Secure mobile multimedia over all-IP wireless heterogeneous networks

A Doctoral Thesis

presented to

to the advising and examining committee

In partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of the department of Information and Communication Systems Engineering

of the University of the Aegean

Giorgos Karopoulos

ADVISING COMMITTEE

OF THIS

DOCTORAL THESIS:

Stefanos Gritzalis , Supervisor

Department of Information and Communication

Systems Engineering

Konstantinos Lambrinoudakis, Advisor

Department of Information and Communication

Systems Engineering

Spyridon Kokolakis, Advisor

Department of Information and Communication

Systems Engineering

University of the Aegean

January 2009

APPROVED BY

THE EXAMINING

COMITTEE:

Stefanos Gritzalis

Professor, University of the Aegean

Konstantinos Lambrinoudakis

Assistant Professor, University of the Aegean

Spyridon Kokolakis

Assistant Professor, University of the Aegean

Georgios Kambourakis

Lecturer, University of the Aegean

Elisavet Konstantinou

Lecturer, University of the Aegean

Dimitrios Lekkas

Lecturer, University of the Aegean

Konstantinos Oikonomou

Lecturer, Ionian University

University of the Aegean

January 2009

ABSTRACT

The specification of IP Multimedia Subsystem (IMS) in the 3rd Generation (3G) of mobile communications signifies the importance of multimedia delivery in future wireless systems. The introduction of a separate subsystem responsible exclusively for multimedia session management is a result of the special characteristics and requirements of multimedia applications; while these applications can tolerate a certain amount of data loss, they cannot tolerate delayed delivery of data. IMS, however, is not only limited to mobile communications since it is IP-based and its core protocols, namely SIP and Diameter, can be found on the Internet world as well. Moreover, current trends reveal that in the near future the convergence of most wireless systems will become reality over a common platform; that is IP, the Internet Protocol. In this heterogeneous network architecture multimedia delivery solutions will converge as well, since they will use the very same protocols; SIP and Diameter. While this convergence will create numerous possibilities for multimedia applications, it will also create security threats since end users will have a large number of operators to interact with in this multidomain environment.

This thesis focuses on security for multimedia delivery over all-IP wireless heterogeneous networks. A number of issues are analyzed and defeated by proposed mechanisms that can operate under very demanding circumstances, e.g. while a mobile node handoffs to a new cell and/or administrative domain. These vulnerabilities and proposed mechanisms are related with the signaling phase of multimedia delivery, i.e. before the actual delivery of multimedia takes place. The effectiveness and efficiency of the proposed mechanisms are qualitatively and quantitatively evaluated through appropriate comparisons and experimental testbed setup.

The vulnerabilities analyzed in this thesis concern the privacy protection of end users when they roam through different administrative domains which are generally considered unknown or untrusted or both. The mechanisms proposed here can be perceived as two modules which can be used either in conjunction or individually. The first one includes two privacy enhanced secure handoff optimization schemes suitable for wireless heterogeneous networks which protect end user's privacy while transferring context information necessary for fast re-authentication and service re-establishment to candidate networks. This context can carry almost every type of information whether this is related to security material or other data required for system and application configuration. The second module is a framework, named PrivaSIP, with a number of variations depending on privacy and performance requirements and is limited to the application layer since it protects end users' identity privacy in the SIP protocol. The combination of these two modules can be realized with the inclusion of SIP re-authentication and re-configuration information into the aforementioned context. The proposed schemes are compared to existing solutions based on well defined criteria; for PrivaSIP an extensive series of experiments were conducted on an appropriately designed testbed in order to measure its performance.

Giorgos Karopoulos

Department of Information and Communication Systems Engineering

UNIVERSITY OF THE AEGEAN

ΠΕΡΙΛΗΨΗ

Ο προσδιορισμός του IP Multimedia Subsystem (IMS) στην 3η Γενιά (3rd Generation - 3G) κινητών επικοινωνιών, σηματοδοτεί τη σπουδαιότητα των υπηρεσιών πολυμέσων στα μελλοντικά ασύρματα συστήματα. Η εισαγωγή ενός ανεξάρτητου υποσυστήματος υπεύθυνου για τη διαχείριση πολυμέσων είναι απόρροια των ειδικών χαρακτηριστικών και απαιτήσεων των εφαρμογών πολυμέσων. Ο τύπος αυτός εφαρμογών, ενώ έχει ανοχή μέχρις ενός σημείου σε απώλεια δεδομένων, δεν δείχνει την ίδια ανοχή σε καθυστερημένη παράδοσή τους. Το IMS, ωστόσο, δεν περιορίζεται μόνο σε κινητές επικοινωνίες κι αυτό διότι είναι βασισμένο στο IP, ενώ και τα βασικά για τη λειτουργία του πρωτόκολλα, τα SIP και Diameter, συναντώνται και στο Διαδίκτυο. Επιπλέον, οι σύγχρονες τάσεις δείχνουν ότι στο κοντινό μέλλον η σύγκλιση των περισσοτέρων ασύρματων συστημάτων θα γίνει πραγματικότητα με τη χρήση μιας κοινής πλατφόρμας που θα είναι το IP, το πρωτόκολλο του Διαδικτύου. Σε αυτή την ετερογενή δικτυακή αρχιτεκτονική οι υπηρεσίες πολυμέσων θα συγκλίνουν και αυτές μιας και θα χρησιμοποιούνται τα ίδια πρωτόκολλα, τα SIP και Diameter. Ενώ, όμως, αυτή η σύγκλιση θα δημιουργήσει πολυάριθμα οφέλη για τις εφαρμογές πολυμέσων, παράλληλα θα ευνοήσει την εμφάνιση απειλών ασφαλείας μιας και οι χρήστες πλέον θα έρχονται σε επαφή με ένα μεγάλο αριθμό παρόχων στα πλαίσια αυτού του πολύ-τομεακού περιβάλλοντος.

Αυτή η διατριβή εστιάζει στην ασφάλεια υπηρεσιών πολυμέσων σε ασύρματα ετερογενή δίκτυα βασισμένα στο πρωτόκολλο IP. Αναλύονται ένας αριθμός από ζητήματα τα οποία αντιμετωπίζονται στη συνέχεια από προτεινόμενους μηχανισμούς οι οποίοι μπορούν να λειτουργήσουν και σε απαιτητικές καταστάσεις, όπως π.χ. κατά τη διάρκεια μιας εναλλαγής (handoff) δικτύου ή/και διαχειριστικού τομέα. Αυτές οι αδυναμίες και οι προτεινόμενοι μηχανισμοί σχετίζονται με τη φάση σηματοδοσίας των υπηρεσιών πολυμέσων, πριν δηλαδή παραδοθούν στον τελικό τους προορισμό τα σχετικά δεδομένα.

