion

In this chapter, we proposed a collaborative handover mechanism for fast moving mobiles by using FMIPv6 protocol with assistance of MIH services. The collaboration between the mobile node and the network allows performing optimized handovers. First we use the information service to collect neighboring networks information before the handover trigger which reduces handover initiation delay. Then, we used MIH subscription primitives to subscribe to link down and link up events, respectively in old and new point of attachment. This procedure allows old access router to forward packets to MN until its physical disconnection. Thus, we reduce handover latency, critical time and dedicated buffer size in access routers. Results of analytical study and simulation showed that our scheme is efficient and allows reduction of handover delay and buffer size in access routers. Finally we exploit MIH services to avoid undesirable handover because of erroneous movement detection.

MAXIMIZING PREDICTIVE MODE PROBABILITY IN FAST HANDOVERS FOR MOBILE IP

1.34. Introduction

Mobility management across multi technology wireless networks is an important feature for ubiquitous access. In such heterogeneous environment, service providers have to deal with service continuity especially for real time applications like voice over IP. MIPv4 and MIPv6 can not meet requirements of these applications. In order to achieve this goal, IETF proposed FMIPv6 to reduce the handover delay. This fast scheme allows MN to anticipate layer 3 handover by acquiring the IP address proactively and involving access routers of old and new access networks. At the opposite of MIPv6 and SIP where several and distant entities intervene in the session update (i.e. AR, home agent, sip registrar, CN), FMIPv6 involves only access routers which are by nature close to the MN. However the role of the ARs is temporary; they do not replace in any case the role of the home agent. Moreover, movement detection is achieved by using layer 2 events which is faster than layer 3 movement detection scheme. Nevertheless, FMIPv6 should deal with layer 2 information in addition to layer 3 information which makes it a cross layer protocol.

FMIPv6 operates in two modes: predictive mode and reactive mode. FMIPv6 handover is achieved seamlessly when the predictive mode is used. The failure of the predictive mode degrades the handover performance and disrupts the perceived quality. Although FMIPv6 standard mentions that it relies on L2 events to trigger mobility operation, it does not specify any mechanism to provide this intelligence. We consider handover optimization done in the previous chapter. Therefore, MIH services are used in order to provide FMIPv6 with the necessary information from lower layers to achieve intelligent handover properly. In this chapter, we propose a maximization of FMIPv6

predictive mode probability by forwarding fast binding acknowledgement (FBack) to the MN in the new network after a successful fast binding update (FBU). Normally, MN is not aware of the success of the binding update if the FBack is not received in the old network. But using the mechanism proposed in chapter 5 PAR can track the MN and forward the FBack according to its location. With our enhancement the predictive mode will depend only on the delivery success of one message (i.e. FBU). Consequently, a new state machine has to be defined for MN to take into account the new enhancement. The impact of the proposed improvement is investigated through an analytical study.

1.35. FMIPv6 Analysis

FMIPv6 may fall into reactive mode if the reverse resolution is not successful or because of the non accomplishment of fast binding update (i.e. sending FBU and receiving FBack). If one of these two messages is not delivered correctly, MN should resend FBU from the new link. In a related work [65] some authors propose to reverse resolve NAR prefix long time before LGD event using Media Independent Information Service. Indeed, this operation will reduce the preparation handover delay and increase the probability of FBU success. However, FBU message itself or FBack can be lost because of unpredicted link down. This leads MN to operate in reactive mode which consists of sending FBU from the new network. Such operation will have many negative consequences on the overall performance of the handover. It does not only increase the handover latency but also may cause packet drops. Packet loss is caused either by the unreachability of the MN in the old link or at least the buffer in the NAR will be overloaded and some packets should be dropped. We differentiate two variants of reactive mode: lossless reactive mode and lossy reactive mode. Lossless mode takes place where FBU delivery is successful and the tunnel is established between PAR and NAR by means of HI/HAck, but FBack is not delivered correctly to the MN, packets are not lost during L2 handover since they are forwarded and buffered in NAR until MN attaches to the new network as long as NAR buffer is not overloaded. Whereas in lossy reactive mode, the tunnel was not established between the two ARs because FBU is lost in the wireless link or one of HI/Hack messages is lost.

1.36. Retransmission System

FMIPv6 is a layer 3 protocol, which means that it has no transport protocol. All FMIPv6 messages are encapsulated over IP either in ICMPv6 message or using mobility headers. Therefore, there is no way for FMIPv6 to ascertain the correct delivery of a message since it does not use a reliable transport protocol like TCP. For this reason a backoff mechanism was defined to provide retransmission of lost messages. After sending FBU message for example, MN starts a backoff timer with a determined timeout. If the FBack is not received before the timer goes off, MN should retransmit FBU and doubles the backoff timer. MN should keep doubling backoff time and retransmitting FBU until it reaches the maximum retransmission retries. If it is the case, MN should conclude that FMIPv6 is not supported by the NAR. Usually backoff time corresponds to the round trip delay between MN and NAR, if this delay is not known, a default value is used.

1.37. Enhanced FMIPv6

MN considers that the predictive mode was unsuccessful once it does not receive FBack in the old link even if the FBU transmission was successful and the tunnel between PAR and NAR was established. In order to increase probability of the predictive mode we propose to forward FBack to MN as soon as it connects to the new network (see Figure 34). In ordinary FMIPv6, forwarding/buffering of data packets is started without any insurance of the current position of MN, this can lead to FBack loss if MN is no more attached to the old link. In the proposal presented in the previous chapter, PoA keeps the AR informed about the position of MN by means of MIH remote service event. Following the location of MN, FBack is delivered accordingly; if it is still in the old link, the message will be delivered directly, if MN is already in the new network FBack will be forwarded via the tunnel.

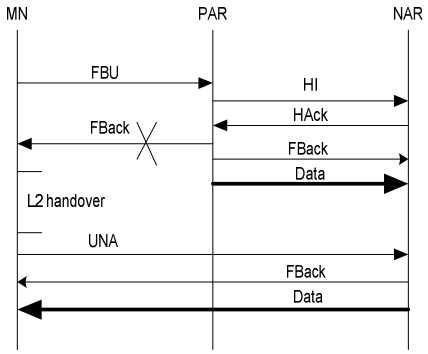

Figure 34: FBack forwarding

1.38. MN State Machine

In the proposed improvement, MN collaborates with the network (i.e. PoA and PAR) to achieve a seamless handover. However, MN should be aware of its state all the time to have a coherent handover operation. To clarify this point we will take the example of a MN that has lost its connection with the PAR immediately after sending a FBU. Although, FBU was successful and media packets are forwarded to the NAR and buffered, MN assumes a fail of the predictive mode and turns on the reactive mode after L2 handover. In order to overcome this inconsistency we propose a new algorithm (see Figure 35) for MN to be in phase with the network. MN will follow the same retransmission system as in the normal FMIPv6 after receiving the LGD trigger. When waiting for FBack message, backoff timer will be doubled each time FBU is retransmitted. If the link goes down without receiving FBack in the old link, MN can not be sure of the delivery of FBU message after the last retransmission. In fact, MN can not know if the non reception of FBack is caused by the loss of this latter or because FBU was not received correctly by the PAR in the first place. Therefore MN persists in retransmitting FBU until FBack is received. In order not to repeat the exchange HI/Hack between the PAR and NAR since the tunnel was already established as FBU was correctly received by PAR, we propose to send FBack locally from the PAR without any new exchange with the NAR as long as the prospective CoA in the FBU message has not been changed. Thus, PAR will ignore the subsequent FBU sent by MN and resend the FBack again.

After a L2 handoff without FBack reception in the old link, MN sends UNA message to NAR and waits for FBack to be forwarded from NAR. This waiting time gives the opportunity to FBack message to reach MN before it resends a new FBU from the new link. Whenever FBack is not received within backoff time, MN concludes that FBU was not successful and the tunnel was not established, consequently, it activates the reactive mode.

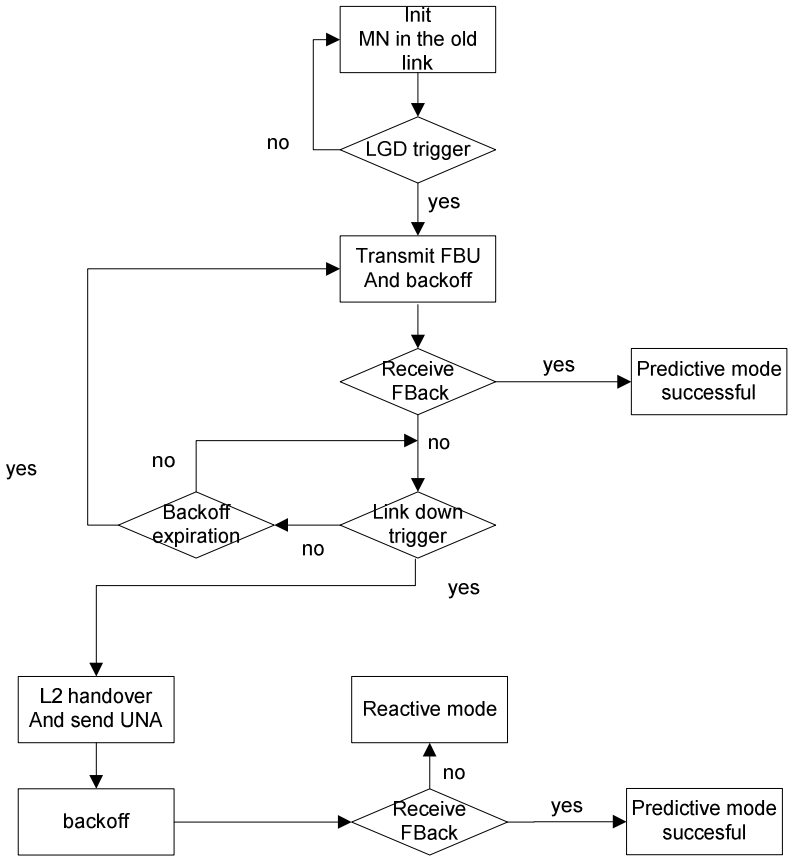

Figure 35: New MN state machine

1.39. Analytical Study

In this study we are interested in the probability of predictive mode success. It should be noted that the reverse resolution of NAR prefix using the discovered PoA MAC address is done before the LGD trigger. Therefore, the success of the predictive

mode depends on the delivery success of FBU and FBack messages in the ordinary FMIPv6, whereas in Enhanced FMIPv6 (E-FMIPv6) it depends only on the success of FBU. We assume that the probability of packet drop in the wired network is null in order to focus only on the wireless link. Packet loss and/or packet error is caused essentially in the wireless link by collision, bit error, or unavailable resources in the PoA depending on the used wireless access technology. We consider p as the probability of any incorrect reception of the packet in the wireless link regardless the reason behind this failure.

Table 12: FMIPv6 messages size

FMIPv6 message	Size (bytes)	Size with options (bytes)
PrRtSol	48	72
PrAdvRt	48	128
HI	46	96
Hack	46	72
FBU	48	96
FBack	48	88
UNA	64	-

MN can operate either in predictive mode or reactive mode, thus we have:

$$P_{\text{pre}} + P_{rea} = 1$$

Where P_{pre} is the probability of predictive mode and P_{rea} is the probability of the reactive mode.

As depicted in Table 12, all FMIPv6 messages do not exceed the fragmentation limit in the IP layer. Therefore, these messages are encapsulated in only one IPv6 packet. FBU transmission is considered unsuccessful if the FBU message is not received correctly or FBack is not received correctly, then the probability of FBU retransmission is:

$$q = p + (1 - p)\,p$$

where p is the probability of non correct reception of FMIPv6 message since each FMIPv6 messages is sent in one IPv6 packet (i.e. no fragmentation is needed). Whereas in E-FMIPv6 the probability of retransmission is

$$q = p$$

because the binding success depends only on correct delivery of FBU to the PAR.

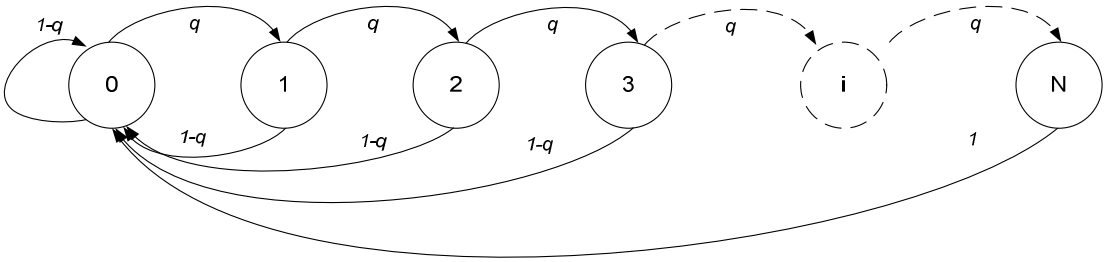

Figure 36: Transition state diagram of FMIPv6 retransmission

Figure 36 represents the transition state diagram of retransmission operation in FMIPv6. The probability of successful FBU transmission after i retransmissions does not depend on the number of retransmissions done before the current transmission. This property is called memoryless property, hence

$$P\,(pre\,/\,i = n+1) = P(pre\,/\,i = n) = (1-q)$$

Hence the delay of a successful FMIPv6 transaction after i retransmissions is:

$$D = 2D_{MN-PAR} + 2D_{PAR-NAR} \qquad if\ \ i=0$$

$$D = 2D_{MN-PAR} + 2D_{PAR-NAR} + \sum_{i=1}^{n} 2^{i-1} \times backoff \qquad if\ \ i>=1$$

Where $D_{MN\text{-}PAR}$ is the delay between MN and PAR and $D_{PAR\text{-}NAR}$ is the delay between PAR and NAR. We assume symmetric delay in uplink and downlink and we neglect the processing delay in PAR and NAR, hence we can write

$$D_{MN-NAR} = D_{MN-PAR} + D_{PAR-NAR}$$

Taking into consideration $D_{LGD\text{-}LD}$ the estimated remaining time before the link is down, the total transaction delay should be less than $D_{LGD\text{-}LD}$ to accomplish a successful predictive mode, which means that the maximum number of retransmissions should be bounded by

$$n = \left\lfloor \log_2 \left(\frac{D_{LGD-LD} - 2D_{MN-NAR}}{backoff} \right) - 1 \right\rfloor$$

1.40. Numerical Results

In order to explore the impact of the estimated time before the link is down on the number of possible retransmission of FBU message, we take the following values for the backoff time and MN-NAR delay respectively: backoff=100ms, $D_{MN\text{-}NAR}$ =50ms.

In Figure 37 , we can observe that the estimated connection time is a very constraining parameter since even for 3 seconds of remaining time the MN has the right to retransmit FBU only 3 times before the link is down. For this reason MN had better succeed in transmitting FBU in the first attempt.

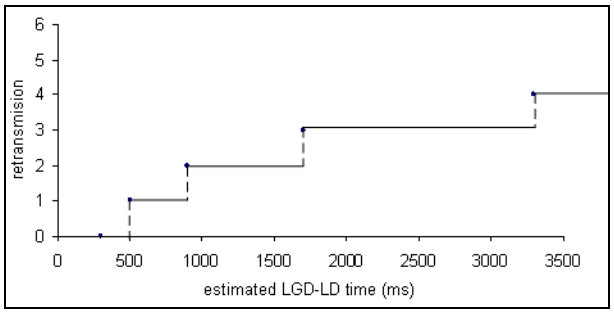

Figure 37: Number of retransmissions versus estimated LGD-LD time

Figure 38 shows the retransmission probability of both ordinary FMIPv6 and E-FMIPv6 regarding retransmission probability. The retransmission probability means also a failure of the fast binding update and consequently the non establishment of the tunnel between the PAR and the NAR. Each time FBU is retransmitted in the ordinary FMIPv6 is a step toward the reactive mode since the number of retransmissions is bounded by the LGD-LD time. Therefore, reducing the probability of FBU retransmissions leads to increase the probability of predictive mode success.

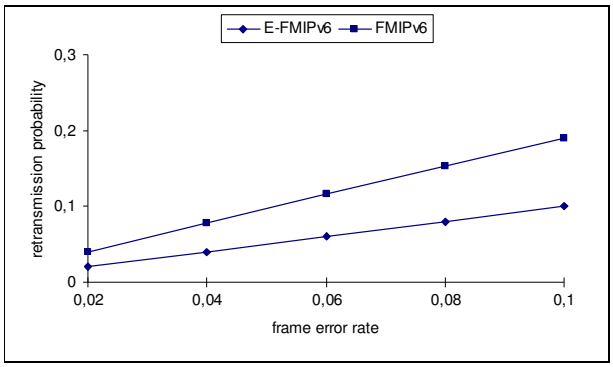

Figure 38: Retransmission probability versus frame error rate

1.41. Conclusion

FMIPv6 is a good solution towards seamless service continuity among different IP subnetworks using different wireless access technologies. Nevertheless, the reactive mode of this protocol penalizes the performance of the handover. In this chapter we focused on the maximization of predictive mode probability by forwarding FBack message to the new location of the MN. This scheme is feasible when MIH services are used in both MN and network side (i.e. NAR and PoA) to localize the MN and then deliver FBack accordingly. This slight modification in the original FMIPv6 has obliged us to adjust the behaviour of MN's state machine in order to achieve the handover coherently. In the analytical study it has been shown that the new changes has brought significant enhancement to the probability of predictive mode.

MOBILITY MANAGEMENT IN HETEROGENEOUS ACCESS NETWORKS IN IMS

In this chapter we tackle the mobility issue in IP Multimedia Subsystem (IMS). Although IMS was designed to integrate different access networks, mobility management among these networks is still unresolved. We propose a novel hybrid mobility management scheme, based on tight cooperation between fast handovers for mobile IPv6 (FMIPv6) and session initiation protocol (SIP) to ensure an uninterrupted real-time service. Moreover, the new Media Independent Handover (MIH) service is integrated into IMS architecture in order to perform intelligent and accurate horizontal and vertical handovers. We investigate two handover cases: selected handover and forced handover. Selected handover takes place when user equipment (UE) is connected to the network via two interfaces at the same time and decides to upgrade the quality of its connection following a given criterion (i.e. cost, bandwidth…etc) without having any difficulty in the previous link. As far as forced handover is concerned, it occurs when the signal reaches a critical level and MN is forced to make a handover in order to maintain the ongoing communication. This case is managed in two phases. The first one or the fast phase is handled by FMIPv6 protocol to preserve as soon as possible packets of the ongoing communication. The second one or the slow phase is handled by SIP protocol to optimize packet delivery route. By doing so, we exploit the benefits of both network layer and application mobility protocols to ensure a continuous session over the two networks without imposing new elements to the network. Through a comparison with other mobility mechanisms, we show in the analytic analysis that our hybrid scheme presents better results in terms of handover latency and packet loss.