Οι αδυναμίες που αναλύονται στην παρούσα διατριβή σχετίζονται με την προστασία της ιδιωτικότητας των χρηστών όταν αυτοί περιηγούνται σε διαχειριστικούς τομείς οι οποίοι θεωρούνται άγνωστοι, μη έμπιστοι ή και τα δύο. Οι μηχανισμοί που προτείνονται εδώ μπορούν να θεωρηθούν ως δύο μονάδες οι οποίες μπορούν να λειτουργήσουν είτε παράλληλα είτε ανεξάρτητα η μία από την άλλη. Η πρώτη περιλαμβάνει δύο σχήματα βελτιστοποίησης ασφαλούς εναλλαγής με ενίσχυση της ιδιωτικότητας κατάλληλα για ασύρματα ετερογενή δίκτυα στα οποία οι απαραίτητες πληροφορίες για γρήγορη επανα-αυθεντικοποίηση και επανασύνδεση υπηρεσιών μεταφέρονται μέσω μιας δομής δεδομένων στα υποψήφια για εναλλαγή δίκτυα. Αυτή η δομή μπορεί να μεταφέρει σχεδόν κάθε τύπο πληροφορίας είτε αυτή σχετίζεται με κρυπτογραφικό υλικό είτε με άλλα δεδομένα χρήσιμα σε εφαρμογές. Η δεύτερη μονάδα είναι ένα πλαίσιο που ονομάζεται PrivaSIP και περιορίζεται στο επίπεδο εφαρμογής μιας και προστατεύει την ιδιωτικότητα χρηστών του πρωτοκόλλου SIP. Ο συνδυασμός των δύο αυτών μονάδων επιτυγχάνεται με την εισαγωγή των κατάλληλων πληροφοριών για το PrivaSIP στην προαναφερθείσα δομή δεδομένων της πρώτης μονάδας. Οι προτεινόμενες λύσεις συγκρίνονται με υπάρχοντα σχήματα με βάση καλά ορισμένα κριτήρια, ενώ για το PrivaSIP μια σειρά από πειράματα εκτελέστηκαν σε κατάλληλα σχεδιασμένη πειραματική διάταξη για τη μέτρηση της απόδοσής του.

ΕΥΧΑΡΙΣΤΙΕΣ

Η διατριβή αυτή εκπονήθηκε στο τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων του Πανεπιστημίου Αιγαίου και κατά τη διάρκειά της υπήρξαν αρκετοί άνθρωποι οι οποίοι βοήθησαν στο, όσο το δυνατό γίνεται, καλύτερο αποτέλεσμα. Σε αυτό το σημείο θα ήθελα να επισημάνω τη σημαντική βοήθεια που είχα από τη χρηματοδότηση της ερευνητικής μου εργασίας από το Πρόγραμμα Ενίσχυσης Νέου Ερευνητικού Δυναμικού (ΠΕΝΕΔ) της Γενικής Γραμματείας Έρευνας και Τεχνολογίας (ΓΓΕΤ) του Υπουργείου Ανάπτυξης, του οποίου ήμουν υπότροφος. Πριν ξεκινήσω, θα ήθελα προκαταβολικά να ζητήσω συγγνώμη από αυτούς τους οποίους έχω ξεχάσει να ευχαριστήσω παρακάτω.

Καταρχήν χρωστάω ένα μεγάλο ευχαριστώ στον επιβλέποντά μου Καθηγητή Στέφανο Γκρίτζαλη για τη σωστή καθοδήγηση, την υπομονή και την εμπιστοσύνη που μου έδειξε καθ' όλη τη διάρκεια αυτής της προσπάθειας. Το γεγονός ότι αφιέρωσε κόπο αλλά και μέρος από τον ανύπαρκτο χρόνο του ήταν πολύ σημαντικό, ενώ και η δημιουργία των κατάλληλων συνθηκών για την ολοκλήρωση της ερευνητικής αυτής εργασίας είναι σαφώς δικό του έργο.

Επίσης, θα ήθελα να ευχαριστήσω τον Επίκουρο Καθηγητή Κωνσταντίνο Λαμπρινουδάκη για την πολύ καλή συνεργασία όλα αυτά τα χρόνια ξεκινώντας από τον πρώτο καιρό που βρέθηκα στο τμήμα, καθώς και για την εμπιστοσύνη που μου έδειξε προτείνοντάς μου τη θέση του υποψήφιου διδάκτορα. Πολύ σημαντική ήταν η συμβολή του Λέκτορα Γεώργιου Καμπουράκη σε όλη τη διάρκεια της εκπόνησης της διδακτορικής διατριβής μου και τον ευχαριστώ ιδιαίτερα για αυτό. Η σχεδόν καθημερινή μας επικοινωνία έβαλε τις βάσεις τόσο για τη γέννηση ερευνητικών ζητημάτων με μεγάλο ενδιαφέρον, όσο και για την αποτελεσματική επίλυσή τους. Επίσης, θα ήθελα να ευχαριστήσω τη Λέκτορα Ελισάβετ Κωνσταντίνου για τη βοήθειά της σε ορισμένα θέματα σχετικά με κρυπτογραφία ελλειπτικών καμπυλών.

Δεν θα μπορούσα να ξεχάσω τους συνάδελφους αλλά κυρίως φίλους Δημήτρη Γενειατάκη και Θοδωρή Ευδωρίδη. Τους ευχαριστώ για τη βοήθεια, τη συνεργασία και τη στήριξη, αλλά και για τις πολλές καλές αναμνήσεις από όλα αυτά τα χρόνια που βρισκόμασταν στη Σάμο. Θα ήθελα επίσης να ευχαριστήσω και τη συνάδελφο κ. Παπαναγιώτου Ευαγγελία για τη βοήθεια που μου προσέφερε σε θέματα στατιστικής.

Ξεχωριστά ευχαριστώ τους φίλους μου Τέλη και Αλέξη οι οποίοι με επανέφεραν στον πραγματικό (τετραδιάστατο) κόσμο όταν αυτό χρειαζόταν. Επίσης, θα ήθελα να ευχαριστήσω και τη φίλη μου Αθηνά για τις συζητήσεις μας και τις παροτρύνσεις της. Το γεγονός ότι μπορούσα να μοιραστώ μαζί τους όλα όσα με απασχολούσαν αλλά και όσα με χαροποιούσαν υπήρξε πολύ σημαντικό για μένα.

Τέλος, δεν ξέρω αν μπορεί να είναι αρκετό, αλλά οφείλω ένα μεγάλο ευχαριστώ στους γονείς μου, Σταμάτη και Βασιλική, και τον αδερφό μου Ηρώδη, οι οποίοι όλα αυτά τα χρόνια με στηρίζουν και συμμερίζονται τις χαρές και τις αγωνίες μου.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 - Introduction

Wireless communications have witnessed tremendous advances in technology and market penetration over the last two decades and are now being a part of everyday life for hundreds of millions of people all around the world. Mobile communications, considered a luxury in the early 1990s, have become a necessary means of communication in less than 20 years. Wireless computer networks transform from extension points of corporate networks into a wireless system that connects not only computers but any kind of portable device. Together with wireless technologies, applications have also changed as well; simple voice telephony of early communication systems has evolved into a wide range of multimedia applications like Short Message Service (SMS), video sharing, video call etc. Hardware advancements have transformed large mobile phones supporting only voice services into an all-in-one pocket size device. All these facts have brought about a demand for high quality multimedia applications and services deliverance which, however, brings a number of issues concerning Quality of Service (QoS); additionally securing these services while maintaining an acceptable QoS is a very challenging research topic.

For a long time, communications, including mobile communications, used circuit switched networks resulting in low utilization of the available bandwidth; on the other hand, computer networks are packet switched, thus providing better throughput. Currently there is a trend towards moving from the Second Generation (2G) of mobile systems, which are mostly based on circuit switching, to the Third Generation (3G), which is based on the packet switched concept; in the future, beyond 3G (b3G) or Next Generation Networks (NGN), as they are known, will continue utilizing packet switching like 3G but in a more open architecture. While this transition will lower the costs for both operators and their customers, improve existing applications and offer the basis for new applications development, it will also pose new challenges regarding quality and security of the offered services.

Wireless and mobile networks today, provide the necessary means to transport and deliver multimedia services to end users. What they do not provide is guaranteed QoS, a unified and modular framework for services charging and security adaptable to specific needs. Throughout this thesis the focus is mainly on QoS and security and especially how multimedia security can be enhanced without severely degrading perceived QoS. Multimedia delivery in wireless computer networks mostly has issues regarding QoS since packet switched wireless networks can have high and variable error rates and delays; on the other hand, multimedia delivery in mobile networks is based mostly on closed and proprietary systems, thus, security issues are inherent since security measures may not be publicly known or tested. The convergence of these two wireless worlds into a NGN will not only combine their advantages but also their disadvantages as well. Consequently, new methods and solutions should be proposed in order to alleviate existing issues and foreseen problems arising from the convergence of these different technologies.

The 3G introduces IP Multimedia Subsystem (IMS) [1] which is a subsystem responsible for multimedia session management; this system or possible descendants will likely exist in NGNs offering a central point of multimedia management to service operators. This way, multimedia delivery enters a new era where multimedia session establishment will be feasible over any type of network technology; this will be made possible using a common framework which is the well known Internet Protocol (IP). Over this common framework, applications and services will be

unified and their management will be easier and offer more options to meet the needs of different groups of customers.

This wireless technologies convergence vision analyzed above will create new business models as well. Currently, the common case is that the network operator also plays the role of content operator offering multimedia content to its subscribers. The upcoming all-IP based wireless networks with their open architecture will offer the possibility to everyone to be able to play the role of content provider while network operators will possibly continue to also provide multimedia content. This new reality creates a multi-domain environment where each operator, no matter whether it is network or content provider, has control of its own administrative domain and different domains are compatible and co-operate with each other by inter-domain agreements or other means of trust transfer. In the described multi-domain environment, security will play a significant role since not all operators will be trusted by the end user; moreover, the end user usually signs a contract, thus agreeing to the specified terms, with one operator making this operator's domain the only domain with which he has explicit mutual trust.

1.1 Research area

In the near future, as already stated above, networks are likely to follow a multi-domain approach where a large number of operators, which can be either network providers or service providers or both, will co-exist and offer their services to end users. Security concerning the exchanged data will be crucial since some of these domains may be unknown or untrusted or both. This opposes to systems used today since, for example, 2G networks are closed systems where subscribers only deal with and trust one operator which is both their network and content provider. Next generation systems will follow an open architecture like the Internet; thus, further research is necessary to protect end users from fraud and data exposure.

Providing security to wireless systems poses a number of challenges since they are more vulnerable compared to wired networks. Wired networks pose certain difficulties to prospective eavesdroppers, the most important of which is that it is physically limited and attackers should attack the wired interface in order to have access to the network. In a wireless network, data are transmitted through the air interface; this means that everyone who has the appropriate equipment can eavesdrop on the exchanged data. These data can be either signaling or media transport data, and security for both of these data categories is important for protecting user’s personal data. Therefore, appropriate measures should be taken and security mechanisms should be applied in order to protect media transport, as well as signaling data.

The security services that need to be offered in wireless multi-domain networks are: confidentiality, integrity and availability of both signaling and media data. Moreover, authentication, authorization and accounting mechanisms should also be employed in order to control network access and billing of the offered services; these mechanisms should be able to co-operate with similar mechanisms of other domains since users will receive services from multiple domains. Users’ roaming among different domains has further implications concerning security; identification of users trying to access some network is important for the protection of the network itself against unauthorized use. Furthermore, the privacy of end users should be protected so that their personal information is only available to entitled entities.

The transition from current incompatible wireless systems into the Next Generation of networks which will be heterogeneous based on an all-IP approach will create complicated trust relationships between users and operators. While in today’s systems there are a number of solutions, their applicability and efficiency in this new environment should first be investigated.

Closed systems, like existing mobile networks, in some cases utilize proprietary and/or outdated mechanisms for the protection of transmitted data; for example Global System for Mobile communications (GSM), which is a 2G system, utilizes A3, A5 and A8 cryptographic algorithms which are proprietary and have been reversed engineered [2]. Even more open systems like Wireless Fidelity (Wi-Fi) have witnessed serious flaws [3] which eventually have been fixed. Previous experience shows that existing security mechanisms utilized in the Internet today are the most appropriate solutions since they are publicly available and thoroughly tested. Under this context the existing security protocols are reviewed, analyzed and evaluated and new or alternative solutions are proposed where the existing schemes prove to be inadequate.

1.2 Problem statement

Probably the highest priority issue when considering multimedia applications, and the characteristic that distinguishes them from other categories of applications, is their low tolerance to time delays. For instance, the acceptable time delay for telephony has been set at 150 msec for one-way transmission path by the International Telecommunications Union – Telecommunications Standardization Sector (ITU-T) in their G.114 recommendation [4]. While multimedia applications can tolerate data losses up to some point, delays higher than 150 msec could render them unacceptable. On the other hand, traditional applications like file transfer are not affected by time delays but cannot tolerate data losses.

It is argued that providing QoS to multimedia applications is a difficult task; however, the situation gets even more complicated when security mechanisms have to be incorporated into such applications. For example, if a solution could adequately face a certain security issue it does not mean that it is appropriate for multimedia applications over heterogeneous networks; a very important factor is that it should introduce time delays within some acceptable bounds. Every security solution has to take into consideration the unique characteristics of multimedia applications and be designed with these in mind so as to be able to meet their specific requirements.

NGNs will be composed of networks with different access technologies forming all-IP heterogeneous networks. It is also foreseen that in this environment user terminals will be equipped with multiple wireless interfaces so that they can connect to the most appropriate network each time based on specific parameters like cost and bandwidth. This will create the need for handoffs, whether these are horizontal or vertical. Horizontal handoffs occur when the user terminal handoffs from a network using an access layer technology to another network using the same technology; in vertical handoffs the terminal handoffs to a network using a different access layer technology. A very important issue when a handoff is taking place is the uninterrupted continuation of multimedia services; the situation is further complicated when considering security as well. For instance, when the mobile terminal executes a vertical handoff then the user/terminal must authenticate and authorized to the new network in order to gain access as well as to re-authenticate to the service provider in order to maintain the running multimedia session. These procedures tend to be expensive in terms of time delay since they normally involve cryptographic operations; the aforementioned observations show that multimedia delivery in all-IP wireless heterogeneous networks is a very challenging area of research.

Another point of interest lies in the protocols used for multimedia delivery and especially the signaling protocols that are used to setup and terminate multimedia sessions. As it has already been stated, IMS is the subsystem that handles multimedia signaling in 3G and it is expected to play central role to NGNs as well. The protocol that plays central role in IMS is the Session

Initiation Protocol (SIP) [49], a text-based application layer protocol responsible for the initiation, modification and termination of multimedia sessions. This transition, however, from present systems to open IP-based protocols and standards will create numerous threats comparable to those encountered on the Internet. The issues encountered here can also be combined with those described in the previous paragraph since the same signaling protocol is used to re-establish the existing sessions during handoff. While solutions to these issues can be developed separately, in order to succeed the goal of smooth operation during handoff these mechanisms should co-operate in order to offer secure and efficient multimedia delivery in NGNs.

1.3 Motivation

Over the last few years a great number of people enjoy the advantages offered by wireless technologies. Users, but system administrators as well, love the freedom and easiness of deployment that wireless technologies can provide compared to wired networks and their use is increasing every day. Nowadays wireless technologies are part of everyday life and influence the work and life of many end users; thus, every technology or issue related to them has an effect on a large number of people.

Ever since computer scientists started considering the importance of systems security, security incidents never stopped appearing and it is likely that they will always exist. While this has not prevented users from adopting new technologies, security is considered a desired characteristic of every system and significantly increases the chances of a new technology to be adopted by users. So, a key requirement in order to increase the market penetration of wireless technologies is to solve the most significant security issues in upcoming wireless systems.