1.42. Introduction

IMS is a framework for delivering multimedia services over IP networks. It was originally designed by the third Generation Partnership Project (3GPP) as a part of the vision of evolving mobile networks beyond GSM. IMS [75] is the new approach adopted by 3GPP towards networks convergence; it was designed to be access independent and ubiquitous.

3GPP adopted the IETF SIP standard for IMS signaling which is considered by nature compatible with internet world. IMS architecture is organized in three horizontal layers that separate user plane, control plane and application plane (see Figure 39). This layering scheme allows isolating the service from network layer by interjecting session control functions for all application types. Therefore any operator can offer a unified interface to service providers to develop a wide range of applications independently. Nevertheless, mobility management through heterogeneous networks is a new challenge that IMS should face. IMS already provides personal mobility for nomadic users, but still need to deal with service continuity within non 3GPP networks. Indeed, IMS inherits mobile faculties from SIP, however, terminal mobility performance is questionable especially for mid session mobility in heterogeneous networks [76]. For this reason, 3GPP has integrated only personal mobility into IMS framework and defined different scenarios for session continuity according to the degree of session control.

In order to overcome SIP limitations in providing seamless service continuity, another mobility management protocol (MMP) should take over the handover management completely or partially to reduce handover latency as well as the number of lost packets. Many mechanisms were proposed in the literature to provide an alternative to SIP or supporting it in achieving a seamless handover. Nevertheless, these mechanisms have themselves some limitations or inflict a big change to the network. In this chapter we propose a new hybrid scheme integrating two mobility solutions with minimum changes in the network. We consider two scenarios: the first one is called selected handover and the second one is called forced handover. Selected handover takes place when the mobile detects a new network using an interface other than the one

currently receiving the stream. Although the mobile does not experience any noticeable degradation in the old network, a new interface gets connected to a network that presents better QoS parameters. Therefore, UE decides to switch the flow from the current interface to the new one. Whereas forced handover is performed where a multi interface UE is obliged to change its point of attachment because of signal fading or QoS degradation. Pre-established thresholds determine the minimal level of signal power required to maintain user experience undisrupted. Under these conditions, our purpose is to minimize handover latency and packet loss with minimal changes in the network. Forced handover consists of two phases; the first phase uses an enhanced FMIPv6 to preserve packets from loss during layer 2 handoff. A tunnel is created between the two access routers in order to forward data packets and buffer them in the new access router. As soon as the UE connects to the new access network, data packets are forwarded to UE on the new link. The second phase uses SIP to update the session and redirect packets to the new destination. We show in the analytical study that the handover performance is improved.

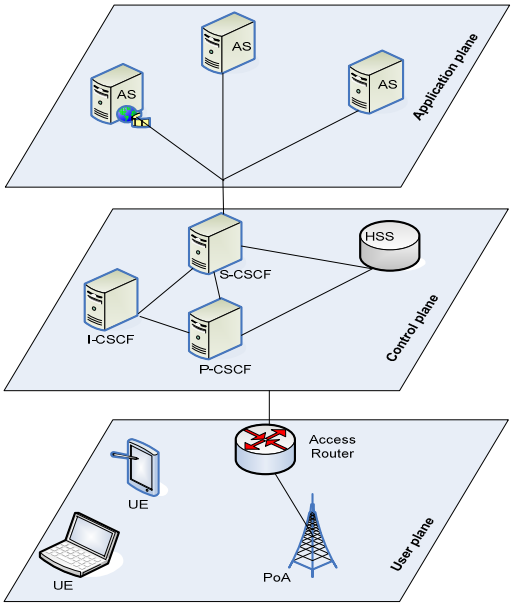

Figure 39: IMS layers

1.43. Related Work

This section gives an overview of IMS architecture and the interworking methods between heterogeneous networks in IMS. Afterwards, some related works to mobility management in IMS are presented and discussed.

1.43.1. IMS Architecture

IMS architecture introduces new elements to establish and control multimedia session. The main components of IMS are the following:

- Proxy-Call Session Control Function (P-CSCF) is considered as the interface between UE and IMS network in order to redirect SIP messages between IMS network and UE.
- Interrogating-Call Session Control Function (I-CSCF) is a SIP proxy server which routes SIP request to the adequate S-CSCF.
- Serving-Call Session Control Function (S-CSCF) is the central point of the IMS control plane. It acts as a SIP registrar which maintains a binding between the user location and the user SIP address.
- Application Server (AS) hosts and executes all services offered by IMS.
- Home Subscriber Server (HSS) dose a similar function as HLR (Home Location Register) in the GSM network. This means that it is a database of all subscribers' profiles.

1.43.2. Interworking

Interworking is the Integration of several heterogeneous access networks. In the context of IMS, interworking is attaching all access networks to the IMS core (see Figure 40). This attachment is done either physically or logically in order to control the session regardless the access network from which UE is connected. There are two levels of integration:

- Tight coupled interworking: also known as 3GPP IP access mode, in this case the WLAN/WMAN is attached physically or via a tunnel to the core network and is totally controlled by the 3GPP network. Two types of IP addresses are necessary in this case. The first one is allocated by the WLAN network and the second one is

allocated by the 3GPP network. An IPsec tunnel is established between UE and packet data gateway (PDG) which is placed in the entry of the 3GPP network and receives all the traffic coming from the WLAN.

- Loosely coupled interworking: or IP direct access, in this mode only high level components especially control and signaling information are exchanged between the two networks, i.e. AAA authentication, mobility management, QoS management…etc.

3GPP has detailed functional incremental scenarios to achieve the ultimate goal, which is a 3GPP subscriber with a WLAN radio interface having the same services as when using a 3GPP radio interface (GSM, GPRS, UMTS) with mobility between 3GPP and non-3GPP radio interfaces similar to intra-3GPP mobility. The six scenarios that 3GPP has defined correspond to incremental interworking functionalities from 1 to 6 as shown in Table 13.

Table 13: incremental interworking scenarios

Service and operational Capabilities:	Scenario 1: Common billing and customer care	Scenario 2: 3GPP system based access control and charging	Scenario 3: Access to 3GPP system PS based services	Scenario 4: Service continuity	Scenario 5: Seamless services	Scenario 6: Access to 3GPP system CS based services
Common billing	X	X	X	X	X	X
Common customer care	X	X	X	X	X	X
3GPP system based access control		X	X	X	X	X
3GPP system based access charging		X	X	X	X	X
Access to 3GPP system PS based services from WLAN			X	X	X	X
Service continuity				X	X	X
Seamless service continuity					X	X
Access to 3GPP system CS based services with seamless mobility						X

For each scenario there are additional functionalities that should be added to the architecture to support the aimed service.

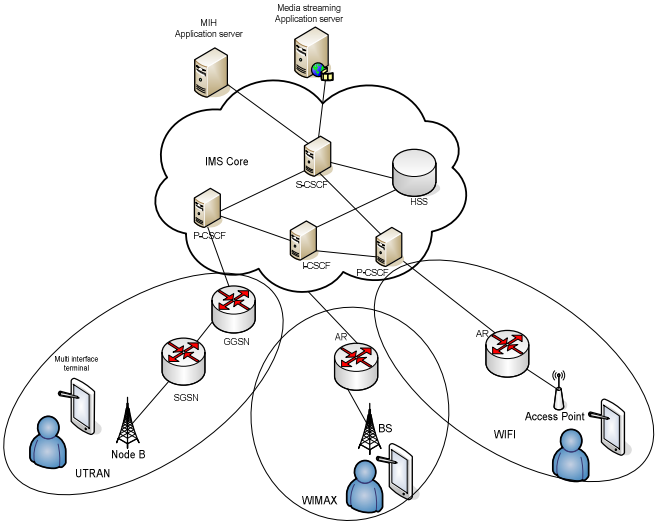

Figure 40: Interworking architecture in IMS

Interworking is the first step towards service continuity in access independent IMS. [77] specifies system description for interworking between 3GPP and WLANs. As for WIMAX interworking with 3GPP, WIMAX forum has proposed an interworking architecture [78] based on the same documents as WLAN interworking. It covers both direct IP access (loosely coupled) and 3GPP IP access (tightly coupled), it does not change the previous one, but only add, the missing elements for WIMAX interworking.

1.43.3. Mobility Within IMS

In [79] three mobility management protocols (MIPv6, SIP, and PMIPv6) are investigated in IMS 3GPP2 context for heterogeneous access networks (i.e. WIFI, CDMA2000). The experimental results show a handover delay of up to 4s for MIPv6 and PMIPv6. This delay reaches 9s when SIP mobility is used.

In [80] the authors use MIP to handle mobility of TCP traffic, and SIP for real time traffic. MIP and SIP create a kind of redundancy in the network and cause double registration and double binding update. Although the authors of [81] tried to reduce this redundancy by merging entities that have similar functionality, still traffic separation is used which adds more complexity in the network. Moreover using only SIP for real time traffic in IMS framework will introduce a big handover delay.

In [82] MIP and SIP are used for real time traffic consecutively. MIP is used first to reroute the traffic to the new network and than SIP updates the route from the correspondent node by re-registering and re-inviting the session. Here again network entities are redundant due to double use of mobility protocols i.e. Home Agent (HA), Foreign Agent (FA) and sip proxy server. Moreover MIP is known to be unsuitable for real-time applications because of its triangular routing and high handover delay. The corresponding UE will update its destination IP address after receiving a re-invite message which cancels the function of HA. Netcape [83] also tries to optimize handover delay by using MIP for underlying traffic redirection.

IHMAS [84] is a proactive solution that aims to reduce vertical handoff latency; it predicts vertical handoffs at client side and starts session reconfiguration before handoff management. This proposal adopts an application layer proxy-based approach, for session signaling and data handoff. It uses application server for session continuity that reduces handoff media losses by decoupling session rebinding and data transfer times. Nevertheless, this solution takes into account only the case of make-before-break, in other words the authors assume that there is always enough time for the UE to prepare the connection to the new network before leaving the original one. While SIP operations (register and invite) in addition to the basic configuration operations need a long time to be achieved.

Media-independent pre-authentication (MPA) is a framework that allows UE to pre-authenticate proactively before layer 2 handover. It allows also a pre-configuration of UE with the appropriate IP address. In addition, the binding update of the corresponding network entities (HA, CN, SIP server etc) is carried out proactively whatever the used MMP. After a handover implying layer 3 modification, UE has to re-register before resuming its communications. Updating the route with SIP proactively has a good impact on handover delay and packet loss, nevertheless, the long time taken by SIP to re-register and update the session is not always easy to predict. Actually, handover predication mechanisms, especially layer 2 triggers such as link going down and link down are the key elements for a successful proactive handover. On the one hand the whole handover procedure is triggered by a warning event (i.e link going down) based on signal strength or QoS degradation. On the other hand, the time

between this first warning and the instant of the actual disconnection or QoS degradation notified by link down event is beyond user control. The remaining connection time should be at least equal to all time needed by the mobile to perform successful proactive update. This time is remarkably high in the case of SIP, given the multitude of network elements intervening in re-registration procedure. The handover preparation may go in vain because of unpredicted disconnection; hence the mobile has to update its session reactively once connected to the new network.

1.44. Hybrid Mechanism

In this section we will present our solution for IMS mobility using enhanced FMIPv6 and SIP.

1.44.1. MIH Integration to IMS Framework

Media independent handover aims to provide MMPs with intelligence in executing both vertical and horizontal handover. MIH as introduced in chapter 5 is independent of the interface technology. Integrating MIH to IMS is done at both access network and core network. In the access network, MIH is implemented in UE, PoA and AR (or GGSN); whereas the information server (IS) is placed in the core network.

As far as media independent information service (MIIS) is concerned, the communication between IS and the remote MIHF is organized following a request response paradigm. In [85], HTTP is used to retrieve the needed information from the IS. The authors of [86] propose SIP to transport MIIS messages, but they don't specify how this information is transported neither the adequate SIP messages to carry them. Actually, SIP seems to be a good solution for transporting MIIS messages and will facilitate the integration of IS into IMS architecture. We propose to deploy IS as an application server (AS) placed in the core network in order to be accessed from all access networks by authenticated and authorized users. IS has a database of information about access networks. Examples of such information are: geographical location, operator name, PoA MAC address, IP address prefix of its AR, operational channel…etc.

Surrounding networks of a given UE can be determined using its IP address or its GPS coordinates if available. Access to this information is conditioned by the authorization of the HSS which allows only users who are already subscribed to this service to reach the needed information. Moreover, the need for neighboring access network knowledge is done at least once after being connected to the network. Updates of the list of neighboring networks can be done in a regular basis depending on UE mobility behavior.

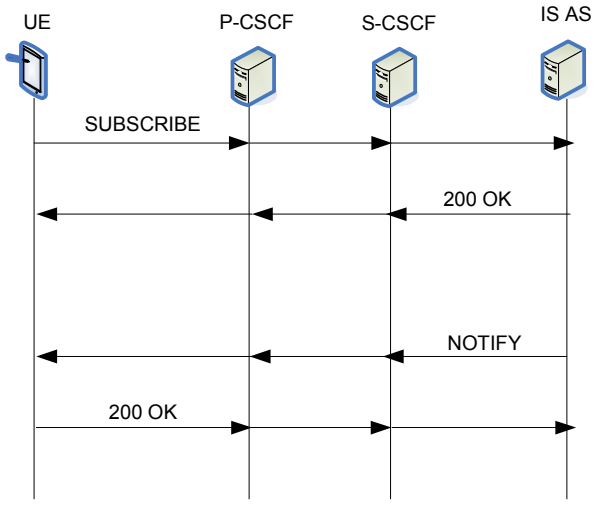

Figure 41: IS integration into IMS

After address configuration and re-registration if necessary, UE should learn about the available wireless networks in its area in order to have a list of candidate networks to handover to in case the signal of the current connection goes down. This anticipation will be very benefic for UE when it needs to reverse resolve the IP address of the AR behind the discovered PoA. Indeed this resolution will be done locally when the handover is imminent, thus reducing the handover preparation delay. In the context of IMS, the list of the neighboring networks will be provided using SUBSCRIBE and NOTIFY messages. It should be noted here that MIH standard does not specify any transport mechanism for information service messages. We propose to carry the info request message on SDP payload of Subscribe and the info response message on notify SDP payload as illustrated in Figure 41.

1.44.2. Hybrid Scheme for IMS Mobility

The goal of any handover scheme is to ensure service continuity with an acceptable perceived quality of experience for the user during the transition. In chapter 4, mobility operations are classified into four main steps: movement detection and address allocation, global location tracking, traffic redirection and handover smoothing. SIP can perform the three first operations by nature. The missing part towards seamless service continuity is the handover smoothing. The idea behind our proposition is to propose a solution based on MIH intelligence to make an appropriate anticipation operation depending on low layers events. Our solution does not imply a big change to the existing network and avoids redundancy among the intervening entities. We will use an enhanced version of FMIPv6 combined with SIP as mobility management protocols. We consider intra domain inter technology and intra domain intra technology handovers. This means that the whole network including all access networks are controlled by the same administrative authority and the mobile can travel among access networks within the same or different technologies (e.g. WIMAX to WIMAX or WIMAX to UMTS). Access networks are connected to IMS core network either loosely or tightly. The advantage of using FMIPv6 is that there is no need for home agent since the only mission of FMIPv6 is to preserve the current session by forwarding and buffering packets from the old AR to the new AR respectively. In other words, FMIPv6 will take in charge the handover smoothing operation, whereas global location tracking and traffic redirection will be performed by SIP.

1.44.2.1. Selected Handover

In selected handover, there is no time constraint on handover achievement since the UE continues in receiving the traffic from the current interface and prepares at the same time all needed updates using the new interface. Special care should be taken in the triggering mechanism of such handover. It is even the major criterion that decides on what type of handover has to be carried out. Therefore, basing on the MIH framework, the adequate triggers that should lead to selected handover are of two types: link up and link detected. After notification of the MIHF, decision algorithm is executed and decides whether the new network is better or not. Obviously, UE would prefer to be

always best connected, therefore when a better network is detected by any other interface, decision engine compares the offered quality of the new network with the quality of the current network. By the term “quality” we refer to a group of criterions that have been configured in the decision engine and should be taken into account when deciding which network is better. Those criterions can include bandwidth, cost, user preferences… etc. UE performs SIP related updates before switching to the new network. It should be noted that always the first operation is IP address acquisition. This operation can be performed by means of different methods: stateful method (e.g DHCP, PDP) or a stateless method (e.g autoconfiguration). IP address acquisition is done normally like if the UE intends to connect to the network using this interface for the first time without taking into account the ongoing handover. After that, UE starts updating the session. First of all, it should discover the new P-CSCF as it could differ from the old one. An additional optimization in the handover that can be performed using MIH framework is that the information about the new P-CSCF can be provided in the info response that was acquired earlier. Afterwards UE negotiates a security association with the P-CSCF in order to construct an IPsec tunnel to secure communications between UE and P-CSCF. Subsequently, UE starts updating the session by re-registering to the S-CSCF which needs first to re-authenticate UE since it is using a different IP address now. Dual use of IP address and being registered to IMS domain with both of them an allowed operation which is known as multiple registrations. Until now traffic is following the old route towards the old IP address since the interface is still connected. Eventually, UE sends re-invite to all its corresponding UEs in order to update the destination address of their streams towards the mobile UE. Finally this latter UE can deregister the previous IP address by sending de-register message from the old interface. Message flow corresponding to selected handover scenario is shown in Figure 42. Message exchange related to IP configuration, security association and I-CSCF/HSS are omitted.

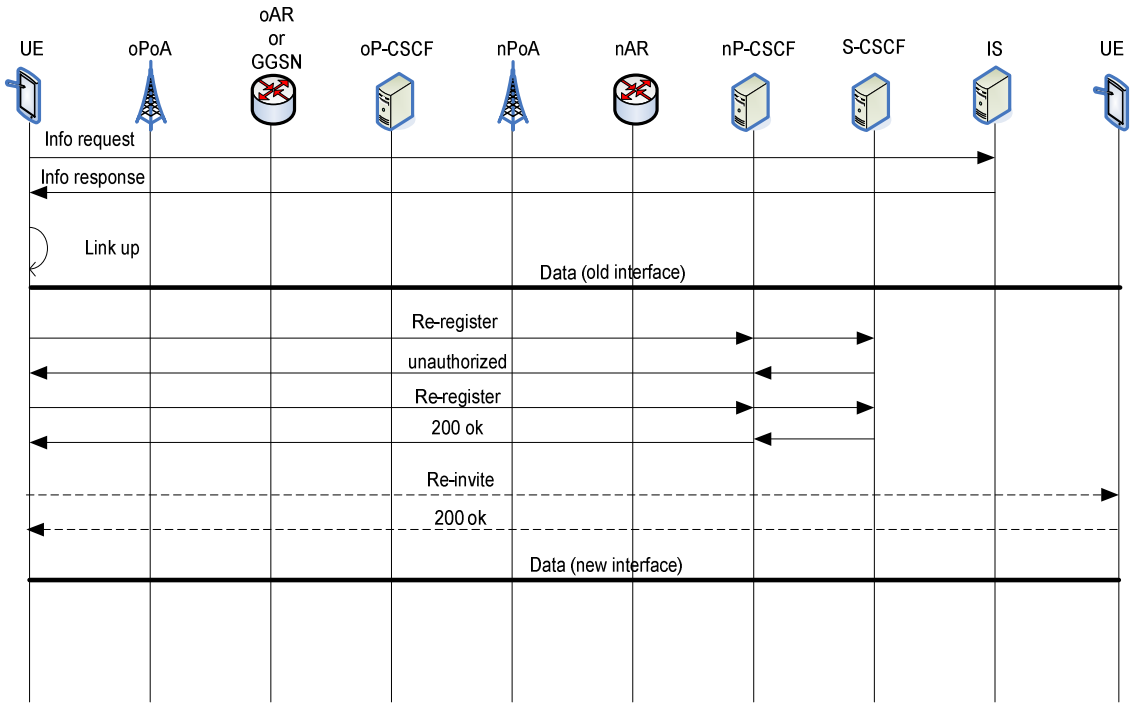

Figure 42: Selected handover message exchange

1.44.2.2. Forced Handover

Forced handover flow chart is shown in Figure 43. In this scenario UE is constrained to change its connection to a new network using the current interface or a second one which is not already connected. The mechanism is carried out in two phases: fast phase and slow phase.

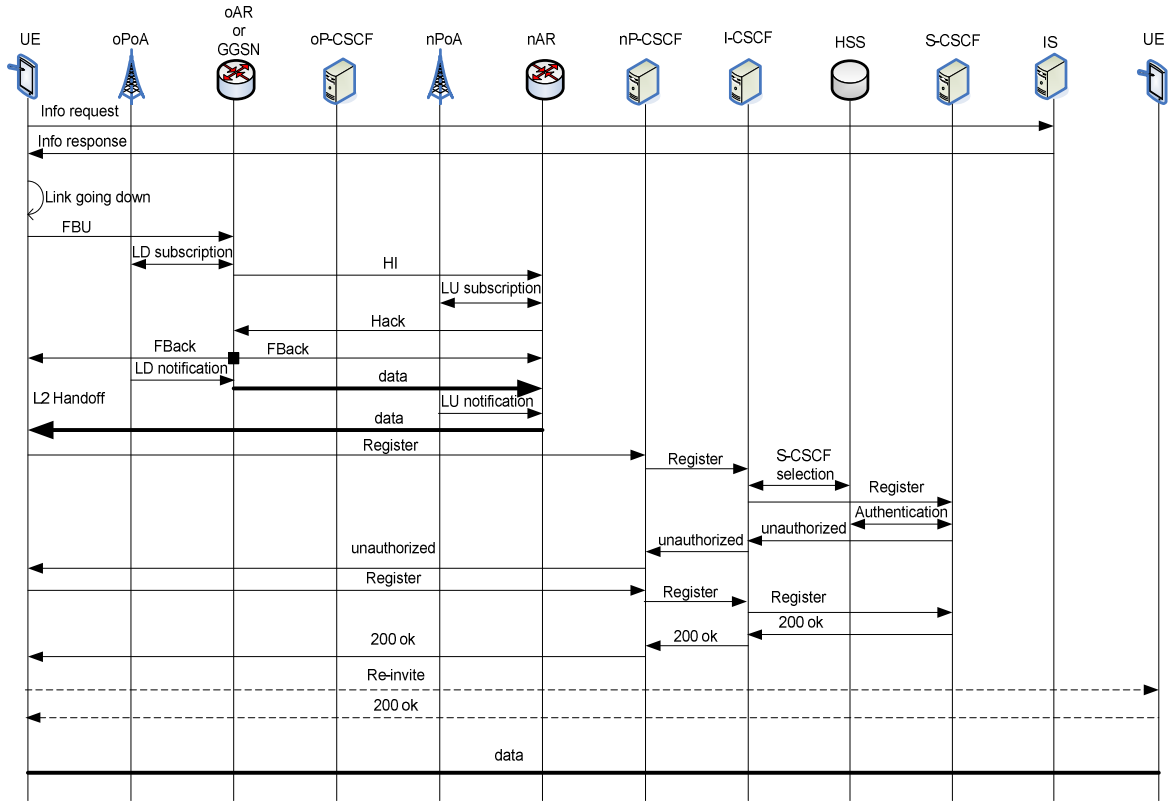

Figure 43: The proposed hybrid handover mechanism

1.44.2.2.1. Fast Phase

Our first priority is to preserve packets of the ongoing session before any other procedure. The first operation is the reverse resolution of the new PoA mac address to acquire the new access router (NAR) prefix. UE uses IS to know about the neighboring networks and their AR prefixes before the handover trigger. After a link going down trigger, UE scans the channels of its interfaces and resolves the NAR address locally using the information provided in the info response of the IS and then constructs a Prospective CoA. After that, FMIPv6 operation is carried out to protect packets from loss. Nevertheless, sending a fast binding update (FBU) message will prevent UE from receiving any packet because they are encapsulated to the NAR. So we need to trigger the tunnel creation by another mechanism which is separated from FBU transmission. The proposed collaborative scheme in chapter 5 is used here to perform this task. Therefore, the terminal can continue its communication in the previous network even

after link going down notification. The tunnel between the two ARs allows preservation of data packets during layer 2 handover and continuity of the service while performing SIP updates.

1.44.2.2.2. *Slow Phase*

As soon as UE connects to the new access network, it starts updating its SIP session while receiving data packets through the tunnel. It should be noted here that SIP related messages are sent with the new IP address, whereas data packets are encapsulated using the old IP address and sent through the tunnel. The reason behind this distinction is that the corresponding UE may drop data packets if they are sent with a source address witch is different from the one that the session was established with in the first place. Only after receiving re-invite message, correspondent UE takes into account the new destination address. On the contrary, SIP messages should be sent directly without encapsulation because this address will be used in security association establishment between UE and the new P-CSCF.

UE starts by discovering its new P-CSCF. This information can also be provided by the IS in the previous information response. Afterwards, UE is re-authenticated, re-registered and finally re-invites the corresponding UE.

Note that this scenario is based on the assumption that FMIPv6 predictive mode is successful. Indeed, FMIPv6 Predictive mode has more chance to succeed than SIP update anticipation, because it only needs to send FBU message. Whereas, the success of SIP session update requires success of several operations. In the case where FBU is not received in the old link, the reactive mode is activated and FBU is sent from the new network.

1.44.3. TCP and UDP Traffic Support

For this hybrid scheme to work well, long sessions like voice calls and multimedia streaming should be transported over a connectionless channel like UDP. For text based web access, traffic model [87] does not require service continuity since it is an on/off behavior (download and read). But for long TCP connections as in the case of downloading big files using FTP, FMIPv6 tunnel should be kept up until the end of the download. Although FTP sessions inflict a long tunnel establishment time, this scheme

is better than MIP where the tunnel is permanent regardless the application currently in use. In FMIPv6, The information about the tunnel establishment time is inserted by the UE in FBU message and transmitted to the NAR by the HI message. The difficulty of this solution resides in the estimation of the remaining TCP connection time and the mechanism to maintain the tunnel established along this period. Actually we need to modify FMIPv6 protocol to add a new mechanism for maintaining and terminating the tunnel between PAR and NAR.

In case of selected handover, the solution is simpler since it suffices that the decision engine postpones the handover execution until all TCP connections are closed. In such cross layer design, the decision engine is in the center of information flows coming from lower layers (i.e. link events) and upper layer (i.e. TCP/UDP session)

1.45. Performance Evaluation

In this section performance of our mobility management scheme is analyzed and compared with SIP and MIP-SIP solutions. The studied performance parameters are handover delay and packet loss.

1.45.1. Handover Latency

In a multimedia entertainment, the handover delay should be reduced to its minimum so that users do not feel any interruption in the session during the handover. The majority of the media are transported via RTP/UDP where no retransmission is available. Therefore, any lost packet will penalize users' quality of experience.

Handover delay or latency is defined as the time separating the reception of the last packet in the old network and the reception of the first packet in the new network. The total handover delay D_H is the addition of 3 components:

$$D_H = D_{H2} + D_{L3} + D_{MMP}$$

Where D_{H2} is layer 2 handover delay. We assume equal values of D_{H2} for both vertical and horizontal handovers since in this scenario the new interface is not up at the time of handover decision.

D_{L3} is the delay of layer 3 operations; it concerns mainly IP address acquisition

D_{MMP} is the delay inflicted by the MMP to update the related entities following the used MMP.

The rest of the notations are reported in Table14

Table 14**:** Notation table

$D_{UE\text{-}AR}$	Delay between UE and AR
$D_{UE\text{-}P}$	Delay between UE and P-CSCF
$D_{P\text{-}I}$	Delay between P-CSCF and I-CSCF
$D_{I\text{-}S}$	Delay between I-CSCF and S-CSCF
$D_{S\text{-}H}$	Delay between S-CSCF and HSS
$D_{P\text{-}S}$	Delay between P-CSCF and S-CSCF
$D_{UE\text{-}HA}$	Delay between UE and HA
$D_{PoA\text{-}AR}$	Delay between PoA and AR

1.45.1.1. SIP

UE acquires the IP address using neighbor discovery protocol, which consists of sending a solicitation to the AR and receiving an advertisement. Then it discovers the new P-CSCF, registers with the S-CSCF and re-invites the corresponding UE which corresponds to the delays D_{disc}, D_{regis} D_{inv} respectively. In case of using SIP as MMP we have:

$$D_{L3} = D_{sol} + D_{adv}$$
$$= 2D_{UE\text{-}AR}$$

We can write D_{SIP} as:

$$D_{SIP} = D_{disc} + D_{regis} + D_{inv}$$

with

$$D_{disc} = D_{UE\text{-}P} + D_{P\text{-}UE}$$
$$= 2D_{UE\text{-}P}$$

and

$$D_{regist} = 4D_{UE\text{-}P} + 4D_{P\text{-}I} + 4D_{I\text{-}S} + 2D_{I\text{-}H} + 2D_{S\text{-}H}$$

and

$$D_{inv} = 4D_{UE\text{-}P} + 4D_{P\text{-}S}$$

We suppose that all paths are symmetric and have the same transmission delay in both directions.

1.45.1.2. MIP-SIP

In MIPv6 the CoA is first acquired from the AR and then binding update is performed within HA. Subsequently packets are forwarded by HA to UE. In MIP-SIP scheme we have:

$$D_{L3} = D_{sol} + D_{adv}$$

$$D_{MIP} = D_{BU} + D_{ack}$$

$$= 2D_{UE\text{-}AR} + 2D_{UE\text{-}HA}$$

1.45.1.3. E-FMIP-SIP

The advantage of using FMIPv6 is that the IP address acquisition is done proactively. So, D_{L3} is null in this case.

Upon detecting a link down in the PoA packets are forwarded through a tunnel between PAR and NAR, and buffered at NAR simultaneously with layer 2 handoff. Once NAR is notified of a link up, packets are delivered. Thus in predictive mode with MIH enhancement (E-FMIP) we have:

$$D_{E\text{-}FMIP} = D_{LUnot} + D_{AR\text{-}UE}$$

$$= D_{PoA\text{-}AR} + D_{AR\text{-}UE}$$

as for reactive mode

$$D_{FMIPrea} = D_{UNA} + D_{FBU} + D_{HI} + D_{Hack} + D_{FBack}$$

$$= 3D_{UE\text{-}AR} + 4D_{AR\text{-}AR}$$

1.46. Numerical Results

1.46.1. Handover Delay

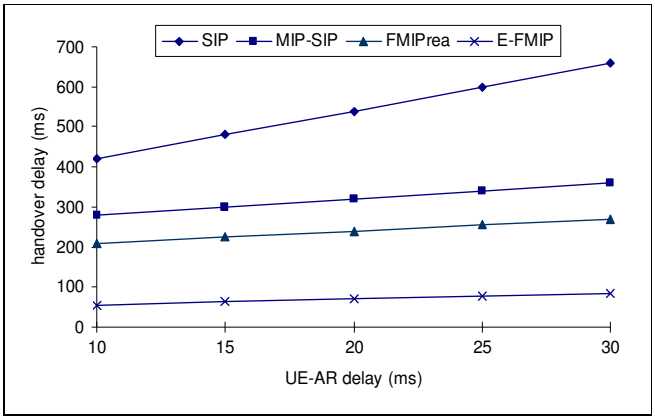

Figure 44: Handover delay vs UE-AR delay

Figure 44 shows that the handover delay of the E-FMIPv6 scheme is low compared to the other schemes even when the access link delay is high which corresponds to complicated access network architecture like GPRS. Nevertheless success of the predictive mode is crucial to perform such low delay. As depicted in Figure 45, FMIPv6 scheme is not sensible to the distance that separates UE from its home network since it involves only the AR of the concerned networks. It also reduces the triangular routing to a small scale.

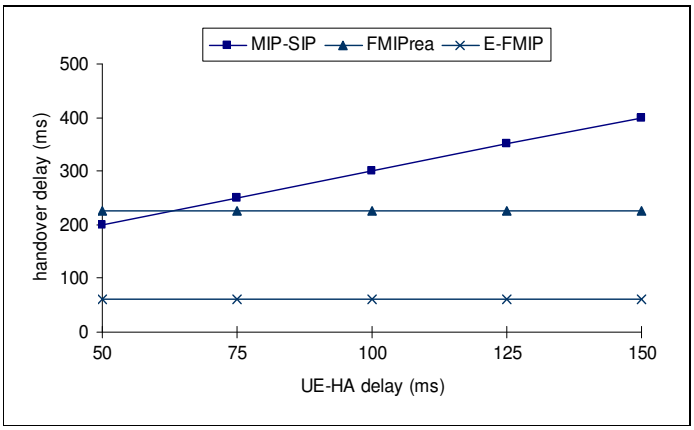

Figure 45: Handover delay vs UE-HA delay

1.46.2. Packet Loss

During the handover some packets may be lost if they are not buffered as in FMIPv6. The amount of lost data is expressed as follows:

$$PL = \int_0^{D_H} R(T_0 + t)dt$$

Where R(t) is the application bit rate and T_0 is the instant of the last received packet on the old link. Figure 46 shows that packet loss in our solution is nearly null. More concretely, in a video streaming application that has a bit rate of 1024 kbit/s there are only one or two packets which are lost during the handover. Such a weak number of lost packets does not have a big impact on the perceived quality. On the contrary, the user will miss a long video sequence during the handover when the other MMPs are used.

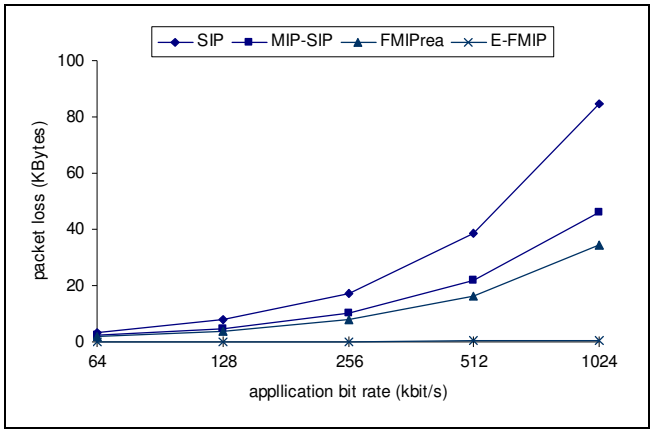

Figure 46: Packet loss vs application bit rate

1.47. Conclusion

Interworking between different access networks is an important step towards networks convergence. But IMS mobility over heterogeneous access networks is still an open issue. In order to offer ubiquitous service to 4G customers, telecom operators should resolve vertical handover issue. In this chapter we proposed a hybrid scheme using enhanced FMIPv6 and SIP. This scheme does not impose any new entities on the network as MIPv6 does. A performance evaluation of the proposed solution was carried out and compared to the other schemes. The results show that our hybrid scheme

reduces the handover delay and packet loss. However, these results are tightly dependant on the success of the predictive mode of FMIPv6.

PERFORMANCE ANALYSIS OF MOBILE IP VARIANTS IN VERTICAL HANDOVER

1.48. Introduction

Mobile IP and its variants present a good solution for mobility in All IP networks. Choosing one variant or the other depends on many parameters and leads to different results in terms of handover performance. Such decision is crucial when it comes to real time applications such as Voice over IP and video streaming. To maintain service continuity when moving, the mobile should perform horizontal handover when the adjacent network is of the same technology or vertical handover when the new network is different from the previous one. FMIPv6 is desirable for homogeneous networks where the MN has to perform a hard handover. In this case, FMIPv6 prevents the ongoing communications from packet loss and reduces the layer 3 handover by preparing the IP address before the layer 2 handover. Nevertheless, in heterogeneous networks where using two or more interfaces, a classical MIPv6 has also the possibility to acquire a new IP address on the new interface using the conventional neighbor discovery protocol [58] or any other mechanism. The heterogeneity of networks gives the mobile an important potential of performing seamless handover since we have at our disposal two wireless interfaces which have the possibility to work simultaneously and independently. In fact, the connection can be established in the new link before disassociating from the current point of attachment. This kind of handover is known as make-before-break handover. Nevertheless, make-before-break handovers are not systematic, too late handover triggering can lead to a disrupting transition. Consequently, the choice of the layer 3 mobility protocol has a direct impact on the performance of the resulting handover. Although the two networks are geographically near from each other, they could be topologically far away from each other. An example of this is the case of two adjacent wireless networks that belongs to different operators.