The previous observations lead to the conclusion that securing wireless technologies is crucial for their approval by end users; considering that multimedia usage over wireless networks is growing, then the study of how to deliver multimedia over all-IP wireless heterogeneous networks securely is a very challenging research area. This research has been motivated by the aforementioned facts and their outcomes; more specifically:

- There is a transition towards an all-IP architecture which offers many advantages to users and operators and new and advanced features to services; on the other hand, the existing security threats found on the Internet are transferred to the new platform while new threats will show up.
- In addition to the above issues, the new generation of networks will also inherit the issues found today in older networks like 2G; this is because 3G is at some degree backwards compatible with 2G.
- In the upcoming NGNs, the number of operators is likely to increase since the existing operators will continue to play the two roles they have in existing networks as both network and service providers, and new (exclusively) service providers will gradually appear. This will create a large number of different administrative domains which should co-operate in order to offer their services and at the same time operate transparently to end users. This multi-domain environment will create a complex trust model and its management will create the need for stable and scalable security solutions.

1.4 Goals

The objective of this thesis is to propose a framework that provides security, and more specifically privacy protection, to end users of multimedia services in all-IP wireless

heterogeneous networks while keeping imposed delays under an acceptable level. The main purpose of this thesis is to review existing mechanisms, propose, implement and evaluate alternative solutions to protect end users' privacy while using multimedia over NGNs. The specific goals of this research are:

1. Review and compare state-of-the-art mechanisms for secure handoff optimization in wireless heterogeneous networks
2. Propose secure handoff optimization schemes that preserve end users' privacy while roaming among different administrative domains
3. Review and compare solutions that preserve users' privacy when using one of the most promising multimedia signaling protocols, namely SIP
4. Propose, implement and evaluate new mechanisms to protect more effectively end users' privacy while using SIP

1.5 Contribution

One of the main findings of this research is the observation that while a large number of security issues are thoroughly investigated and more or less solved, end users' privacy is not receiving as much attention as other security properties in multimedia delivery over NGNs. This thesis mainly contributes towards protecting end users' privacy when they roam between different administrative domains and at the same time receive multimedia services over all-IP wireless heterogeneous networks.

First, the existing state-of-the-art mechanisms for secure handoff optimization in NGN are reviewed and compared to each other [5]. These mechanisms not only provide user and network security during handoff but try to minimize the handoff delays as well. The evaluation of these schemes showed that while security, in general, and handoff delays are at an acceptable level, no method takes into consideration the privacy protection of end users. This observation led to the proposal of two privacy preserving secure handoff optimization schemes [6] - [8]. One of the most promising protocols from the previous review was chosen, more specifically "AAA Context Transfer", and was enhanced in order to protect end users' privacy as well. AAA Context Transfer, and consequently the two privacy enhanced proposed methods, are generic solutions that offer secure handoff optimization to any kind of service; if, however, the protocols operating at a higher layer do not protect users' privacy as well, then it is possible that no privacy at all is offered. Since the focus of this thesis is multimedia delivery, the next step was to find solutions to protect end users' privacy in one of the most popular multimedia signaling protocols, which is SIP. An extensive review of existing solutions that can be used to protect privacy in SIP has shown that they are inadequate and/or cannot be applied to certain environments [9]. The inadequacy of existing methods led to the proposal of a framework [9] - [11] that offers identity privacy in SIP using a number of different cryptographic algorithms, thus giving the opportunity to system administrators to choose the right combination in order to get privacy, acceptable delays and easiness of deployment. The qualitative and quantitative evaluation showed that the aforementioned framework offers adequate privacy to end users while at the same time the performance penalty is within an acceptable range. The two proposals, that is privacy enhanced context transfer and privacy framework for SIP, can either be combined, since they operate at different layers, or utilized individually depending on the environment and the situation. The contribution of this thesis by chapter is summarized in Table 1-1.

Chapters	Description	Contribution
Chapter 1	This introduction.	
Chapter 2	Overview of the research area of this thesis. The concepts of Next Generation Networks, multimedia signaling protocols and Authentication, Authorization and Accounting are studied.	
Chapter 3	Review and evaluation of secure handoff optimization schemes. The state-of-the-art mechanisms are reviewed and compared to each other based on a number of criteria.	[5]
Chapter 4	Design and evaluation of privacy preserving secure handoff optimization schemes for wireless heterogeneous networks.	[6] - [8]
Chapter 5	Review and evaluation of SIP privacy solutions. All existing SIP privacy solutions were examined, even those not designed for privacy specifically but can offer such services as well.	[9]
Chapter 6	Design, implement and evaluate PrivaSIP, a framework for protecting end users' privacy in SIP. The proposed framework was evaluated through testbed experimentation and qualitatively compared with existing solutions.	[9] - [11]
Chapter 7	The conclusions of this thesis and related open research issues are described.	

Table 1-1: Thesis contribution by chapter

1.6 Thesis structure

The *second chapter* of this thesis provides an overview of the environment of multimedia delivery in the Next Generation of Networks. Currently, a number of different access layer technologies for wireless networking co-exist and this is not likely to change; at least too soon. This chapter analyses the dominant wireless networking technologies which are WLAN and 3G, looking beyond those technologies in the future, examining 3G-WLAN convergence and NGNs as well. Next, the most important multimedia delivery protocols are summarized and an overview of Authentication, Authorization and Accounting (AAA) mechanisms is provided. Finally, an implementation combining all these concepts, the IP Multimedia Subsystem (IMS) which is part of the 3G specification and expected to be an important part of NGNs, is analyzed.

Chapter three provides a critical review of secure handoff optimization schemes utilized in all-IP wireless heterogeneous networks. Handoff schemes are very important in wireless systems because they provide continuation of received services; the optimization of such schemes is even more important when demanding applications like multimedia are in place. Adding security features to such mechanisms make even more difficult the goal of providing smooth and uninterrupted continuation of multimedia services. Such schemes are reviewed and compared to each other based on well defined criteria showing the advantages and disadvantages of each one of them.

The findings of the previous review showed that no secure handoff optimization scheme takes into consideration end users' privacy. In *chapter four*, two privacy enhanced mechanisms are proposed, each one having its own characteristics so that they can be utilized according to the system environment. First, the privacy issues in secure handoff optimization schemes are

discussed, followed by the description of the two mechanisms and the justification of their effectiveness.

Chapter five deals with a specific multimedia signaling protocol, namely Session Initiation Protocol (SIP), and the ways that is possible to protect end users' privacy in this protocol. This chapter begins with the problem statement concerning user identity privacy in SIP and how this issue can be misused for malicious acts. A number of security solutions for SIP can be utilized for protecting end users' privacy even if they were not designed for this purpose specifically. These solutions are reviewed and compared to each other focusing mainly on their privacy protecting qualities. The outcome of this evaluation is that these mechanisms can be inefficient and/or inadequate under certain conditions.

The inadequacies of SIP security schemes revealed the need for a new mechanism that could protect end users' privacy and at the same time be efficient enough to be used in wireless heterogeneous networks. In *chapter six* a new framework is proposed, namely PrivaSIP, which effectively protects user identities in SIP sessions. After the description of the framework and its variations, the experimental testbed setup is presented which was used to measure the imposed delays by the proposed framework using different combinations of cryptographic algorithms. At the end of the chapter, PrivaSIP is compared with existing schemes based on the same criteria of the previous chapter.

Chapter seven provides the conclusions of this thesis and possible open issues stemming from this research in the area of multimedia security in all-IP wireless heterogeneous networks.