Packets routed between the two networks should follow the route established by the autonomous system according to a particular policy. In this chapter, we focus our interest on the use of MIPv6, HMIPv6, and FMIPv6 in heterogeneous environment. We show that mobility protocol performance depends essentially on: 1) available time before the link with old network is broken, 2) the delay between visited network and home network or the delay between the visited network and the CN in case of route optimization and finally 3) the delay between previous access network and new access network. When the MN has the possibility to perform a make-before-break handover, MIPv6 and HMIPv6 seem to be the best layer 3 mobility protocols to use. However, when the handover is triggered by a serious fading in the received signal, then the choice will depend on the protocol operation delay before the link is down. Performance evaluation takes into account handover latency, disruption time and packet loss. Finally we give the guidelines for choosing the high-performance protocol according to the numerical results.

1.49. Related Work

It could appear obvious that FMIPv6 by principle performs faster handovers than any other IP based mobility protocol. This is true in horizontal handover but not necessarily true in vertical handover as we will demonstrate. Although many works treats the problem of vertical handover but the majority propose new scheme to perform seamless handover and few of them make comparative study between them as we are suggesting. The authors of [88] conducted a study based on simulations to compare different variations of mobile IP but it concerns only the case of horizontal handover. In [89] the authors define a classification of vertical handovers and makes performance evaluation of SIP and MIP.

1.50. Vertical Handover Process

1.50.1. Vertical Versus Horizontal Handover

Handover or handoff event is the transition from one point of attachment to another. We talk about horizontal handover when both PoAs belong to the same technology (e.g. WIFI to WIFI). Horizontal handover can be initiated by the mobile or the network as it is the case in 3G and WIMAX networks. Vertical handover takes place when the mobile has more than one interface each one belongs to a particular technology. The ability of roaming between the different networks while continuing the ongoing session is called vertical handover.

1.50.2. Triggering Mechanism

Vertical handover is only possible if the MN is equipped with two or more wireless cards of different technologies (i.e. WIFI, WIMAX, UMTS…etc), when the current network is not providing the required QoS any more or the signal strength is becoming weak, the MN can proceed to a vertical handover.

The vertical handover can be triggered by one of the following events: link_parameters_report, link_going_down (LGD), link_detected and link_up, considering the use of MIH event service. The considered triggers take into account the seamless behavior of the handover. When the mobile experiences difficulties in maintaining an acceptable level of QoS for its ongoing applications, MIH user will be notified of a link going down event or link_parametrs_report event to perform a scan for next network to handover to. Two scenarios are possible: i) the new interface is already up and likely has already acquired its IP address, or ii) the interface is down and the MIH function turns it on by means of link_action command (see chapter 5) with certain execution delay. Handover is performed seamlessly only if the remaining connection time is long enough for handover preparation whatever the used mobility protocol is. There are many estimation algorithms [90] that can predict the link down (LD) event within certain period of time, but the connection can be broken after a severe fading in the signal.

Link detected or link up events could happen at any time in any interface other than the one currently in use. If the new detected network presents a better opportunity for the user in terms of QoS or cost for example, the decision engine may choose to switch to this network.

1.50.3. Address Acquisition

MN starts resolving the new network prefix using the discovered Access Point identifier. When an information server is available there is no need for sending a solicitation message because the resolution is done locally for all mobility protocols.

This step is done in FMIPv6 by sending a PrRtSol to the AR to acquire the prefix address of the AR behind the candidate AP. This operation is known as reverse address resolution. We should note here that this operation is successful only if the PAR has already the correspondent layer 3 information of the candidate AP. Otherwise Candidate Access Router Discovery (CARD) [91] should be implemented between ARs to exchange and update access router information. Moreover, when an entry is expired or a MN wants to handover to a new network for the first time, PAR needs to request a CARD server to resolve AR's info for the unknown PoA. Thus, a proxy router solicitation will take an important delay to perform the reverse address resolution. At the opposite, MIPv6/HMIPv6 always requests AR info directly from the NAR without any proxiying process, and just by using the conventional neighbor discovery protocol. Thus, the resolution operation takes only one RTT.

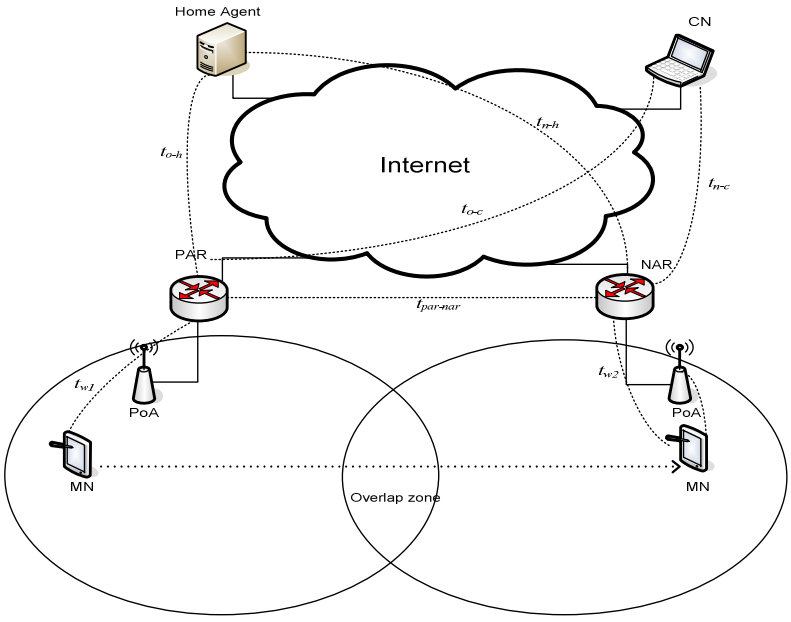

Figure 47: delays between different entities

1.51. Performance Evaluation

In this section, we will focus on the following performance parameters: protocol operation delay, handover delay, disruption delay, and packet loss. We consider the mobility scenario in Figure 47.

1.51.1. Protocol Operation Delay

Flow charts of the considered mobility protocols are sketched below. We distinguish between bidirectional tunneling operation of MIPv6 (Figure 48) and route optimization with return routability procedure (Figure 49).

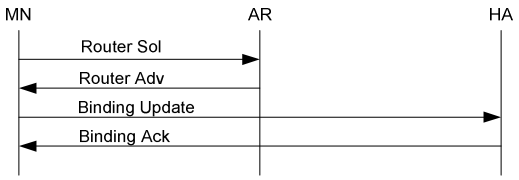

Figure 48: MIPv6 bidirectional tunneling

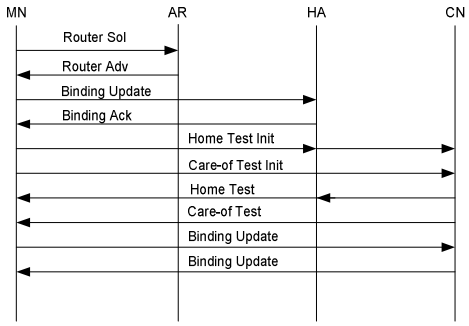

Figure 49: MIPv6 route optimization

We also consider the case of interdomain mobility in HMIPv6 (Figure 50) and predictive mode/reactive mode of FMIPv6 (Figure 51).

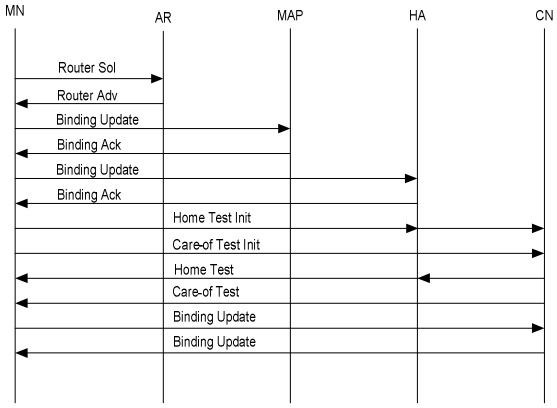

Figure 50: HMIPv6 inter domain mobility and route optimization

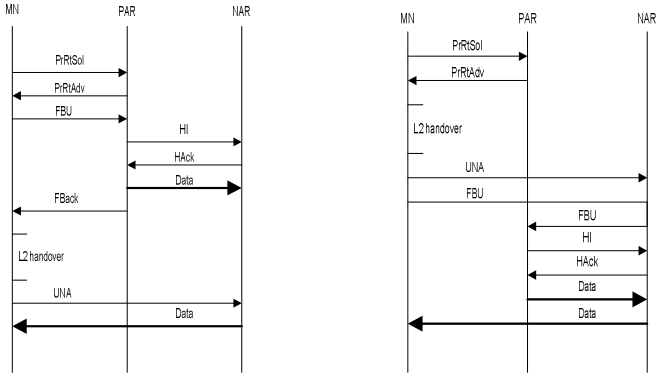

Figure 51: FMIPv6 operation modes (predictive left, reactive right)

Protocol operation delay T is the necessary time for the MN to exchange all messages relevant to the mobility protocol before the data flows switch to the new interface. Measuring this delay is important because it should be less than LD estimation time to have a seamless handover. We consider that there is no information server in the network, so that MN should solicit the AR to obtain new network prefix.

Following are protocol operation delays for: MIPv6 with reverse tunnelling T_{MIP}, MIPv6 with route optimization T_{MIPro}, FMIPv6 in predictive mode $T_{FMIPpre}$, FMIPv6 in reactive mode $T_{FMIPrea}$, and HMIPv6 in inter domain mobility with route optimization T_{HMIPro} :

$$T_{MIP} = 2t_{w2}+2t_{w2}+2t_{n\text{-}h}=4t_{w2}+2t_{n\text{-}h}$$

$$T_{MIPro} = 2t_{w2}+2t_{w2}+2t_{n\text{-}h}+2t_{w2}+2t_{n\text{-}h\text{-}c}+4t_{w2}+4t_{n\text{-}c}$$

$$= 10t_{w2}+4t_{n\text{-}h}+2t_{h\text{-}c}+4t_{n\text{-}c} \text{ with } t_{n\text{-}h\text{-}c}=t_{n\text{-}h}+ t_{h\text{-}c}$$

$$T_{FMIPpre} = 4t_{w1} + 2t_{par\text{-}nar} + t_{w2}$$

$$T_{FMIPrea} = 2t_{w1} + 2t_{w2} + 3t_{par\text{-}nar}$$

$$T_{HMIPro} = 2t_{w2} + 2t_{nar\text{-}map} + 2t_{w2} + 2t_{n\text{-}h} + 2t_{w2} + 2t_{n\text{-}h\text{-}c} + 4t_{w2} + 4t_{n\text{-}c}$$

$$= 12t_{w2} + 2t_{n\text{-}map} + 4t_{n\text{-}h} + 2t_{h\text{-}c} + 4t_{n\text{-}c}$$

Where:

t_{wx} is the delay between MN and its AR, it includes wireless link delay and wired link delay between the point of attachment and AR . x=1,2 for the old network and new network respectively.

$t_{n\text{-}h}$ is the delay between NAR and HA.

$t_{h\text{-}c}$ is the delay between HA and CN.

$t_{n\text{-}map}$ is the delay between NAR and MAP (HMIPv6).

$t_{par\text{-}nar}$ is the delay between PAR and NAR.

$t_{n\text{-}c}$ is the delay between NAR and CN.

HMIPv6 intra domain mobility has the same behavior as MIPv6 with reverse tunneling, because the only operation that the MN should perform is MAP update. This latter plays the role of the HA in the MAP domain.

1.51.2. Handover Delay

Handover delay *T'* is the delay between the last data packet received on the previous interface and the first data packet received on the new interface.

Handover delay for FMIPv6 will depend on the mode in which this protocol is operating.

$T'_{FMIPpre} = t_{par\text{-}nar} + t_{w2}$ which correspond to forwarding delay of received packets

$$T'_{FMIPrea} = 2t_{w2} + 3t_{par\text{-}nar} + t_{par\text{-}nar} + t_{w2}$$

$$= 3t_{w2} + 4t_{par\text{-}nar}$$

Handover delay of MIPv6 and HMIPv6 will depend on $t_{Lgd\text{-}Ld}$ which is the period between LGD trigger and LD. If this period is sufficient for the mobile to perform the

necessary binding update then the handover delay will be null. Otherwise, there will be a disconnection period until the protocol operation achievement.

$$T'_{MIP} = T_{MIP} - t_{Lgd\text{-}Ld}$$
$$T'_{MIPro} = T_{MIPro} - t_{Lgd\text{-}Ld} \qquad if \quad T > t_{Lgd\text{-}Ld}$$
$$T'_{HMIPro} = T_{HMIPro} - t_{Lgd\text{-}Ld}$$
$$T' = 0 \qquad otherwise$$

Here we neglect the end to end delay difference between old and new route.

1.51.3. Disruption Delay

Handover disruption delay *T''* is the delay between the last data packet received by the old interface and the first data packet received by the new interface directly from the CN either by means of tunneling via the HA or directly from the CN when the route optimization is used. We call it disruption delay because during this period, the perceived quality is disrupted because of: 1) layer 2 transition, during which the terminal is not able to receive or transmit any packet, 2) forwarding buffered packets from the old link to the new link is not always done at the same rate as the original flow, moreover, additional end to end delay is added to the delivered packets.

For FMIPv6 handovers, the disruption delay will concern both the current FMIPv6 operation and the following MIPv6 or HMIPv6 updates.

$$T''_{FMIPpre} = T'_{FMIPpre} + T_{MIP}\ (or\ T_{HMIP})$$
$$T''_{FMIPrea} = T'_{FMIPrea} + T_{MIP}\ (or\ T_{HMIP})$$

Disruption time of MIPv6 and HMIPv6 is equal to handover delay *T'= T''*. A forwarding scheme could be used in the case of HMIPv6 to forward packets from old MAP to the new MAP during the binding update of the different entities. In this case, handover delay and packet loss will be reduced but the disruption time remains unchanged.

1.51.4. Packet Loss

In the downlink, packets sent from HA or CN continue in taking the old route until the HA or CN sends an acknowledgement of the binding update. The following packets are sent to the new location immediately. Data packets are lost during the

handover due to two reasons: loss of the connection on the previous interface before accomplishing the updates with HA/CN for MIPv6/HMIPv6, or before receiving FBack message in FMIPv6 (i.e. reactive mode). In FMIPv6 predictive mode there is no packet loss because of packet buffering in the NAR. For the other mobility protocols the number of lost packets is equal to:

$$L = \int_0^{T'} R(t)dt$$

where R(t) is the instantaneous bit rate of the active flow. This value is proportional to the handover delay *T'* and bit rate of applications currently in communication.

The moment from which the MN decides to switch data packets from the old to the new interface is decisive for uplink data packets. During the period when the MN is waiting for the binding acknowledgement in the new interface, packets are still sent using the old interface, therefore with the old address as source address. These packets could be dropped for security reasons at the CN since it was informed that the IP address of its correspondent has been changed, consequently any received data packet with the old address will be rejected. The number of dropped packets due to this inconsistency is proportional to the old route end to end delay. The optimal way to redirect flows with minimal dropped packets is that MN continues sending data using the old interface until the binding update message is sent to HA or CN (following the use of reverse tunnelling or route optimization), at this moment precisely and without waiting for the binding update acknowledgement, it puts all data relevant to the ongoing communications in the new interface. Nevertheless some packets could be lost due to difference in end to end delay between old and new route (i.e. old route end to end delay is bigger than the new route end to end delay). Moreover, reliability of the mobility protocol is affected since MN decides to switch data packets on the new link without having the "evidence" of binding update accomplishment.

1.52. Numerical Results And Discussion

1.52.1. Numerical Results

We consider the following values:

Table 15: Simulation values

parameter	t_w	*Average* t_{n-h}	t_{h-c}	*Average* $t_{par-nar}$	t_{n-c}	t_{n-map}
Value(ms)	15	80	100	80	100	20

In order to compare the performance of the different MIP variants in vertical handover, we calculate the protocol operation delay using increasing values of t_{n-h} and fixed value of $t_{par-nar}$. In a second time, we fixe t_{n-h} and vary $t_{par-nar}$.

In Figure 52, lower values of t_{n-h} correspond also to HMIPv6 intra mobility domain, and $t_{par-nar}$ is fixed to its average. It is clear that FMIPv6 is not affected by the distance from the home network since it involves only the ARs. We observe that when the MN is not very far from the home network MIPv6 performs shorter handovers than FMIPv6. This is due to the low number of exchanged messages in reverse tunneling mode of MIPv6. But when route optimization is used, protocol operation delay is considerably high in both MIPv6 and HMIPv6.

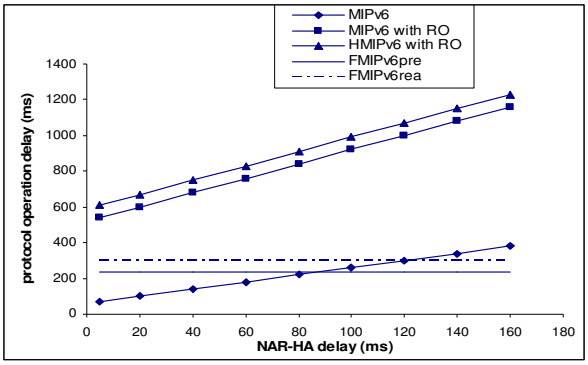

Figure 52: Protocol operation delay vs NAR-HA delay

Figure 53, t_{n-h} is fixed to its average value. This figure shows that FMIPv6 is sensitive to the delay between new access network and old access network. Less than certain limit of $t_{par-nar}$, FMIPv6 has shorter operation delay than MIPv6, but over this limit, especially when the home network is at the same distance from old and new access networks MIPv6 surpass FMIPv6. Moreover when PAR-NAR delay becomes extremely high route optimization use can be also envisaged.

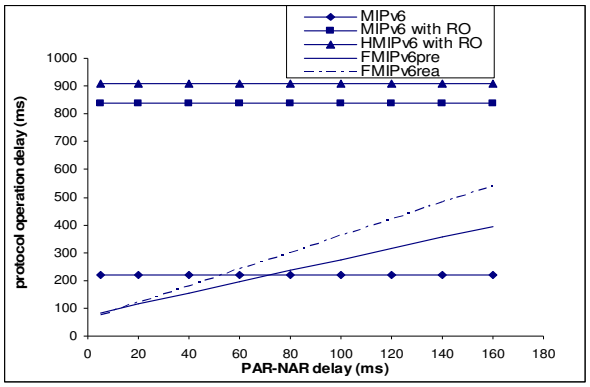

Figure 53: Protocol operation delay vs PAR-NAR delay

To explore the impact of connection remaining time before the old link is broken, we vary LGD-LD time and fix the other parameters and observe its impact on the handover delay. Since the make before break policy is adopted, MN starts establishing the new connection using the new link. Therefore, the handover delay will depend on how long this operation will take which was shown in previous graphs and how long the old link will last in order to allow this establishment simultaneously with data traffic on the old link. In Figure 54, FMIPv6 moves from reactive mode to predictive mode at $t_{Lgd\text{-}Ld}$ =235 ms which corresponds to predictive FMIPv6 average value. Surprisingly, we notice that MIPv6 and HMIPv6 even with route optimization have a chance to perform seamless handover with zero delay, thing that predictive FMIPv6 can not ensure whatever the remaining connection delay. This is due to the communication between the two interfaces and forwarding of data packets via access routers instead of direct switching in the case of the other protocols.

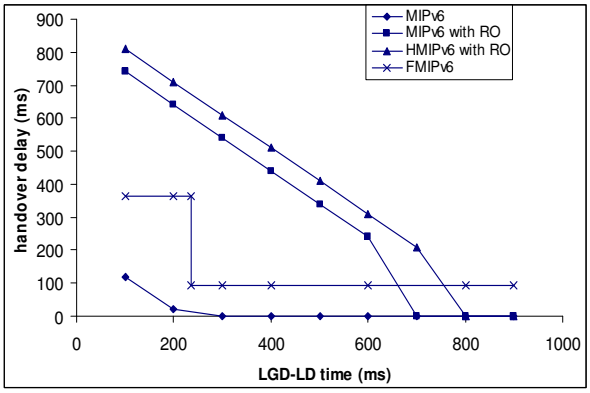

Figure 54: Handover delay vs LGD-LD time

1.52.2. Discussion

The table below gives some guidelines that should help in choosing the mobility protocol to use according to the triggering event and the delay between entities intervening in layer 3 handover.

Table 16: guidelines for mobility protocol choice

Protocol variant	When to use?
MIPv6	Link up or link detected events, short NAR-HA delay, medium tLgd-Ld
MIPv6 with RO	Link up or link detected events, long tLgd-Ld
HMIPv6 intra domain mobility	Link up or link detected events, short NAR-MAP delay, medium tLgd-Ld
HMIPv6 inter domain mobility	Link up or link detected events, long tLgd-Ld
FMIPv6	Link going down event, short NAR-PAR delay, short $t_{Lgd\text{-}Ld}$

1.53. Conclusion

At the opposite of horizontal handover, where the use of FMIPv6 is preferable to other variants because of its proactivity in acquiring IP address and buffering mechanism that prevents packet loss during the link layer handover, vertical handover gives the other variants the opportunity to prepare network layer handover using the new interface before the old interface disconnection. We showed that handover performance depends on the delay between PAR and NAR and the delay between NAR and HA/MAP for FMIPv6 and MIPv6/HMIPv6 respectively. As far as handover disruption time is concerned, FMIPv6 has always longer period of disruption, because the mobile node should perform MIPv6 or HMIPv6 updates after FMIPv6 operation. Moreover, the remaining connection time after a link going down event is important in determining handover delay and the amount of packet loss. On the one hand, it determines whether FMIPv6 operates in predictive or reactive mode. On the other hand, the handover delay and packet loss of MIPv6/HMIPv6 are important when the

remaining connection time is low, and null when it covers at least the protocol operation delay. Following the impact of these parameters on mobility protocols, we gave some guidelines to choose the best protocol that meets QoS constraints during the handover.

CONCLUSIONS

In this thesis different aspects of service continuity have been investigated. Session mobility is an important aspect of service continuity that gives the user more flexibility and liberty when dealing with multimedia services. To achieve this purpose in the context of video streaming, we proposed SESSAMO as a protocol for session transfer between terminals. The main characteristics of this new protocol are its lightweight and its peer to peer character which make it easy in deployment and use. This protocol was validated by an implementation in commercial mobile device (i.e. Nokia 770) and the results from the testbed are vey encouraging. Another issue of session mobility which is media renegotiation was also addressed. In order to adapt the media to the new terminal capacity, there should be a way to renegotiate the adequate media format accurately. As SDP is unable to achieve this task, we proposed to use SDPng along with MPEG21-DIA in order to describe terminal context. Different scenarios was defined when renegotiation is needed and special RTSP messages was specified to convey the description according to user situation.

Afterwards terminal mobility was tackled. After a bibliographic study we drew a classification of mobility operations that the majority of mobility protocols over IP follow to guarantee service continuity through different sub networks. We determined four operations that are not necessarily present in all mobility protocols. This classification is important to analyze and diagnose any mobility protocol to define its limitations. Afterwards we proposed a new handover mechanism based on collaboration between the mobile and the network using MIH services. The innovation of this mechanism resides in the efficient use of MIH events in both mobile node and point of attachment. Analytic and simulation results showed that the proposed solution is efficient in terms of handover delay and buffer size. Moreover, this solution presents a resiliency to ping pong effect. For the same purpose, a minor modification in FMIPv6 protocol allows maximization of predictive mode probability. A simple forwarding of FBack message through the tunnel between the access routers can ensure a successful predictive mode.

IMS is a concrete example where service continuity represents an important link to realize the desired objectives. Although important steps are achieved towards access networks convergence, seamless handover between heterogeneous networks is still matter of research. As IMS relies on SIP in the signaling level, we proposed to use FMIPv6 during SIP updates after layer 2 handover. This solution corresponds to bringing handover smoothing operation to IMS mobility according to the classification done before. SIP and FMIPv6 combination does not only avoid redundancy in the network but improves the performance of the handover as well. Moreover, MIH enhancement can also be applied in the case of IM. This scheme does not impose any new entities on the network as MIPv6 does. Performance evaluation of the proposed solution was carried out and compared to the other schemes. The results show that our hybrid scheme reduces the handover delay and packet loss. However, these results are tightly dependant on the success of the predictive mode of FMIPv6.

Finally, we conducted a comparative study between mobile IP variants in vertical handover. This study demonstrates that the choice of the most convenient mobility protocol should respect certain conditions. After highlighting the specificity of vertical handover and the assets brought to the mobile node compared to horizontal handover, we compared the performance of mobile IP variants in different scenarios. This study proved that having multiple interfaces, the mobile node can perform faster handovers with the conventional MIPv6 than FMIPv6 in certain cases. Guidelines that should help in choosing the best mobility protocol are provided.

Perspectives:

A unique mobility solution for both UDP and TCP connexions is still an open issue. A possible way to resolve this problem is the use of temporary tunnels and a full cross layer communication including layer 2 layer 3 and application layer vertical layers at the client side. In addition, Buffering is a necessary operation towards seamless service continuity, but forwarding buffered packets can cause losses if the transmission rate is higher than what the new link can support. A policy of flushing buffers is necessary to prevent any loss or disturbance at the user level.

Although security was not tackled in this thesis, it should be taken into consideration in order to secure all mobility operations for both session mobility and

terminal mobility. Furthermore security mechanisms should be conceived in order not to compromise the improvement brought to mobility protocols.

REFERENCES

[1] www.jiwire.com

[2] IEEE, "IEEE Std 802.11b™-1999: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band"

[3] IEEE, "IEEE Std 802.11a™-1999: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHz Band".

[4] IEEE, "IEEE Std 802.11g™-2003 Part 11: Wireless LAN Medium Access Control (MAC) and Physical

[5] 802.11n-2009 IEEE Standard for Local and metropolitan area, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput

[6] IEEE 802.11e IEEE Standards for Local and Metropolitan Area Networks: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. Medium Access Control (MAC), Quality of Service Enhancements

[7] IEEE Std 802.16.-2004, IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems.

[8] 802.16a-2003 IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems

[9] 802.16e-2005 IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems

[10] 802.15.3-2003 IEEE Standard Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs)

[11] 802.15.4-2003 IEEE Standard Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)

[12] Zigbee Alliance, "Zigbee specification: Zigbee document 053474r13 Version 1.1," 1 Dec. 2006.

[13] Bluetooth Special Interest Group," Bluetooth Network Encapsulation Protocol (BNEP) Specification", June 12, 2001

[14] Bluetooth Special Interest Group, "Bluetooth Core", Specification of the Bluetooth System, Version 1.1, February 22, 2001

[15] http://www.itu.int/ITU-R

[16] 3GPP TS23.107 V7.1.0 (2007-09):"Quality of Service (QoS) concept and architecture (Release 7)"

[17] 802.11r-2008 IEEE Standard Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Fast Basic Service Set (BSS)

[18] A. kumar "mobile broadcasting with WIMAX, principles, technology and applications" focal press 2008

[19] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, "Locator/ID Separation Protocol (LISP)", draft-ietf-lisp-9, October 11, 2010

[20] R. Moskowitz, P. Nikander, P. Jokela and T. Henderson, "Host Identity Protocol," RFC 5201, April 2008

[21] C. smith and J. Meyer "3G wireless with WiMAX and WIFI 802.16 AND 802.11" The McGraw-Hill Companies 2005

[22] A.T. Campbell et al., "Design, Implementation, and evaluation of Cellular IP," IEEE Pers. Commun., Aug. 2000, pp. 42-49.

[23] A. Misra et al., "IDMP-based Fast Handoffs and Paging in IP-based 4G Mobile Networks, " IEEE Commun. Mag., Mar 2002, pp 138-145

[24] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan and L. Salgarelli, "IP micro-mobility support using HAWAII," draft-ietf-mobileip-hawaii-01.txt, July 2000, Work in Progress.

[25] C. Perkins, Ed., "IP Mobility Support for IPv4," RFC 3344, Aug. 2002

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

[27] R. Koodli, Ed., “Fast Handovers for Mobile IPv6”, RFC 5568, IETF Network Working Group, July 2009

[28] H. Soliman, et al, “Hierarchical Mobile IPv6 Mobility Management”, RFC 4140, IETF Network Working Group, August 2005

[29] Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K.,and B. Patil, "Proxy Mobile IPv6", RFC 5213, August 2008.

[30] H. Yokota, K. Chowdhury, R. Kooldi, B. Patil, F. Xia, “Fast Handovers for Proxy Mobile IPv6”, RFC 5949, September 2010

[31] M. Boutabia, E. Abdelrahmane, H. Afifi , “A hybrid mobility mechanism for heterogeneous networks in IMS”, in the proceedings of the 11th IEEE international conference on Multimedia and Expo ICME 2010, July 19-23, Singapore

[32] A. C. Snoeren, H. Balakrishnan, “TCP Connection Migration”, draft-snoeren-tcp-migrate-00, November 2000

[33] R. Stewart, “Stream Control Transmission Protocol” RFC 4960, September 2007

[34] M. Riegel and M.Tuexen, “Mobile SCTP”, Internet Draft, work in progress, November 2007, draft-riegel-tuexen-mobile-sctp-09.txt

[35] 3GPP TS 23.206 V7.5.0 (2007-12) Voice Call Continuity (VCC) between Circuit Switched (CS)and IP Multimedia Subsystem (IMS);stage 2 (Release 7)

[36] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M.Handley, and E. Schooler. “SIP: Session Initiation Protocol”, RFC 3261, June 2002.

[37] H. Schulzrinne and E. Wedlund, “Application-layer mobility using SIP”, SIGMOBILE Mob. Comput. Commun. Rev. 4 (2000), no. 3, 47–57

[38] H. Schulzrinne, A. Rao and R. Lanphier, Real Time Streaming Protocol (RTSP), RFC 2326, April 1998.

[39] H. Schulzrinne, A. Rao, R.Lanphier, M. Westerlund, A. Narasimhan, “Real Time Streaming Protocol 2.0”, work in progress, IETF Draft, March 6, 2006

[40] Schulzrinne, H., Casner, S., Frederick, R. and V. Jacobson,"RTP: A Transport Protocol for Real-time Applications", RFC 3550, July 2003.

[41] R. Shacham, H. Schulzrinne, S. Thakolsri, and W. Kellerer, “SIP session mobility”, Internet Draft (2006).

[42] M. Rawashdeh, A. Karmouch, "Seamless video handoff in session mobility over the IMS network", WoWMoM 2009.

[43] C. Yun, J Park, and Y. Lim “Session Mobility of IP Multimedia Subsystem (IMS) using Modified Assured (MA) Session Transfer”, 15th Asia-Pacific Conference on Communications (APCC) 2009

[44] Crocker, "Standard for the Format of ARPA Internet Text Messages", STD 11, RFC 822, UDEL, August 1982

[45] Live Networks inc., live555 Streaming Media,USA, 1999

[46] http://www.videolan.org/vlc/

[47] http://www.scratchbox.org/

[48] http://www.wireshark.org/

[49] D. Kutscher, J. Ott, C. Bormann "Session Description and Capability Negotiation ", IETF work-in-progress, draft-ietf-mmusic-sdpng-08, February 20, 2005.

[50] M. Handley, V. Jacobson, C. Perkins, “SDP: Session Description Protocol”, RFC 4566 July 2006

[51] A. Vetro, C. Timmerer, S. Devillers (eds.), ISO/IEC 21000-7:2004, "Information Technology - Multimedia Framework (MPEG-21) - Part 7: Digital Item Adaptation", October 2004.

[52] T. Guenkova-Luy et al, “Harmonisation of Session and Capability Description between SDPng and MPEG-21 Digital Item Adaptation”. IRTF work-in-progress, draft-guenkova-mmusic-mpeg-21-sdpng-00, Feb 2005.

[53] P. Assuncao and M. Ghanbari, “Post-processing of MPEG2 coded video For transmission at lower bitrates,” in Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing, vol. 4, Atlanta,GA, May 1996, pp. 1998- 2001.

[54] A.Vetro, Charilaos Christopoulos, and Huifang Sun , “Video Transcoding Architectures and Techniques: an Overview”, IEEE Signal Processing Magazine, March 2003, pp. 18-29.

[55] A. Vetro, H. Sun, and Y. Wang, “Object based transcoding for adaptive video content delivery,” IEEE Trans. Circuits Syst. Video Technol., vol 11, pp. 387-401, Mar. 2001.

[56] R. Rejaie, Mark Handley, and Deborah Estrin, “Layered Quality Adaptation for Internet Video Streaming”, IEEE Journal on Selected Areas in Communications, vol. 18, No. 12, December 2000.

[57] www.itea-sumo.org/.

[58] T. Narten E. Nordmark, W. Simpson, H. Soliman, “Neighbor Discovery for IP version 6 (IPv6)”, RFC 4861, September 2007

[59] K. El Malki, “Low-Latency Handoffs in Mobile IPv4” RFC 4881 June 2007

[60] N.Nakajima, A.Dutta, Subir Das, Henning Schulzrinne “Handoff Delay Analysis and Measurement for SIP based mobility in IPv6”, In ICC 2003, May 2003.

[61] Seok Joo Koh and Wook Hyun, "mSIP: Extension of SIP for Soft Handover with Bicasting," IEEE Communications Letters, Vol. 12, No.7,pp. 532-534, July 2008.

[62] H. Izumikawa and R. Lillie, “SIP-based bicasting for seamless handover between heterogeneous networks,” IETF Internet Draft, draft-izumikawa-sipping-sipbicast-01, February. 2008.

[63] Park, S., Kim, P., and Volz, B., "Rapid Commit Option for the Dynamic Host Configuration Protocol version 4 (DHCPv4)," RFC4039, IETF, March 2005.

[64] A. McAuley et al., "Dynamic Registration and Configuration Protocol (DRCP) for Mobile Hosts," Internet draft, draft-itsumo-drcp-01.txt, July 2000, work in progress.

[65] Mussabbir, Q.B.; Wenbing Yao; Zeyun Niu; Xiaoming Fu, "Optimized FMIPv6 Using IEEE 802.21 MIH Services in VehicularNetworks" IEEE Transactions on Vehicular Technology, Volume 56, Issue 6, Part 1, Nov. 2007, pp. 3397 - 3407

[66] Yoon Young An; Byung Ho Yae; Kang Won Lee; You Ze Cho; Woo Young Jung, "Reduction of Handover Latency Using MIH Services in MIPv6", AINA 2006, Volume 2, 18-20 April 2006, pp. 229 – 234.

[67] Byungjoo Park; Sunguk Lee; Latchman, H.; "Performance analysis of enhanced-mobile IPv6 with fast handover over end-to-end TCP", WCNC 2006, Volume 1, pp. 581 – 586.

[68] Ruidong Li; Jie Li; Kui Wu; Yang Xiao; Jiang Xie; "An Enhanced Fast Handover with Low Latency for Mobile IPv6", IEEE Transactions on Wireless Communications, Volume 7, Issue 1, Jan. 2008, pp. 334 – 342

[69] Hancheng Lu; Xiaolei Tie; Peilin Hong; "A Novel Buffer Mechanism for Fast Handovers in Mobile IPv6", WOCN '07 2-4 July 2007, pp. 1 – 5.

[70] T. S. RAPPAPORT "Wireless Communications: principles and practice" Prentice Hall 2002

[71] K.Won-Ik, L.Bong-Ju, S.Jae-Su, S.Yeon-Seung, K.Yeong-Jin, "Ping-Pong Avoidance Algorithm for Vertical Handover in Wireless Overlay Networks", vehicular technology conference VTC 2007, pp 1509-1512

[72] IEEE Standard for local and Metropolitan Area Networks, Part 21: Media Independent Handover Services", January 2009

[73] A. Rahman, U. Olvera-Hernandez, M. Watfa, H.W. Kim, "Transport of Media Independent Handover Messages Over IP", draft-rahman-mipshop-mih-transport-03, July 2007

[74] V. Mhatre, K. Papagiannaki, "Using smart triggers for improved user performance in 802.11 wireless networks", in Proceedings of the 4th international conference on Mobile systems, applications and services MobiSys'06, June 19-22, 2006, Uppsala, Sweden.