Chapter 2 - Secure multimedia over Next Generation Networks

As the Internet has become an essential part of our everyday life, the progress of communication technologies will offer an "always connected" opportunity to everybody. To support this vision, the research community is suggesting a move towards an all-IP platform in order to take advantage of the high bandwidth of WLANs and the broad coverage of cellular networks and WMANs. The convergence of these heterogeneous wireless technologies will eventually lead to the Next Generation Networks (NGN).

The deliverance of multimedia services over these NGNs poses new challenges and requirements over existing standards used for multimedia delivery. These standards exist today and used in the Internet; however, their adoption for NGNs requires them to offer additional features like Quality of Service (QoS) and charging. Such features can be offered by Authentication, Authorization and Accounting (AAA) protocols which offer a central point of control in the core network.

This chapter provides an overview of the most significant wireless technologies and the vision of NGN. A discussion of the dominant multimedia delivery protocols existing today is followed by a summary of AAA protocols and their advantages and disadvantages. Finally, it is presented how these protocols can be combined in one framework in order to have secure multimedia services delivery over all-IP heterogeneous networks.

2.1 Next Generation Networks

In the past few years a transition from wired to wireless networks has been observed and this trend is likely to increase in the future. This way the realization of better applications and services targeting a very large number of people will be possible. End users wish to receive services anytime and everywhere and this makes more apparent the need of a transparent and seamless solution for the users to access the Internet.

In today's networked world, however, a high degree of heterogeneity is observed and this holds especially for wireless networks. While in the wired networks the domination of the Internet Protocol (IP) is unquestionable, wireless networks use a number of different technologies making them incompatible to each other. These different technologies mainly concern the link layer while over the last years there are efforts for the convergence of different types of networks on a higher layer and more specifically on the network layer utilizing IP.

It is foreseen that in the near future all these types of networks with different access technologies will converge into heterogeneous networks that will base their operation on an all-IP approach. This all-IP approach will allow the multiple types of wireless networks to interoperate with each other and also with wired networks without the need of gateways or any other translation means.

Nowadays, the dominant wireless systems are (a) the third generation (3G) of mobile telecommunications system (although it can also be argued that we are in a transition period from the second generation to the third) and (b) the Wireless Local Area Network (WLAN). The next generation of wireless systems will be the so called fourth generation (4G) or beyond 3G (B3G), a term which however has no formal definition but its characteristics can be summarized

in the following: "*The 4G will be a fully IP-based integrated system of systems and network of networks achieved after the convergence of wired and wireless networks as well as computer, consumer electronics, communication technology, and several other convergences that will be capable of providing 100 Mbit/s and 1 Gbit/s, respectively, in outdoor and indoor environments with end-to-end QoS and high security, offering any kind of services anytime, anywhere, at affordable cost and one billing."* [12]. Another term used for future telecommunications networks is "Next Generation Networks" (NGN) which is a broad term and according to the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T) [13] is defined as: *"A Next Generation Network (NGN) is a packet-based network able to provide services including Telecommunication Services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies. It offers unrestricted access by users to different service providers. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users."* [14]. The following sections provide a synopsis of the main today's wireless technologies and the approaches for their convergence which will eventually lead to the NGN.

2.1.1 3G

One of the main technologies that lead the proliferation of wireless communications is mobile communications systems which are being extensively used throughout the world. A great penetration has been achieved by the second generation (2G) of cellular mobile systems leading to the familiarization of users with wireless networks. However, the limitations of 2G and mainly the high costs and low bit-rate necessitated the transition to the third generation (3G) of mobile systems which can offer better services, in higher speeds and at a lower cost. Currently two 3G systems are mainly in use worldwide: the Universal Mobile Telecommunications System (UMTS) and Code Division Multiple Access 2000 (CDMA2000). From this point forward UMTS will be used as the representative 3G system keeping in mind that the same principles also apply to CDMA2000.

While 3G is the evolution of 2G, an intermediate system known as 2.5G has been used. 2.5G systems use the infrastructure of 2G in order to offer data services in higher speeds than 2G at approximately 100 kbps. This speed is considered an improvement over the 9.6 kbps offered by 2G systems, but it cannot satisfy the requirements of today's users and multimedia applications. 3G provides higher speeds while retaining the advantages of its predecessors; it offers data services at speeds ranging from 300 kbps to 2Mbps depending on the movement of the user, and retains the wide coverage, high mobility and established user base of 2G and 2.5G systems.

3G mobile system is well defined and has reached a satisfactory level of standardization; this, together with all previous advantages, makes it an appealing system for end users and network providers. In spite of its advantages and new capabilities, however, 3G has seen a rather slow deployment and market penetration; at least slower that the expected. It comes as no surprise that the reasons for this are the high costs; these costs concern: (a) radio spectrum licensing, (b) purchase or upgrade of network equipment, (c) relatively higher operational and maintenance costs. Still, mobile network operators have in their future plans the full transition to 3G systems.

Looking into the architecture of 3G systems it can be argued that each operator's network constitutes an autonomous administrative domain. This network is divided into the core network (CN) and the access network (AN). The AN is responsible for the connectivity of mobile terminals with the CN. The CN is divided into a circuit switched domain, which is used for voice related traffic, and a packet switched domain, which is used for data traffic. It is further enriched

with IP Multimedia Subsystem (IMS) which is used to deliver multimedia services to UMTS end users. The CN is responsible, among others, for managing the subscriber base, providing security related functions like authentication, authorization and billing, connecting this operator's network with the Internet and provide multimedia services to end users.

2.1.2 WLAN

A little bit later than 2G, WLAN started evolving and nowadays it has successfully penetrated the wireless systems market. The main purpose of WLAN is to enable enterprises and individual users to setup LANs without the need to install wiring which costs time and money. Apart from homes and enterprises, places with substantial concentration of people like ports, airports, hotels and cafes are suitable for hot spot areas – areas covered by WLAN. The two main WLAN standards are: 802.11 based networks and HIPERLAN/2. Here, the family of 802.11 based networks will be considered as representative of WLAN.

WLAN's advantages include: (a) low cost of equipment, operation and maintenance, (b) no radio spectrum licensing, (c) simplicity, (d) easiness of deployment and (e) high speeds. Comparing WLAN with 3G the following points can be deducted. WLAN operates at speeds ranging from 1 to 108 Mbps which is significantly higher than the 300 kbps to 2 Mbps found in 3G. While WLAN utilizes the unlicensed radio spectrum reducing the costs in comparison to 3G, this feature also decreases its availability and reliability since the shared radio spectrum can result in substantial interference from other users. Another disadvantage of WLAN is that it covers much more limited areas compared to 3G while it can tolerate only low mobility by the users.

In contrast to UMTS, WLANs do not have standardized network architecture. WLAN has two main operational modes: an infrastructure-based architecture and an ad hoc one. The majority of deployed WLAN systems utilize the first one, where an Access Point (AP) offers IP connectivity to wireless hosts with a wired backbone network. Obviously here only packet-switched services are offered together with security related functions like authentication, authorization and accounting (AAA).

2.1.3 WLAN-3G convergence

It might be natural for 3G and WLAN, both being wireless technologies, to compete with each other but in reality this is not the case. Considering their characteristics these two technologies are rather complimentary than competitive. While 3G can support higher user mobility and cover greater geographical areas, WLAN can support higher speed with significantly lower cost, making the convergence of the two types of wireless networks an appealing perspective.

This convergence vision presents a number of advantages for all involved parties. From the 3G operators' point of view, WLAN can assist them in offering better services with higher speeds to their already established customer base. For some operator of a WLAN network, 3G can offer continuity in service delivery by providing connectivity to users between hot spots. The users can benefit from such cooperation by enjoying higher speeds at low cost when they reside in a hot spot area like an airport without losing their connectivity when moving to more rural areas.