[75] 3GPP, “IP Multimedia Subsystem (IMS); stage 2 (Release 9)” TS23.228 V9.0.0, Dec. 2009

[76] Dutta, A.; et al, “A-IMS architecture analysis and experimental IPv6 testbed” International Conference on IP Multimedia Subsystem Architecture and Applications, 2007

[77] 3GPP, “3GPP system to Wireless Local Area Network (WLAN) interworking,stage 2 (Release 9)” TS23.234 V9.0.0, Dec. 2009

[78] WIMAX forum “WiMAX - 3GPP Interworking”, WMF-T37-002-R010v3. Nov 2008

[79] Chiba, T.; et al; “Performance Analysis of Next Generation Mobility Protocols for IMS/MMD Networks” IWCMC '08.6-8 Aug. 2008 .pp. 68 - 73

[80] Wong K., et al,"A Multilayered Mobility Management Scheme for Auto-Configured Wireless IP Networks". IEEE Wireless Communications, vol. 10, no. 5,pp. 62 - 69. Oct 2003.

[81] Wang Q. and Abu-Rgheff M. "Mobility Management Architectures based on Joint Mobile IP and SIP Protocols". IEEE Wireless Communications, pp 68-76, Dec 2006

[82] Munasinghe, K.S.; Jamalipour, A.: “A Unified Mobility and Session Management Platform for Next Generation Mobile Networks”, IEEE GLOBECOM '07. pp 4979 – 4983

[83] A. Udugama et al., “NetCAPE: Enabling Seamless IMS Service Delivery across Heterogeneous Mobile Networks,”IEEE Commun. Mag., vol. 45, no. 7, July 2007,pp. 84–91.

[84] Bellavista, P., Corradi, A., Foschini, L., “IMS-Compliant management of vertical handoffs for mobile multimedia session continuity”IEEE Commun. Mag., vol. 48, no. 4, April 2010, 114 - 121

[85] Dutta, et al, "An Experimental Study of Location Assisted Proactive Handover ", IEEE GLOBECOM, November 2007, Washington.

[86] Silvana, G.P.; Schulzrinne, H.; “SIP and 802.21 for Service Mobility and Pro-active Authentication”, 6th Annual Communication Networks and Services Research Conference, 2008. pp: 176–182

[87] Hyoung-Kee Choi, Limb, J.O., "A behavioral model of Web traffic" Seventh International Conference on Network Protocols, 1999. (ICNP '99) pp 327 - 334

[88] S. Haseeb, A. F. Ismail “Handoff latency analysis of mobile IPv6 protocol variations”. Computer Communications, v 30, n4, Feb 26, 2007, Nature-Inspired Distributed Computing, p 849-855.

[89] Polidoro, A.; Salsano, S.; Niccolini, S.; "Performance Evaluation of vertical handover mechanisms in IP networks", IEEE Wireless Communications and Networking Conference, WCNC 2008, pp 2783 – 2788

[90] M. Yang, K. Jung, A. Park, S. Kim, “Definitive Link Layer Triggers for Predictive Handover Optimization”, Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE, 11-14 May 2008 Pages: 2326 – 2330 Singapore

[91] Liebsch, M., Ed., Singh, A., Ed., Chaskar, H., Funato, D., "Candidate Access Router Discovery (CARD)", RFC 4066, July 2005.

[92] http://www.isi.edu/nsnam/ns/ , may 2008

[93] the network simulator NS-2 NIST add-on, IEEE802.21 model, NIST/ITL/ANTD/HSNTG – draft 1.0, january 2007

APPENDIX A : LIST OF PUBLICATIONS

- M. BOUTABIA, L.R. CARDENAS and H. AFIFI, "A Cross Layer Architecture for Real time Applications", (2007), Seventh International Workshop on Applications and Services in Wireless Networks, ASWN 2007, Santander, Spain, May 2007
- M. BOUTABIA, L.R. CARDENAS and H. AFIFI, "Client Driven QoS Adaptation for Multimedia Session Mobility", 1st IEEE international Global Information Infrastructure Symposium GIIS 2007, Marrakech, July 2007, pp. 141-145
- L.R. CARDENAS, M. BOUTABIA and H. AFIFI, "An Infrastructure-based Approach for Fast and Seamless Handover", Third International Conference on Digital Telecommunications ICDT 08, July 2008, Bucharest, Romania, Best paper award
- M. BOUTABIA and H. AFIFI “MIH-based FMIPv6 Optimization for Fast-moving Mobiles”, Third International Conference on Pervasive Computing and Applications, 2008. ICPCA 2008. , 6-8 Oct. 2008 Alexandria, Egypt, pp. 616 – 620
- L.R. CARDENAS, M. BOUTABIA and H. AFIFI, "A Cross-layer Mechanism Based on Dynamic Host Configuration Protocol for Service Continuity of Real-Time Applications", International Journal on Advances in Systems and Measurements, vol 2 no 1, year 2009. pp. 76-83
- M. BOUTABIA and H. AFIFI, “Performance Analysis of Mobile IP Variants in Vertical Handover”, The 12th International Symposium on Wireless Personal Multimedia Communications, WPMC’09 , September 7-10, Sendai, Japan

- M. BOUTABIA and H. AFIFI “Collaborative Handover Mechanism for Real time Services”, The 9th International Conference on ITS Telecommunications 2009. October 20-22 Lille France
- M. BOUTABIA, E. ABDELRAHMAN and H. AFIFI , “A hybrid Mobility Mechanism for Heterogeneous Networks in IMS”, in the proceedings of the 11th IEEE international conference on Multimedia and Expo (ICME) 2010, July 19-23, Singapore
- E. ABDELRAHMAN, M. BOUTABIA and H. AFIFI, “Video Streaming Security: Reliable Hash Chain Mechanism Using Redundancy Codes”, ACM 8th International Conference on Advances in Mobile Computing and Multimedia (MoMM) Paris, France, November 2010. pp: 69-76
- E.ABDELRAHMAN, M.BOUTABIA and H.AFIFI “Hash Chain Links Resynchronization Methods In Video Streaming Security: Performance Comparison” accepted for publication in Journal of Mobile Multimedia (JMM), JMM Vol.7 No.1&2. pp089-112 Rinton Press 2011
- M.BOUTABIA and H.AFIFI, “Towards Service Continuity in IMS Heterogeneous Access Networks”, accepted for publication in International Journal of Computer Networks and Communications IJCNC 2011
- M.BOUTABIA, L.R. CARDENAS, H.AFIFI, “SESSAMO: Session Mobility for Video Streaming Applications”, accepted for publication in International Journal of Ubiquitous Computing IJU 2011

APPENDIX B: LIST OF ACRONYMS

3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project 2
ABC	Always Best Connected
AP	Access Point
AR	Access Router
BU	Binding Update
BS	Base Station
CBR	Constant Bit Rate
CN	Correspondent Node
CoA	Care of Address
CIP	Cellular IP
CSCF	Call Session Control Function
CPU	Central Proceesing Unit
DAD	Duplicate Address Detection
DHCP	Dynamic Host Configuration Protocol
FA	Foreign Agent
FBU	Fast Binding Update
FMIP	Fast handovers for Mobile IP
GPRS	General Packet Radio Service
HA	Home Agent
HI	Handover Initiation
HIP	Host Identity Protocol
HMIP	Hierarchical Mobile IP
IDMP	Intradomain Mobility Management Protocol
IMS	IP Multimedia Subsystem
ISP	Internet Service Provider
LAN	Local Area Network

LCoA	Local CoA
LGD	Link Going Down
MAC	Media Access Control
MAP	Mobile Anchor Point
MIH	Media Independent Handover
MIP	Mobile IP
MMP	Mobility Management Protocol
MN	Mobile Node
MTU	Maximum Transfer Unit
MPEG	Motion Picture Expert Group
NAR	New Access Router
PAN	Personal Area Network
PAR	Previous Access Router
PoA	Point of Attachment
PLMN	Public Land Mobile Network
PMIP	Proxy Mobile IP
PrRtAdv	Proxy Router Advertisement
QoS	Quality of Service
RA	Router Advertisement
RCoA	Regional CoA
RO	Route Optimization
RR	Return Routability
RtsolPr	Router Solicitation proxy
SAP	Service Access Point
SDP	Session Description Protocol
SDPng	Session Description Protocol new generation
SIP	Session Initiation Protocol
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UNA	Unsolicited Neighbour Advertisement
WIFI	Wireless Fidelity

WIMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless LAN

APPENDIX B: RÉSUMÉ LONG

1.54. Introduction

Les réseaux de communication sans fil et mobile ont connu un énorme progrès et un succès remarquable au cours des dernières années. Le système de quatrième génération de télécommunication vise à fournir un accès large bande sans fil aux utilisateurs à tout moment et n'importe où. Les utilisateurs du 4 G auront la possibilité d'utiliser différentes technologies d'accès au réseau allant des réseaux étendu aux réseaux locaux sans fil comme WIFI. Le faible coût de déploiement et d'exploitation des réseaux locaux sans fil les rend très attrayant pour les fournisseurs de services et pour clients. Actuellement, la France est parmi les pays qui ont un grand nombre de points d'accès publics WIFI déployés avec plus de 30000 points d'accès en 2010. Par conséquent, il occupe la troisième place après les USA et la Chine. Ce nombre ne tient pas compte des réseaux offerts par certains fournisseurs de services Internet à leurs clients pour profiter de la connexion internet quand ils sont loin de la maison (ex: "Wi-Fi gratuit", "Neuf WiFi") en partagent la connexion avec d'autre utilisateurs du même opérateur. Le nombre total de points d'accès public Wi-Fi déployé dans le monde a atteint 310.000 en 2010, selon la même étude, avec un taux de croissance de 20%. En parallèle au développement du réseau, les appareils mobiles ont connu une transformation complète. Téléphonie mobile, assistant numérique personnel (PDA), tablettes Internet, les ordinateurs portables ... etc, ont acquis des capacités matérielles de plus en plus sophistiquées en termes de vitesse de traitement, de l'espace mémoire, des interfaces de communication et de l'espace de stockage. Ces capacités permettent aux appareils mobiles, non seulement de communiquer à travers différentes technologies de réseau, mais aussi de choisir le plus adéquat dans le cas de plusieurs réseaux disponibles; cette dernière caractéristique est connue sous le nom toujours mieux connecté. Cela signifie qu'à tout moment, le mobile doit être connecté au meilleur réseau disponible. "mieux" ici peut se référer à de nombreux critères comme le coût, débit, préférences de l'utilisateur, etc ... Le mobile doit décider quel réseau répondra aux besoins de ses applications à un moment donné. En outre, l'utilisateur peut même choisir

le dispositif à utiliser en fonction de sa situation. Transférer la session en cours entre les différents terminaux donne un degré de liberté élevé à l'utilisateur et réalise une véritable ubiquité du service. Néanmoins, la coexistence de multiples technologies d'accès au réseau pose le problème de l'interconnexion et la gestion de la mobilité à travers eux. Heureusement, la large utilisation de protocole IP dans tous les réseaux d'accès et de cœur permet aux utilisateurs de se déplacer entre les différents réseaux d'accès tout en profitant de leurs services. Toutefois, la croissance du multimédia et des applications en temps réel l'utilisation par les clients Internet impose des contraintes supplémentaires aux protocoles de gestion de la mobilité. Ce genre d'application nécessite une grande attention à maintenir le même niveau de qualité de service. Par exemple, dans les applications en temps réel, le délai de transfert doit être aussi court que possible pour garantir la continuité de service sans affecter la qualité d'expérience perçu par l'utilisateur.

1.55. Motivation

La Continuité de service est le défi le plus important que l'opérateur doit faire face afin d'offrir un service universel. Si l'hétérogénéité du réseau est bénéfique du point de vue utilisateur, cela complique davantage la tâche du fournisseur du service. D'une part, résoudre le problème de la continuité de service permettra aux opérateurs de diversifier leurs réseaux d'accès et de profiter de l'infrastructure à faible coût tout en conservant le même niveau de qualité de service. D'autre part, ils accordent plus de souplesse aux utilisateurs pour choisir leur réseau favori ou encore le transport de la session en cours à un autre terminal avec des capacités du matériel de meilleure qualité.

La nécessité de maintenir la continuité de service est particulièrement important dans les applications qui ont une certaine continuité dans le temps. Par exemple la navigation sur Internet est n'a pas besoin d'assurance de la continuité du service, car il s'agit d'une application discontinue: l'utilisateur demande une page web et attend la réponse commence alors la lecture des informations affichées. A l'inverse, lorsque l'application a lieu pendant un certain temps comme dans le cas de téléchargement ou le streaming vidéo le support de la continuité du service est obligatoire sinon le service s'arrête. La manière dont la mobilité est prise en charge est lié à l'application elle-même.

L'utilisateur ne sera pas affecté par un retard de transfert élevée dans le cas du téléchargement du fichier, car le résultat n'est pas perceptible jusqu'à la fin du téléchargement. Au contraire, tout retard excessif dans la réalisation de l'opération de mobilité aura un impact sur la qualité des applications sensibles aux délais telles que le streaming vidéo ou la voix sur IP. Bien que de nombreux travaux ont été menés dans le domaine de la mobilité, il reste toujours des problèmes qui ne sont pas complètement résolus.

1.56. Mobilité de session

L'objectif de la mobilité de session est de donner aux utilisateurs la possibilité de passer d'un terminal à un autre au cours de la même session multimédia sans aucune interruption. Cette opération s'est révélée utile pour les utilisateurs qui sont en mobilité. Un scénario typique de la mobilité de session est un utilisateur qui se sert de son assistant numérique personnel (PDA) pour regarder son jeu préféré quand il est dehors. Le PDA est connecté à Internet via le réseau mobile terrestre public, qui fournit des services 3G. Une fois chez lui l'utilisateur souhaite profiter de son accès à large bande et son écran haut définition pour regarder le même programme sans l'initialisation ou l'interruption du service. Le transfert de la session peut être commandé par le dispositif à partir duquel la session est transférée, ce mode est appelé mode «poussé», ou par le dispositif vers lequel la session est transféré, ce mode est appelé mode « tiré »l.

La continuité de service signifie qu'il ne devrait y avoir aucune interruption lorsque le flux est transféré entre les deux appareils. Par conséquent, la mobilité de session ajoute une contrainte temporelle à l'opération de transfert. Le délai du handover est la période de temps écoulé entre l'instant où l'utilisateur choisit de passer au nouveau dispositif en déclenchant le transfert (ex: appuie sur un bouton) et l'instant où le media commence à être joué dans le dispositif cible. En d'autres termes, le délai de transfert représente la réactivité du mécanisme de transfert. Dans le cas idéal il ne devrait y avoir aucun décalage de temps entre le moment où le média disparaît du premier dispositif et le moment où il apparaît à nouveau sur le dispositif cible. En pratique, le délai de transfert n'est pas nulle en raison du retard dans la transmission de paquets de signalisation et de paquets de données entre les entités concernées (i.e. serveur

multimédia, le dispositif cible et originaire). Néanmoins, le retard de transfert devrait être réduit au minimum afin de ne pas perturber la qualité de l'expérience de l'utilisateur. En outre, la continuité de service ajoute une autre contrainte liée à la synchronisation des médias entre les deux appareils. Pour réaliser un transfert de session précis, il faut que le flux commence dans le dispositif cible à partir de l'instant où il a été laissé dans le premier dispositif. Ce problème est une conséquence directe du délai de transfert. Résoudre ce problème est important afin de préserver la cohérence du service. Cela signifie que nous devons nous assurer que l'utilisateur ne manque aucune séquence du média et en même temps minimiser les chevauchements cela veut dire ne pas jouer une séquence déjà affiché dans le dispositif d'origine).

1.56.1. Adaptation du média

Un autre défi de la mobilité de session est l'adaptation des média étant transférés. Vu la variété des dispositifs et leurs capacités, il est obligatoire que les flux multimédia soient adaptés aux dispositifs auxquels ils sont destinés en termes de capacités et débit de connexion. Par exemple, le flux supporté par un écran haute définition ne convient pas à un appareil mobile sans aucune adaptation, sinon le système sera surchargé et la qualité sera médiocre.

La difficulté de l'adaptation des médias réside dans la façon d'adapter le flux multimédia contenus sur des serveurs à la grande variété d'écrans, capacités du processeur, débit des cartes réseau, la duré des batteries, etc… Ce processus implique un certain nombre de tâches pour lesquelles des procédures de signalisation ne sont pas complètement défini. Pour réaliser cette tâche, deux procédures sont distingués: la négociation et l'adaptation. La négociation est engagée par le client et a lieu au début de la session ou pendant la session lorsque l'un des paramètres change. L'adaptation est l'action menée par le serveur après la négociation afin d'apporter les changements nécessaires sur le flux servi.

1.56.2. Description du protocole proposé

SESSAMO (SESSion And MObility) est un protocole simple qui opère entre les deux dispositifs dans une relation d'égal à égal. Cela signifie que les deux appareils implémentent les deux modules serveur et client. SESSAMO est basé sur l'échange de

messages texte [44] tels que HTTP et RTSP. Deux principaux types de messages sont identifiés: les messages de requête et les messages de réponse. Le premier véhicule la commande, tandis que le second transporte le statut du résultat de l'opération. Le format général d'un message SESSAMO est le suivant:

Header 1 CRLF

Header 2 CRLF

...

Header n CRLF

Un en-tête est toujours composé de deux éléments: l'identifiant de l'entête et la valeur ou l'attribut. A l'intérieur du message, l'en-tête peut prendre n'importe quel ordre. D'autre part, il y a un nombre réduit d'entêtes afin que la longueur totale du message n'excède pas le Maximum Transfer Unit (MTU).