This convergence, however, should be realized in a way so that the switching between different network technologies is as transparent and seamless to the end user as possible. The reality today is that the coverage areas of these two types of networks are in many cases overlapping; however, it is not easy for the end users to take advantage of this fact. A truly

converged environment would allow single sign on [15], user and terminal mobility, session and application continuity as well as smooth handoff between networks.

There is a number of proposals [16][17][18] on how WLAN-3G convergence could be possible; a common requirement of all proposals is that the mobile equipment of the user should be dual-mode supporting both 3G as well as WLAN connectivity. The two main types of WLAN-3G interworking are: tight coupling and loose coupling. In tight coupling, WLAN is considered as an addition to 3G's access network so that all traffic from WLAN is routed through 3G's core network where all AAA procedures are executed. In any case an interworking network element is needed in order to bridge the two networks and make the WLAN appear like an internal element to 3G system. The main advantage of tight coupling is easiness of deployment since all 3G elements and mechanisms are reused in the new converged network; this, however, can only be made possible if the 3G operator is the same as the WLAN operator. In loose coupling only certain functionality of 3G core elements is utilized such as subscriber management and AAA functions. Here, an interworking element is not needed for the bridging of the two different types of networks and the deployment, operation and management of 3G and WLAN systems are independent. Since, however, operators should have business level agreements in order to support seamless roaming of their users there should be some kind of interworking mechanisms in order to allow WLAN and 3G to cooperate seamlessly; these mechanisms are usually IP-based because it is easier in terms of deployment and complexity. The advantages of loose coupling are the easiness of deployment and the independency it offers to 3G and WLAN operators; this, however, comes at the cost of performance degradation.

2.1.4 Beyond 3G

The research community shows great interest in the next step in wireless communications moving towards the so-called beyond-3G (b3G) or 4G architectures. While there are a number of definitions trying to identify the concept of 4G communications, it seems that the most common points of these definitions argue that a 4G system has the ability to offer purely packet switched data services over any wireless access system in a transparent and seamless way. Another term used for the future networks is Next Generation Network (NGN) which is mainly defined as a type of network that can carry all types of services, including voice, video, text and images over a common platform, which is IP. In any case, the convergence of WLAN and 3G technologies is a significant step towards the new generation of networks even if it has inherent deficiencies stemming from the fact that it depends on the currently defined architectures of WLAN and 3G.

The definition of NGN by ITU-T was given previously; the same organization defines the following as the fundamental aspects of any NGN [14]:

- Packet-based transfer
- Separation of control functions among bearer capabilities, call/session, and application/ service
- Decoupling of service provision from network, and provision of open interfaces
- Support for a wide range of services, applications and mechanisms based on service building blocks (including real time/ streaming/ non-real time services and multi-media)
- Broadband capabilities with end-to-end QoS and transparency
- Interworking with legacy networks via open interfaces
- Generalized mobility
- Unrestricted access by users to different service providers
- A variety of identification schemes which can be resolved to IP addresses for the purposes of routing in IP networks

- Unified service characteristics for the same service as perceived by the user
- Converged services between Fixed/Mobile
- Independence of service-related functions from underlying transport technologies
- Compliant with all Regulatory requirements, for example concerning emergency communications and security/privacy, etc.

The fundamental idea of NGN is to carry all types of service over a single packet-switched network; the advantages of such an approach are numerous. Operators save money since they have to manage and maintain only one network platform; it also offers them the possibility to provide new services that combine different types of data. NGNs will be more versatile than today's networks because they do not have to be physically upgraded in order to support new service types; the network simply transports data while services are controlled by computer software which is easily upgradable. This also means that apart from network operators, third parties can launch new services as well; in such an environment the user will have to choose between a large number of service operators with any implications this may have in quality of service, cost, security etc. Another important point here is that while NGNs will be based on IP, they will have features that the Internet does not have, such as an assured quality of service and level of security.

The types of services that an NGN will be able to offer include the following:

- VoIP. Voice-over-IP (VoIP) will become reality since it will be possible for a user to make a call either from a telephone or a softphone on a PC using the same underlying platform. These calls will be transferred over a packet-switched network in a more efficient way than today's voice calls since they will not need a dedicated line anymore; these calls will share bandwidth with other types of data.
- IPTV. Internet Protocol Television (IPTV) is a term for digital television deliverance over IP; this technology is expected to be one of the main commercial drivers increasing the penetration of NGN's worldwide.
- Converged services. Here the possibilities are numerous and each service provider can combine different services running on the NGN to create a unique representation of information to the user. For example, a user could have a single mailbox that collects e-mails, voice mails and video mails.
- Personalized services. Nowadays providers offer combinations of broadband internet, TV, phone and mobile services to users; however, for managing reasons these combinations are limited. With NGNs the end users can have finer control over the services they receive, be billed only for the services they want, and all this through a single system.
- Mobility services. Through NGNs the vision to access services anytime, anywhere will become reality since different types of mobility will be easily supported by future networks; these types include session, application, user and terminal mobility.

2.2 Multimedia delivery protocols

Regarding multimedia storage, playback and transmission the past decade has been a transition period from analog to digital technology. Now, that this analog-to-digital multimedia revolution is nearly complete, a new multimedia-over-IP revolution has started where all types of multimedia like radio, television, telephony and stored media are being delivered over IP using wired and wireless networks. This new revolution will not only make cheaper and easier the distribution of multimedia content, but it will create the conditions for novel applications as well.

Despite the fact that multimedia are being used for long time, the Internet and wireless networks provide only limited support for multimedia applications. This is due to the characteristics of both the Internet and wireless networks which are inherently unpredictable and have varied time delays and packet losses. While these characteristics might not be a problem for applications like file transfer, they can have considerable consequences for real time multimedia applications when the delays and packet losses are perceived by the end user. Multimedia applications tend to be delay sensitive, bandwidth intense and loss tolerant and these properties highly define the requirements by transport networks. This is the reason why in recent years the area of multimedia communication and networking is an active and challenging area and has seen significant research interest.

Multimedia delivery is made possible through a number of protocols operating over IP; these protocols can be divided into three general categories:

- Signaling protocols
- Media transport protocols, and
- Other protocols that offer complementary services like authentication, QoS and Network Address Translation (NAT) traversal.

Figure 2-1 summarizes the main protocols found in each category. For the rest of this thesis, the main focus is on the security of signaling protocols and how this can be accomplished with no perceived delay by the end user; the category of other protocols is also investigated in order to discover how AAA operations, which are in general considered expensive in terms of time delay, can be improved so as to be efficient enough to be used for multimedia delivery.

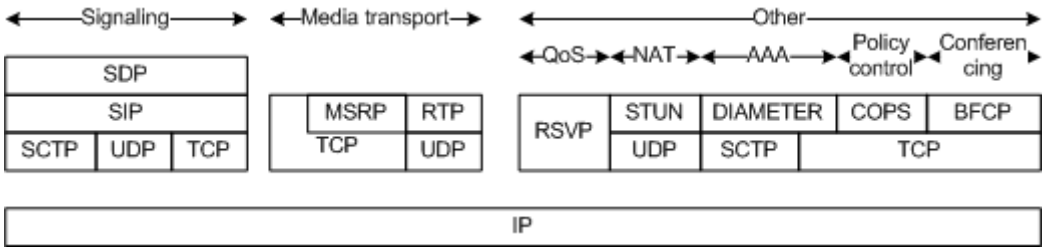


Figure 2-1: Protocols for multimedia delivery over IP

2.2.1 H.323

The first, and for some years the only, signaling protocol for multimedia communications over IP was the ITU recommendation H.323 [19]. H.323 is part of the H.32x series of protocols and was originally proposed for video conferencing over LAN; soon, it was extended to cover telephony as well as other types of multimedia over the Internet. From 1996 until 2006, six versions of H.323 have been adopted which are backwards compatible to each other; however, most deployed systems use version 2.

H.323, together with other ITU and IETF protocols, defines and standardizes a number of elements which are shown in Figure 2-2. The main elements of an H.323 architecture are: terminals, gatekeepers, gateways and Multipoint Control Units (MCUs). From these elements terminals, gateways and MCUs are called end points because they are network end devices. An end point can originate and terminate media streams which could be audio, video, data or a combination of these; the minimum that an end point must support is audio, while video and

data are optional. The gatekeepers are servers that control a zone which is the smallest possible administrative domain in H.323 and their presence is optional; the terminals' capabilities, however, are limited when a gatekeeper is not present. A gateway provides an interface between an H.323 network and a network using a different protocol, like the Public Switched Telephone Network (PSTN), and it is also an optional element in the H.323 architecture. Finally, an MCU provides conferencing services to the H.323 network terminals.

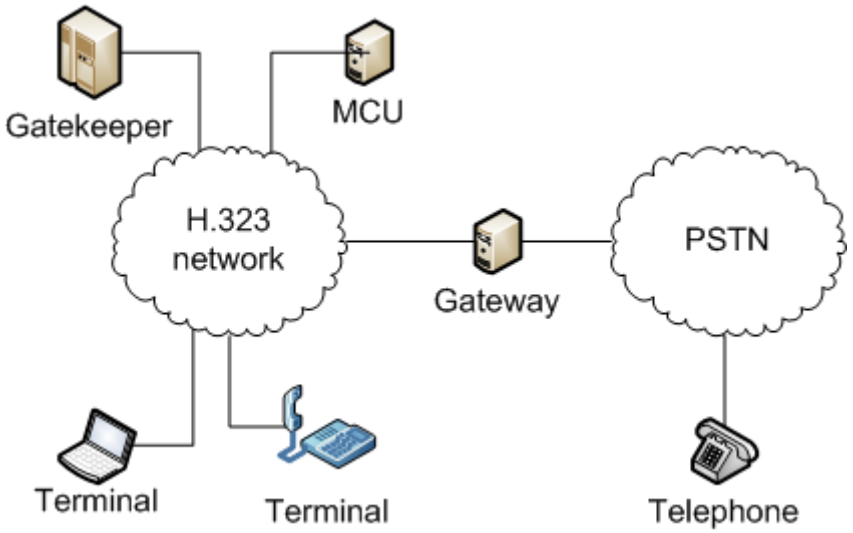


Figure 2-2: General architecture of an H.323 network

Figure 2-3 shows an example call flow for H.323. The exchanged messages before "RTP media session" constitutes the call setup while the rest of the messages the tear down of the call. In the following, the most important of these messages will be summarized. First both terminals must have been registered to the gatekeeper. The calling terminal sends an Admission Request (ARQ) to the gatekeeper containing the address of the called terminal. If the gatekeeper decides that the call can be continued it sends back an Admission Confirmation (ACF) message. The caller sets up a TCP connection to the callee through the "Setup" and "Call proceeding" messages. The callee must also get permission from the gatekeeper before the acceptance of the call, so ARQ and ACF messages are exchanged between them. What follows is a number of handshake messages for capabilities determination. When the media session is over either of the terminals can tear down the current call. The correspondent messages are Disengage Request (DRQ) and Disengage Confirmation (DCF) and these messages should be directed to the gatekeeper so that it knows that the resources used in this call have been freed up.

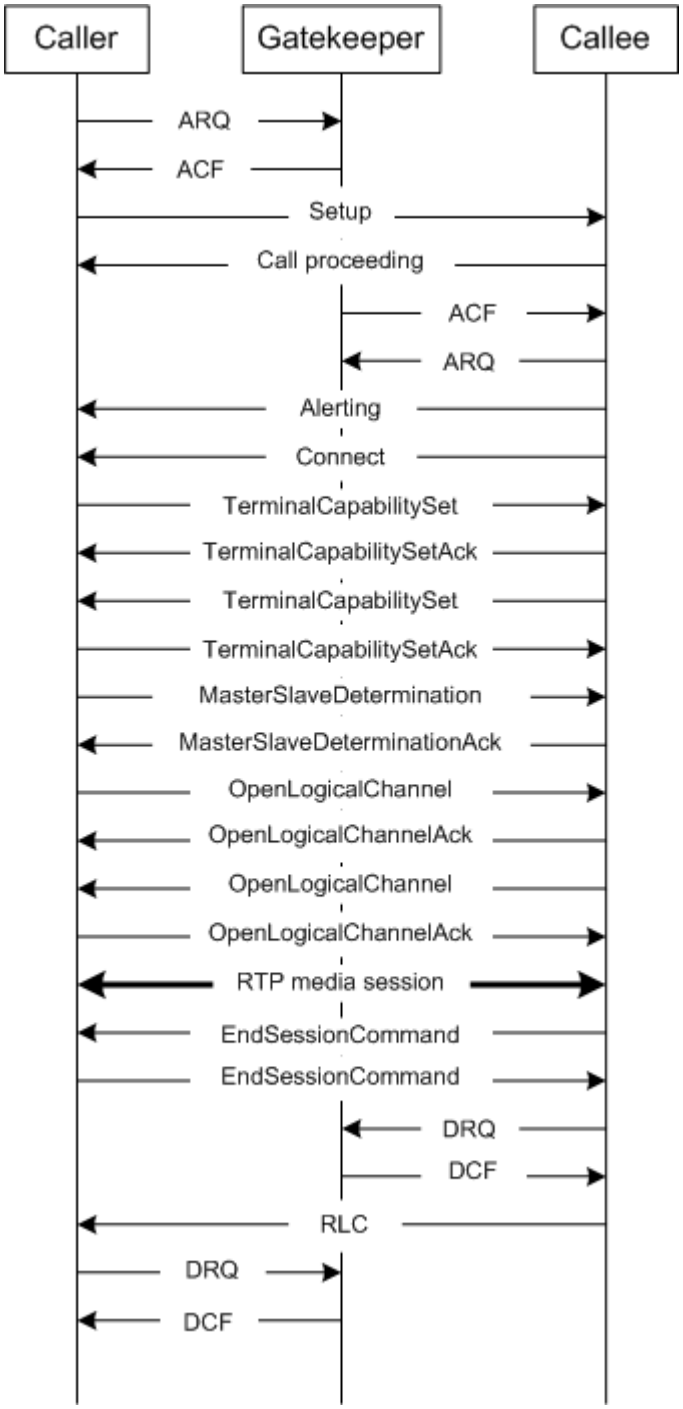


Figure 2-3: H.323 call flow

2.2.2 SIP

The Session Initiation Protocol (SIP) [49] is an application-layer control protocol for the creation, modification and termination of multimedia sessions with one or more participants. SIP is a signaling protocol and it cannot be considered a full communications system; it is rather a component which can be used in conjunction with other IETF protocols to build a complete multimedia architecture. Such protocols are the Real-time Transport Protocol (RTP) [20] for transporting real-time data and providing QoS feedback, the Real-Time Streaming Protocol (RTSP) [21] for controlling delivery of streaming media, the Media Gateway Control Protocol

(MEGACO) [22] for controlling gateways to the PSTN, and the Session Description Protocol (SDP) [23] for describing multimedia sessions. While SIP requires all these protocols in order to offer complete services, its functionality and operation does not depend on any of these protocols.