1.56.3. Implémentation

Nous utilisons la bibliothèque open source LIVE555 [45] sous linux pour le protocole RTSP et VLC [46] comme un lecteur vidéo sur les terminaux. LIVE555 est utilisé comme serveur vidéo et aussi inclus un module lors de la compilation de VLC pour le support client de RTSP. SESSAMO coordonne entre les deux entités participant à l'opération de mobilité SESSAMO: Le dispositif d'origine (DO) et le dispositif cible (DC). SESSAMO est un protocole d'égal à égal, pour cette raison un appareil qui prend en charge la mobilité session a deux types de programmes: un programme serveur qui écoute en permanence les demandes provenant d'autres terminaux, et un programme client qui permet de transférer la session du terminal actuel. Ainsi, SESSAMO fonctionne en mode « poussé ». Le programme client est équipé d'une interface graphique. Il comprend un ensemble de ressources graphiques pour établir les paramètres de fonctionnement. D'une part, il contrôle VLC pour accomplir les commandes telles que lecture, arrêt, pause, etc ... D'autre part, il récupère les rapports de d'état sur la session actuelle de la vidéo, comme l'instant du dernier RTP reçus. Cette information est utilisée plus tard dans l'opération de transfert afin de reprendre le média du meme endroit ou il a été dans l'ancien terminal. L'échange d'informations requis par le protocole est toujours commencé par DO. Cela se produit lorsque l'utilisateur appuie

sur la touche "transfert de session en cours" (voir figure 10). Cette action permet au protocole SESSAMO d'envoyer un message de demande à DC. Ce message contient les informations nécessaires pour récupérer la session vidéo. Si la DC ne reçoit pas une réponse avant RTT seconde, il retransmet le même message; on répète l'opération si nécessaire mais pas plus de sept fois. Après sept tentatives, la communication avec le DC est considérée comme impossible.

1.56.4. Scenario Make Before Break

L'approche « make before break » comme illustré à la figure 14 consiste à établir la session sur le DC avant que l'accusé de réception est renvoyé au DO. Accusé de réception ici signifie que le trafic de données commence à arriver et la session a été transféré avec succès. L'avantage de ce mode est la garantie que le flux est réellement reçu par l'DO, si ce dernier ne peut pas établir la session avec le serveur vidéo, l'accusé de réception ne sera pas envoyé et la session continuera sur le DO.

Tableau 17: résultats du make before break

Délai	valeur en ms
Délai du handover de la session	453
Chevauchement de la session vidéo	455
Temps de démarrage de VLC	442

Le tableau 1 résume les résultats de transfert de la session. Le délai de transfert est court et le transfert est presque instantané pour l'utilisateur. L'inconvénient de ce mode est que le chevauchement de la session vidéo est élevé. A noter que le temps de démarrage de VLC représente 97% du délai de transfert total.

1.56.5. Scenario Break Before Make

En scénario « break before make » (voir figure 15) l'accusé de réception est envoyé immédiatement après réception de la demande de transfert, puis la session prend fin sur le DO. L'avantage de cette approche est que le chevauchement vidéo est réduit par rapport au mode précédent comme le montre le tableau 2. Quant au délai de transfert, il est presque égal à scenario précédent puisque la principale cause de ce retard est temps de démarrage de VLC qui constitue plus de 97% du délai. L'inconvénient de

ce scénario est le temps de silence qui se produit pendant l'opération de transfert, en particulier lorsque le délai transfert est élevé.

Tableau 18: résultats break before make

Delay	value in ms
Session handover delay	499
Chevauchement de la session video	9
VLC starting time	487

1.56.6. Renégociation des paramètres de la qualité de service

Le défi maintenant consiste à faire face à la variété de dispositifs mobiles. En effet, nous pouvons trouver dans le marché des dispositifs différents qui ont différentes caractéristiques en termes de taille d'écran, de puissance de calcul, de systèmes d'exploitation, d'interfaces réseau, de codecs pris en charge, de la durée de batterie ... etc Un serveur offrant un service donné doit être capable de satisfaire chaque terminal en fonction de ses propres capacités. Les protocoles de signalisation tels que SIP et RTSP transportent dans leur message charge les contraintes imposées par les clients et de leurs capacités et le serveur transcode ou re-quantifie le contenu convenablement. Il est évident que l'adaptation du média est nécessaire en cas de mobilité session. Néanmoins, nous nous attaquons au problème de la gestion de la QoS dans la mobilité de session dans différentes phases. La première phase est la négociation qui a lieu au début de la session, la seconde est l'adaptation au cours de la session lorsque certains paramètres changent au sein du même terminal, comme le niveau de la batterie et la vitesse de connexion. La troisième phase est la renégociation lors du transfert de la session. La particularité de notre travail réside dans la manière ces tâches sont réalisées au moyen de SDPng/MPEG-21 et RTSP. En particulier, l'adaptation des paramètres de la qualité de service n'est pas fondée sur une approche classique sur lequel le serveur adapte le débit sans la participation du client. En effet, dans notre approche le client guide le serveur pour obtenir la qualité de service la plus approprié.

Il convient de noter que l'adaptation aux conditions du réseau n'est pas l'objet de notre étude. Si la qualité de la vidéo se dégrade en raison de la congestion sur la route vers le terminal, d'autres techniques de bout en bout doivent intervenir pour adapter les médias. RTCP [40] est l'une des solutions possibles pour contrôler le flux en utilisant les

rapports périodiques. Ces rapports sont renvoyés au serveur pour avoir une idée sur bande passante disponible le long de la route au terminal.

1.57. Analyse des protocoles de mobilité sur IP

Beaucoup d'appareils mobiles utilisent actuellement les réseaux IP pour accéder à une grande variété d'applications, y compris ceux qui ont besoin d'assurance de la tenue de la session lors du passage d'un réseau à l'autre. Le problème de la mobilité dans les réseaux IP vient du fait que ce protocole n'a pas été conçu pour gérer la mobilité. En effet, les protocoles Internet ne sont pas adaptés pour supporter les communications mobiles en raison de leurs principes d'adressage et de routage. Toute adresse de l'hôte doit être déduite de l'adresse réseau où il est physiquement attaché et aucun changement dans cette adresse au cours de la session n'est considéré. En vertu de ce régime, quand un mobile se déplace de son réseau d'origine à un réseau étranger, il rencontra au moins les problèmes suivants: 1) quand il atteint un nouveau réseau, toute communication devient impossible. Étant donné que son adresse n'est pas valide dans le réseau étranger, il ne peut être acceptée ni par les nœuds étrangère ni routeurs correspondants. L'Obtention d'une nouvelle adresse valable dans le réseau étranger est alors nécessaire. 2) Les associations de la communication en cours sont perdus à cause de incohérence de l'adresse, 3) le nœud mobile disparaît du réseau global. Normalement, les hôtes se retrouvent dans le réseau au moyen d'un répertoire de localisation (RL). Il s'agit d'une base de données distribuée contenant le nom d'hôte et son adresse IP, un exemple de cette base de données est le fameux service DNS (Domain Name System). Pour rester en contact avec le réseau global, le mobile doit informer le RL chaque fois qu'il acquiert une nouvelle adresse IP.

Afin de faire face aux limitations de l'IP dans les communications mobiles, un certain nombre d'approches ont été proposées. Bien qu'elles abordent le problème sous des angles différents, ils sont d'accord sur la façon dont ils y font face. En effet, les principales approches reposent sur un certain nombre de procédures qui peuvent être classés en: détection de mouvement et l'allocation des adresses, redirection de trafic, mise à jour de la localisation globale et le lissage du handover. Le tableau suivant

recense les protocoles de mobilité sur IP les plus connus et les classifie selon les étapes cité auparavant :

Protocol de mobilité	*SIP*	*MIPv4*	*MIPv6*	*PMIPv6*	*TCP-migrate*	*HIP*
couche	**application**	**réseau**	**réseau**	**réseau**	**transport**	**transport**
Detection de movement et allocation d'adresses	**DHCP , PDP**	**FA advertisement (FA-CoA) – DHCP (co-located CoA)**	**Neighbor discovery, autoconfiguration, DHCPv6**	**Home network prefix allocation and auto-configuration or DHCP**	**DHCP**	**DHCP**
Redirection du Traffic	**Re-invite method**	**Route Optimization,** **Reverse tunneling**	**Route Optimization with RR,** **Reverse tunneling**	**Proxy binding update**	**End to end**	**HIP update**
Lissage du handover	**Bicast**	**FMIPv4**	**FMIPv6**	**FMIPv6**	**-**	**-**
Agent de localization global	**SIP registrar**	**Home agent**	**Home agent**	**Local mobility anchor**	**-**	**Rendez vous server**

1.58. Mécanisme collaboratif de handover

Les réseaux de prochaine génération ont l'intention d'offrir des services omniprésents et ubiquitaires aux clients où qu'ils soient et quelle que soit l'application qu'ils utilisent. Pour atteindre cet objectif, la continuité de service sur des réseaux homogènes et hétérogènes doit être garantie. Grâce à l'IP mobile et ses variantes, la continuité du service dans les réseaux tout IP peut être assurée. MIPv6 a été conçu par l'IETF pour fournir des nœuds mobiles avec la possibilité de maintenir la communication en cours lors de la modification du réseau d'accès et l'adresse IP de MN. Bien que MIPv6 présente quelques améliorations par rapport à MIPv4, son retard de transfert à long rend impropre pour les applications temps réel. FMIPv6 tente de réduire

ce délai en utilisant la couche de liaison des déclencheurs pour effectuer l'acquisition d'adresses avant de la couche 2 de transfert. En outre, FMIPv6 empêche la perte de paquets en créant un tunnel entre le routeur d'accès précédente (PAR) et le Nouveau routeur d'accès (NAR) pour transmettre les paquets jusqu'à ce que le nœud correspondant et l'agent d'accueil mise à jour le soin d'adresse (CoA). Toutefois, FMIPv6 montre certaines limites dans les terminaux en mouvement rapide. D'une part, le terminal n'a pas assez de temps pour échanger tous les messages de PAR dans la phase d'initiation. D'autre part, routeurs d'accès nécessite une taille de tampon pour amortir grands paquets envoyés à la MN entre le moment de la transmission FBack et le moment de l'avertissement UNA, respectivement par la RAP et NAR.

Dans ce travail nous utilisons les services MIH plus efficacement aux différentes étapes du FMIPv6, non seulement dans l'étape de découverte. En outre, en utilisant des indicateurs d'événement le mobile sera servi jusqu'à sa déconnexion physique, et dès qu'il se connecte au nouveau réseau, les paquets seront transmis par le nouveau routeur d'accès, sans attendre l'annonce de l'attachement FMIPv6. Par conséquent la taille mémoire dédié aux paquets dans le NAR est réduite. Un autre avantage de cette proposition est la résilience à l'effet ping-pong. En effet, le handover est initiée par le MN, mais finalisé par le réseau permettant l'annulation du handover s'il est déclenché par une "fausse alerte".

Nous montrons à travers l'analyse d'analyse et de simulation, que notre régime réduit transfert retard ainsi que la perte de paquets dans les deux relèves prédictive et réactive. En outre, la taille du tampon nécessaire pour le tampon de paquets dans les routeurs d'accès lors de la remise est plus petite. Enfin, le mécanisme proposé évite des événements de ping-pong excessifs entre les réseaux sans fil.

1.58.1. Limitations de FMIPv6

L'avantage de FMIPv6 réside dans son mode d'opération prédictif. Toute défaillance dans l'exécution de ce mode conduit à un mode réactif qui ajoute plus de retard au handover. Par conséquent, en cas de mobile en mouvement rapide, il est probable qu'il ne reste pas connecté au réseau précédent assez longtemps pour envoyer et recevoir tous les messages FMIPv6. Ainsi, le terminal devrait achever l'étape

d'initiation en une courte période de temps afin de ne pas perdre aucun message FMIPv6. En fait, le message de sollicitation est envoyé après le déclenchement du handover au niveau 2 ce qui entraine une période d'initiation plus longue. En outre, si la connexion est interrompue avant la réception du message FBack, tous les paquets seront perdus jusqu'à ce que la connexion au nouveau réseau soit établit et le tunnel entre NAR et PAR est crée. Le retard important qui prolonge la période d'interruption du service est comprise entre le moment ou MN reçoit FBack et le moment de la dissociation de l'ancien réseau, tous les paquets destinés à MN au cours de cette période sont transmises à NAR mise en mémoire, alors que MN peut encore recevoir ces paquets sur l'ancien lien. En fait FMIPv6 propose également d'entamer la mise en mémoire de paquets et de les transmettre en même temps lors de la réception du FBU avant même la création du tunnel. Cette solution est bonne pour s'assurer que tous les paquets adressés à MN pendant l'opération FMIPv6 seront reçus soit dans l'ancien ou le nouvel emplacement. En fait, cette solution peut être considérée comme une solution bicast abordées au chapitre 4, où le trafic est envoyé à la fois à l'ancienne et à la nouvelle adresses. Néanmoins, cette solution a au moins deux inconvénients: elle augmente l'overhead et a besoin d'un mécanisme dans le système d'exploitation du MN pour éviter la duplication de paquets au cas où elle a lieu.

1.58.2. Media Independent Handover

IEEE 802.21 ou MIH (Media Independant Handover) est un nouveau standard qui a été approuvée par l'IEEE au début de 2009. Il fournit des renseignements sur la couche de liaison et d'autres informations réseau aux les couches supérieures afin d'effectuer des handovers optimisé entre les réseaux hétérogènes. Les technologies de réseaux couverts par MIH sont à la fois des technologies IEEE et non IEEE comme les réseaux 3GPP. Le but de cette norme est d'améliorer l'expérience des utilisateurs mobiles en facilitant le handover entre réseaux hétérogènes.

1.58.3. Procedure du Handover Collaboratif

FMIPv6 entend protéger les paquets de données contre la perte. Néanmoins, l'envoi du message FBU après l'événement link going down empêche MN de recevoir les paquets de données, car ils sont encapsulés et envoyés vers NAR, même si MN est

encore capable de recevoir des paquets. Par conséquent, nous avons besoin de déclencher la création de tunnel par un autre mécanisme qui est séparée de la transmission du FBU. Dans ce travail, nous proposons un mécanisme collaboratif qui implique à la fois le terminal et le réseau. Le routeur d'accès et le point d'attachement utilisent les services MIH pour donner au terminal la possibilité de poursuivre sa communication dans le réseau précédent, même après l'évènement LGD. L'événement link down est le seul événement qui doit déclencher la transmission de paquets via le tunnel. Un tel événement est également détectable au niveau du point d'attachement qui va informer le nouveau routeur d'accès pour commencer la transmission des paquets à travers le tunnel. Préserver les paquets de la session en cours est la première priorité que nous avons l'intention de réaliser, avant toute autre procédure.

Nous considérons le scénario de la mobilité montré à la figure 24. MN est relié au réseau de service et connecté au serveur d'information. L'emplacement du serveur de l'information peut être fourni par un serveur DHCP dans la réponse à la requête de MN lors de sa première connexion au réseau. La fonction MIH commence avec l'opération d'initialisation par découvrir les entités du réseau qui fournissent des services MIH et leurs capacités, puis elle s'enregistre avec d'autres MIHF et souscrit à des événements localement ou à distance. MN demande au serveur d'informations des informations sur les réseaux voisins en envoyant la requete MIH_Get_Information (voir figure 25). La réponse contient des couples (AP-ID, AR-Info). Ainsi MN connaît ses réseaux voisins longtemps avant le handover, et non après le déclenchement, comme c'est le cas dans FMIPv6 classique. Une fois que le signal devient faible, la couche MIH est notifiée par l'événement Link_going_down, puis procède à l'exécution de l'algorithme de décision pour décider vers quel réseau il faut faire le handover, basée sur la réponse du serveur d'informations déjà disponibles.

Figure 26 explique les procédures de souscription et l'indication d'événements dans les deux entités locales et distantes. FMIPv6 doit inclure l'adresse MAC du MN dans le message MIH de souscription en utilisant l'adresse source du message FBU. Ainsi, MIHF au PoA peut filtrer l'événement Link down de ce MN spécifique. L'événement Link down peut être notifié par le PoA lorsque la connexion est perdue soit explicitement par la procédure de dissociation ou après l'expiration des

temporisateurs des accusés de réception. Après avoir reçu HI, NAR peut effectuer la détection de duplication d'adresse pour les futurs CoA, et en même temps, il souscrit à link_up au nouveau PoA. Après le handover au niveau liaison, MN se connecte au nouveau réseau et le PoA notifie cet événement à NAR, par conséquent, NAR commence la transmission des paquets de MN sans attendre le message UNA. La principale diffère du mécanisme classique par le fait que PAR ne commence à rediriger les paquets que si il reçoit une notification Link_down du PoA, d'autre part les paquets sont transférer du NAR dès que MN se connecte au nouveau réseau.

1.58.4. Simulation

Le mécanisme proposé est évalué en utilisant NS2 (ns2.29) [92] avec le module MIH qui a été développé par le National Institute of Standards and Technology (NIST) [93]. Nous avons développé le module FMIPv6 comme étant un nouvel agent dans le nœud mobile et les routeurs d'accès. En outre MIH est modifié pour prendre en compte les améliorations proposées. Afin d'évaluer MIH-FMIPv6 par rapport à FMIPv6 telle qu'elle est décrite dans [65] (la résolution d'adresse inverse est faite localement à l'aide du service d'information MIH), nous considérons le scénario suivant: MN est relié avec l'interface WIFI et se déplace hors de la zone WIFI et entre au secteur WIMAX. MN reçoit un trafic de type CBR (débit binaire constant) depuis un nœud fixe.

délai du handover :

Le délai du handover est la période de temps séparant la réception du dernier paquet sur l'interface WIFI et le premier paquet reçu sur l'interface WIMAX. Figure 30 représente les résultats de cette simulation. La simulation est exécutée pour différentes valeurs du coefficient LGD avec et sans l'amélioration proposée. Les résultats du délai du handover montrent que pour le protocole FMIPv6, le délai de transfert augmente avec l'augmentation du coefficient LGD. A l'inverse, le délai du handover avec MIH-FMIPv6 n'est pas affecté par le niveau du coefficient LGD. Ce résultat est conforme à nos objectifs car le lien est encore utilisable même après un événement LGD. Le délai du handover élevé est dû au retard de la connexion de l'interface WIMAX

Taille de la mémoire :

Les paquets qui sont transmis vers le nouvel emplacement du MN doivent être mis en mémoire dans le NAR puisque le délai de handover au niveau de la couche 2 est trop

long. la taille de mémoire dédié à ces paquets dépend du délai de transfert ci-dessus. Ceci est approuvé par l'allure des courbes de la figure 31 qui suivent le même comportement que dans le délai de handover. FMIPv6 a besoin d'une taille mémoire qui augmente avec l'augmentation du coefficient LGD car il couvre plusieurs paquets de données. Quant à MIH-FMIPv6 la taille de mémoire est réduite et reste constante lorsque le coefficient LGD augmente.

Perte de Paquets :

La figure 32 montre le nombre de paquets perdus durant le handover pour différentes valeurs du coefficient de LGD. En FMIPv6 la perte de paquets est nul parce que les paquets sont transmis par le tunnel et mises en mémoire dans le NAR. Quant à MIH-FMIPv6, 8 paquets sont perdus quelle que soit la valeur du coefficient LGD. La perte de paquets dans notre solution est justifiée par la transmission de paquets en file d'attente de l'interface. Afin de réduire la perte de paquets une solution consiste à récupérer les paquets restants dans la file d'attente et les transmettre au NAR. Dans notre proposition le handover n'est pas systématique après l'envoi de message FBU, mais ne devient effective qu'après une déconnexion physique du terminal. Le temps supplémentaire au cours duquel MN est relié au PoA peut corriger une erreur de détection de mouvement causée par la fluctuation rapide de la force du signal.

Effet Ping Pong :

La force du signal reçu (RSS) est utilisée en tant que critères pour choisir le meilleur réseau, trois seuils sont définies pour chacune de ces technologies: i) LGD est le seuil à partir duquel le processus FMIPv6 est déclenchée par le moyen de message FBU; ii) ThrE est le seuil d'entrée au réseau, il correspond ainsi à l'événement link_up; iii) ThrO est le seuil de sortie et correspond également à la notification link_down. Figure 33 présentes l'algorithme qui permet d'éviter les handover inutiles en raison de la faiblesse temporaire de l'intensité du signal. L'avantage d'utiliser les services MIH est le contrôle du protocole FMIPv6. Toute amélioration dans le niveau du signal après une indication LGD désactive l'ensemble du processus par l'annulation des inscriptions aux événements dans les deux PAR et NAR. Ce mécanisme d'annulation peut être mis en œuvre au moyen de temporisateur, quand expiré, l'inscription à l'événement est supprimée.

1.59. Maximisation du mode predictif de FMIPv6

MN considère que le mode opératoire prédictif a échoué une fois qu'il ne reçoit pas le message FBack dans l'ancien lien, même si la transmission du FBU a réussi et le tunnel entre le PAR et NAR a été créé. Afin d'augmenter la probabilité du mode prédictif, nous proposons de transmettre FBack à MN dès qu'il se connecte au nouveau réseau (voir Figure 34). Dans FMIPv6 ordinaire, le transfert et la mise en mémoire de paquets de données est lancé sans aucune assurance de la position actuelle du MN, cela peut conduire à la perte du FBack si MN n'est pas plus attaché à l'ancien lien. Dans la proposition présentée dans le chapitre précédent, PAR reste informé de la position de MN au moyen des événements MIH à distance. Suite à l'emplacement de MN, FBack est livré, si il est encore à l'ancien lien, le message sera livré directement, si MN est déjà dans le nouveau réseau FBack sera transmis via le tunnel.

Dans cette amélioration, MN collabore avec le réseau pour parvenir à un handover sans couture. Toutefois, MN doit être conscient de son état tout le temps pour avoir une opération de handover cohérente. Pour clarifier ce point, nous prendrons l'exemple d'un MN qui a perdu sa connexion avec le PAR immédiatement après l'envoi d'un FBU. Bien que FBU a réussi et le tunnel a été crée, MN suppose un échec du mode prédictif et active le mode réactif après l'établissment du lien avec le nouveau lien. Afin de surmonter cette contradiction, nous proposons un nouvel algorithme (voir Figure 35) pour que MN soit en phase avec le réseau. MN suivra le meme système de retransmission que dans le FMIPv6 ordianire après avoir reçu le déclencheur LGD. En attendant le message FBack, le temporisateur de backoff sera doublé chaque fois le FBU est retransmis. Si la liaison est coupée sans recevoir FBack sur l'ancien lien, MN ne peut pas être sûr de la livraison du message FBU après la dernière retransmission. En fait, MN ne peut pas savoir si la non réception de FBack est causée par la perte de ce dernier, ou parce que FBU n'a pas été reçu correctement par le PAR en premier lieu. Par conséquent MN persiste à retransmettre FBU jusqu'à ce que FBack est reçu. Afin de ne pas répéter l'échange HI / Hack entre le PAR et NAR puisque le tunnel était déjà établi comme FBU a été bien reçu par PAR, nous proposons d'envoyer FBack localement à partir du PAR sans aucun nouvel échange avec le NAR si la CoA prospective dans le

message FBU n'a pas été modifié. Ainsi, PAR ignore le FBU envoyée ultérieurement par MN et renvoyer le FBack de nouveau. Après un handover au nouveau de la couche liaison sans réception du FBack dans l'ancien lien, MN envoie un message UNA à NAR et attend que FBack soit retransmis du NAR. Ce temps d'attente donne l'occasion au message FBack pour atteindre MN avant qu'il envoie un FBU à partir de la nouvelle liaison. si FBack n'est pas reçu dans un délai égal au backoff, MN conclut que FBU n'a pas été transmis correctement et par conséquent le tunnel n'a pas été établi, ce qui donne lieu à l'activation du mode réactif. Afin d'explorer l'impact de la durée estimée avant que le lien est coupé par rappor au nombre de retransmission possible du message FBU, nous prenons les valeurs suivantes pour le backoff et le délai MN-NAR respectivement: backoff = 100ms, DMN-NAR = 50 ms .

Dans la figure 37, nous pouvons observer que le temps de connexion restant est un paramètre très contraignant puisque même pendant 3 secondes de temps restant, MN a le droit de retransmettre le FBU seulement 3 fois avant que le lien soit coupé. Pour cette raison, MN doit réussir la transmission de FBU depuis la première tentative.

La figure 38 montre la probabilité de retransmission de FMIPv6 [65] et notre FMIPv6 amélioré (E-FMIPv6). la retransmission est un échec de FBU par conséquent la non mise en place du tunnel entre le PAR et le NAR. Chaque fois FBU est retransmis dans lest une étape vers le mode réactif, puisque le nombre de retransmissions est délimité par le temps LGD-LD. Par conséquent, la réduction de la probabilité de retransmission de FBU conduit à augmenter la probabilité de réussite du prédictif.

1.60. Mobilité dans l'IMS

Dans ce chapitre, nous abordons la question de la mobilité dans IP Multimedia Subsystem (IMS). Bien que IMS a été conçu pour intégrer les différents réseaux d'accès, la gestion de la mobilité entre ces réseaux n'est pas encore résolu. Nous proposons un nouveau mécanisme hybride pour la gestion de la mobilité, basé sur une coopération étroite entre FMIPv6 et SIP pour assurer un service continu en temps réel. En outre, MIH est intégré dans l'architecture IMS afin d'effectuer des handovers intelligents et précis aussi bien horizontale que verticale. Nous étudions deux cas de handover: le handover choisi et le handover forcé. Les handovers choisi ont lieu lorsque l'équipement

utilisateur (UE) est connecté au réseau via deux interfaces en même temps et décide de mettre à niveau la qualité de sa connexion selon un critère donné (à savoir le coût, la bande passante, etc ...) sans avoir aucune difficulté dans le lien précédent. Quant au handover forcé, il se produit lorsque le signal atteint un niveau critique et MN est forcé de faire un handover afin de maintenir la communication en cours. Ce cas est géré en deux phases. La première phase ou la phase rapide est assurée par le protocole FMIPv6 qui préserver le plus tôt possible les paquets de la communication en cours. La seconde phase ou la phase lente est géré par le protocole SIP pour acheminer les paquets et optimiser la livraison. Avec cette méthode, nous exploitons les avantages des protocoles de mobilité de la couche réseau et les protocoles de mobilité de la couche applications afin d'assurer une session continue à travers des deux réseaux sans imposer de nouveaux éléments au le réseau. Grâce à une comparaison avec d'autres mécanismes de mobilité, nous montrons dans l'analyse analytique que notre système hybride présente de meilleurs résultats en termes de latence du handover et de perte de paquets.

1.61. Comparaison des variantes MIP

Mobile IP et de ses variantes présentent une bonne solution pour la mobilité dans tous les réseaux IP. Choisir une variante ou l'autre dépend de nombreux paramètres et conduit à des résultats différents en termes de performances de handover. Cette décision est cruciale quand il s'agit d'applications temps réel telles que la Voix sur IP et le streaming vidéo. Pour maintenir la continuité de service lors du déplacement, le mobile doit effectuer le handover horizontal lorsque le réseau adjacent est de même technologie ou un handover vertical lorsque le nouveau réseau est de technologie différente du précédent. FMIPv6 est souhaitable pour les réseaux homogènes où le MN doit effectuer un handover dur. Dans ce cas, FMIPv6 préserve les communications en cours de perdre des paquets et réduit le handover au niveau de la couche 3 en préparant l'adresse IP avant le handover niveau 2. Néanmoins, dans des réseaux hétérogènes où l'utilisation de deux ou plusieurs interfaces, un protocole MIPv6 classique a également la possibilité d'acquérir une nouvelle adresse IP sur la nouvelle interface en utilisant le protocole classique de découverte de voisin (neighbor discovery). L'hétérogénéité des réseaux donne aux mobiles un potentiel important d'effectuer un handover sans couture car nous

avons à notre disposition deux interfaces qui ont la possibilité de fonctionner simultanément et indépendamment. En fait, la connexion peut être établie dans le nouveau lien avant d'etre dissocier du point d'attachement actuel. Ce type de transfert est connu comme make before break. Néanmoins ce type de handover n'est pas systématique, un déclenchement tardif peut conduire à une perturbation lors de la transition . Par conséquent, le choix du protocole de mobilité de couche 3 a un impact direct sur la performance du handover qui en résulte. Bien que les deux réseaux sont géographiquement proches les uns des autres, ils pourraient être topologiquement éloignés les uns des autres. Un exemple de cela est le cas de deux réseaux sans fil adjacents qui appartient aux différents opérateurs. Les paquets routés entre les deux réseaux devraient suivre la route établie par le système autonome conformément à une politique particulière. Dans ce chapitre, nous concentrons notre intérêt sur l'utilisation de MIPv6, HMIPv6, et FMIPv6 dans un environnement hétérogène. Nous montrons que les performances des protocole de mobilité dépend essentiellement de: 1) le temps disponible avant que le lien avec le réseau précédent est coupé, 2) le délai entre le réseau visité et le réseau domestique ou le délai entre le réseau visité et le noeud correspondant dans le cas de l'optimisation de la route et, enfin, 3) le délai entre le réseau d'accès précédent et nouveau réseau d'accès. Lorsque MN a la possibilité d'effectuer un make before break, MIPv6 et HMIPv6 semble être les meilleur protocoles de niveau 3 à utiliser. Toutefois, lorsque le handover est déclenché par un évanouissement grave dans le signal reçu, alors le choix dépendra du temps d'opération du protocole avant le lien est coupé. L'évaluation de performance prend en compte la latence du hadover, temps de perturbation et la perte de paquets. Enfin, nous indiquerons les directives pour le choix du protocole avec les meilleures performances en fonction des résultats numériques.

1.61.1. Resultats numériques

Dans la figure 52, on observe que lorsque le MN n'est pas très loin du réseau domestique MIPv6 effectue un handover plus court que FMIPv6. Cela est dû au faible nombre de messages échangés dans le mode tunnel inverse de MIPv6. Mais lorsque

l'optimisation de route est utilisée, le retard de fonctionnement du protocole est très élevé aussi bien pour MIPv6 que HMIPv6.

La figure 53 montre que FMIPv6 est sensible au délai entre le nouveau réseau d'accès et l'ancien réseau d'accès. Inférieur certaine limite FMIPv6 a un délai d'opération plus court que MIPv6, mais au delà de cette limite, en particulier lorsque le réseau domestique est à la même distance des réseaux d'accès ancien et nouveau, MIPv6 dépasse FMIPv6. En outre lorsque le délai PAR-NAR devient extrêmement élevé l'utilisation d'optimisation de la route peuvent également être envisagés.

Dans la figure 54, curieusement nous remarquons que MIPv6 et HMIPv6 même avec l'optimisation des routes ont une chance d'effectuer un handover sans couture avec un délai nul, ce que FMIPv6 ne peut assurer quel que soit le temps de connexion restant. Cela est dû à la communication entre les deux interfaces et à la transmission des paquets de données via des routeurs d'accès au lieu d'être transmis directement dans le cas des autres protocoles.

1.61.2. Discussion

Le tableau ci-dessous donne quelques directives qui devraient aider à choisir le protocole de mobilité à utiliser en fonction de l'événement déclencheur et le délai entre les entités intervenant dans le handover du niveau 3.

Tableau 3: directives pour le choix du protocole de mobilité

Varinate du protocole	Quand utiliser?
MIPv6	Evénement Link up ou link detected, delai NAR-HA court, temps Lgd-Ld moyen
MIPv6 avec RO	Evénement Link up ou link detected, temps Lgd-Ld long
HMIPv6 intra domain mobility	Événement Link up ou link detected, délai NAR-MAP court, temps Lgd-Ld moyen
HMIPv6 inter domain mobility	Evenement Link up ou link detected, temps Lgd-Ld long
FMIPv6	Evenement Link going down, délai NAR-PAR court, $t_{Lgd\text{-}Ld}$ court

1.62. Conclusions

Dans cette thèse, différents aspects de la continuité de service ont été étudiées. La mobilité de session est un aspect important de la continuité de service qui donne la possibilité à l'utilisateur plus de liberté. Pour atteindre cet objectif dans un contexte de streaming vidéo, nous avons proposé SESSAMO comme un protocole de transfert de session entre les terminaux. Les principales caractéristiques de ce nouveau protocole sont sa légèreté et son caractère pair à pair qui le rend facile à déployer et à utiliser. Ce protocole a été validé par une implémentation dans un dispositif mobile commerciale (Nokia 770) et les résultats du banc d'essai sont très encourageants. La renégociation des médias qui constitue un effet direct de la mobilité é de session a été également abordée. Afin d'adapter les médias à la capacité du nouveau terminal, il devrait y avoir un moyen de renégocier le format adéquat du média avec précision. Comme SDP est incapable de réaliser cette tâche, nous avons proposé d'utiliser SDPng avec MPEG21-DIA, afin de décrire le contexte du terminal avec précision. Différents scénarios ont été définis lorsque la renégociation est nécessaire et des messages spéciales RTSP ont été spécifiés pour transmettre la description selon la situation de l'utilisateur.

Dans un deuxième temps, la mobilité de terminal a été abordée. Après une étude bibliographique nous avons établi une classification des opérations de mobilité que la majorité des protocoles de mobilité sur IP suivent pour garantir la continuité de service à travers des sous-réseaux différentes. Nous avons déterminé quatre opérations qui ne sont pas forcément présents dans tous les protocoles de mobilité. Cette classification est importante pour analyser et diagnostiquer tout protocole de mobilité et en déterminer les limites. Ensuite, nous avons proposé un nouveau mécanisme de handover basé sur la collaboration entre le mobile et le réseau en utilisant les services de MIH. L'innovation de ce mécanisme réside dans l'utilisation efficace des événements MIH dans le mobile et le point d'attachement. les résultats de l'étude analytique et de la simulation ont montré que la solution proposée est efficace en termes de délai de handover et la taille de l'espace mémoire réservé dans les routeurs d'accès. En outre, cette solution présente une résistance à l'effet du ping pong. Dans le même objectif, une modification mineure dans le protocole FMIPv6 permet de maximiser la probabilité du mode prédictif. Un

simple transfert du message FBack à travers le tunnel entre les routeurs d'accès peut éviter au mobile de tomber dans le mode réactif de FMIPv6.

IMS est un exemple concret où la continuité de service représente un maillon important pour réaliser l'objectif de convergence visé par IMS. Bien que d'importantes mesures soient déjà réalisées en matière de convergence des réseaux d'accès, un handover sans couture entre les réseaux hétérogènes n'est toujours pas résolu. la solution proposée consiste à apporter une opération de lissage lors de la mobilité IMS par l'introduction de FMIPv6, conformément à la classification fait auparavant. La combinaison entre SIP et FMIPv6 permet non seulement d'éviter la redondance dans le réseau, mais aussi d'améliore les performances du handover. En outre, l'amélioration MIH peut être également appliquée dans le cas d'IMS. L'évaluation des performances de la solution proposée a été effectuée et comparée aux autres propositions. Les résultats montrent que notre système hybride permet de réduire le délai de handover et le taux de perte. Cependant, ces résultats dépendent étroitement de la réussite du mode opératoire prédictif de FMIPv6.

Enfin, nous avons mené une étude comparative entre les variantes de MIP dans le cas du handover vertical. Cette étude démontre que le choix du protocole de mobilité le plus optimale doit respecter certaines conditions. Après avoir souligné la spécificité du handover vertical et le potentiel apporté au nœud mobile par rapport au handover horizontal, nous avons comparé les performances des variantes MIP dans différents scénarios. Cette étude a prouvé que la présence de multiples interfaces dans le nœud mobile permet aux protocoles de mobilité classiques comme MIPv6 d'achever des handovers plus rapidement que FMIPv6 dans certains cas. En fin des directives qui devraient aider à choisir le meilleur protocole de mobilité sont fournis.

Perspectives:

Une solution de mobilité unique pour les connexions TCP et UDP est encore une problématique non résolu. Une manière possible de résoudre ce problème est l'utilisation de tunnels temporaires et une communication cross layer complète, comprenant la couche 2, la couche 3 et l'application côté client. En outre, la mise en mémoire tampon des paquets de donnés est une opération nécessaire à la continuité de service sans coupure, néanmoins la transmission des paquets tamponnée peut causer des

pertes si le taux de transmission est plus élevé que ce que le nouveau lien peut prendre en charge. Une politique de purge de la mémoire tampon est nécessaire pour éviter toute perte ou perturbation au niveau des utilisateurs.

Bien que la sécurité n'ait pas été abordée dans cette thèse, elle doit être prise en considération afin de sécuriser toutes les opérations de mobilité aussi bien pour mobilité de session que pour la mobilité des terminaux. En outre les mécanismes de sécurité doivent être conçues de telle manière ne pas compromettre les améliorations apportées aux protocoles de mobilité.