SIP is a text encoded protocol based on the same principles of the HyperText Transport Protocol (HTTP) [24], which is used for web browsing, and the Simple Mail Transport Protocol (SMTP) [25], which is the main protocol used for exchanging e-mails over the Internet. While SIP is used for P2P communications, it uses a client-server transaction model similar to HTTP. When a User Agent (UA), that is a software or hardware terminal that acts on behalf of the end user, initiates a SIP request it is considered a client and the called party a server; these roles can be reversed since each one of the two parties can initiate any kind of request. Each request can be of one type of SIP methods; the ones defined in the base SIP specification are shown in Table 2-1 but there other types as well defined in subsequent RFCs that extend functionality of the base protocol. The responses to those requests are numerical and highly similar to those of HTTP; for instance a 200 OK response means that the request has been completed successfully.

Method	**Description**
INVITE	Session setup
ACK	Acknowledgement of final response to INVITE
BYE	Session termination
CANCEL	Pending session cancellation
REGISTER	Registration of a user's URI
OPTIONS	Query of options and capabilities

Table 2-1: SIP methods

A SIP message has the following form (here an INVITE request is depicted):

```
INVITE sip:obrien@miniluv.org SIP/2.0
Via: SIP/2.0/UDP 195.251.161.144:5060; branch=z9hG4bK74b43
Max-Forwards: 70
From: Smith <sip:smith@minitrue.org>; tag=9fxced76sl
To: O'Brien <sip:obrien@miniluv.org>
Call-ID: 3848276298220188511@minitrue.org
CSeq: 1 INVITE
Contact: <sip:192.168.1.8@minitrue.org>
Content-Type: application/sdp
Content-Length: 151
```

The most important fields here are <From>, which indicates the caller, <To>, which indicates the callee and <Contact> which indicates where the callee can contact the caller in order to establish a P2P communication for media transport. These fields use a SIP URI which is a URI of the form user_id@domain_id, where user_id is a unique ID assigned to the user by the SIP service provider and domain_id which is this operator's unique domain name.

The three main elements in a SIP network are: User Agents (UAs), servers and location services. As it was already discussed, UAs originate SIP requests to establish media sessions, and send and receive media. They are the end entities in a SIP network and can play the role of

either a client or a server; they behave as clients when they initiate requests and as servers when they respond to such requests. SIP servers are intermediary entities that assist UAs, among others, in session establishment. There are three types of SIP servers defined in [49]:

- A SIP proxy receives SIP requests from a UA or another proxy and forwards the request to another entity.
- A redirect server receives a request from a UA or proxy and returns a response indicating where the request should be sent.
- A registrar server receives SIP registration requests and updates the UA's information into a location service.

A location service is some kind of database where information about users is kept such as URIs, IP addresses, host names and other. It may also contain routing information about proxies, gateways and other entities. UAs do not directly interact with a location service and the communication between a SIP server and a location service is not based on SIP and thus it is not fully considered SIP element.

Figure 2-4 shows an example message flow of a SIP session initiation and termination; the messages prior to "Media session" belong to the initiation of the call while the rest of them to the termination procedure. Here only one proxy server is used, but a number of other proxies could also be present in the path from the caller to the callee. This message flow could also use no proxies at all, if the caller is aware of the IP address of the callee.

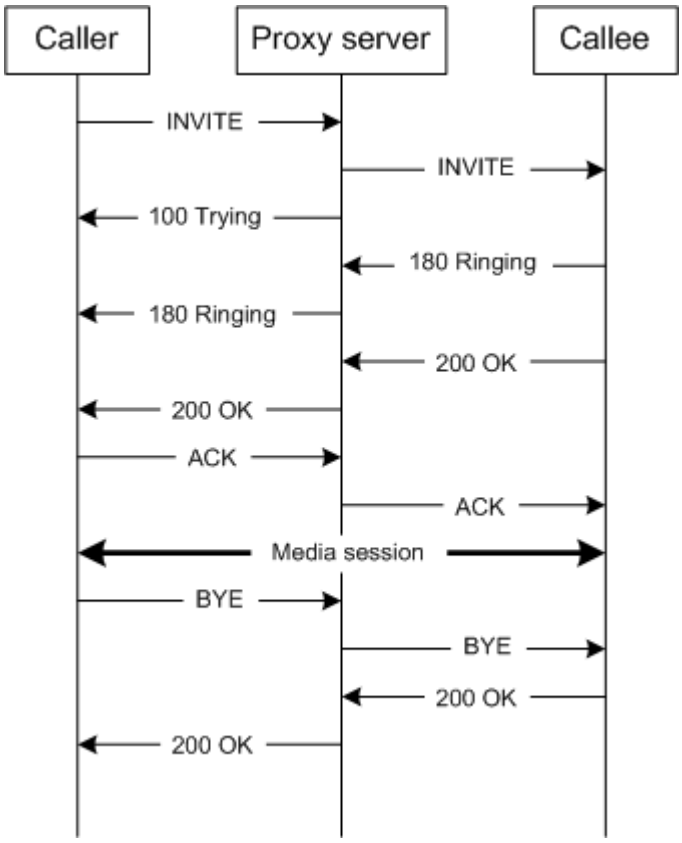


Figure 2-4: SIP call flow

A comparison of SIP and H.323 would reveal a number of differences mainly stemming from the fact that they were developed for different purposes by standards bodies with very different requirements. H.323 was developed by ITU and its design and implementation reflect its PSTN background. On the other hand SIP was developed by the IETF and thus is more close to the Internet logic. The first key difference is the encoding scheme used by each protocol. H.323 uses binary encoding resulting in small message sizes but higher implementation complexity. SIP is text-based and each message can be easily interpreted with no additional tools. Another difference is the level of security. SIP is an Internet protocol so it reuses well known and tested security solutions like Transport Layer Security (TLS) [26]and Secure / Multipurpose Internet Mail Extensions (S/MIME) [85]. Another difference is vendor support; while H.323 has an established base in the industry, SIP is taking over not only in new installations but also in vendors who already use H.323. This is further amplified by the adoption of SIP by mobile operators as the call signaling protocol for 3G networks.

The major strength of SIP is its simplicity; this stems from the fact that it is an Internet protocol following the Internet's architecture. While there are some similarities with H.323 and some existing markets where H.323 dominates, SIP with its simple logic and multitude of extensions is expected to be the signaling protocol for multimedia delivery in any kind of device that use the Internet in the future.

2.3 Authentication, Authorization, Accounting

Authentication, Authorization and Accounting (AAA) are three important security related blocks used in the construction of a network architecture that help operators control access to their networks. A generic AAA architecture [27], especially in the NGNs, will have to support multidomain architectures where different operators interact with end users; these operators can be either network or service providers or both. In such an environment the AAA architecture and the entities implementing it should provide operators a single point of controlling the network which should also be able to interoperate with AAA entities of other operators for roaming reasons.

Authentication can be divided into two types: client and message authentication. Client authentication means that a client presents its identity along with a set of credentials in order to gain access and connect to a network. Message authentication, on the other hand, ensures and verifies the authenticity of the exchanged data. Another concept here is the mutual authentication where the network should also prove its authenticity to the client as well.

Authentication can follow either a two-party or a three-party authentication model. The most prominent example of a two-party model is a client-server interaction where a single or mutual authentication can take place. Figure 2-5 shows the three-party authentication model when the network operator utilizes a AAA infrastructure for authenticating access to the network. At the edge of the network there are a number of entities called Network Access Servers (NASes) which are responsible for authenticating the client. NASes usually have limited resources; at the same time operators prefer to have a single point for controlling their network. Thus, authentication messages pass through NAS and are forwarded to the AAA server who responds back to NAS with the result of the authentication process.

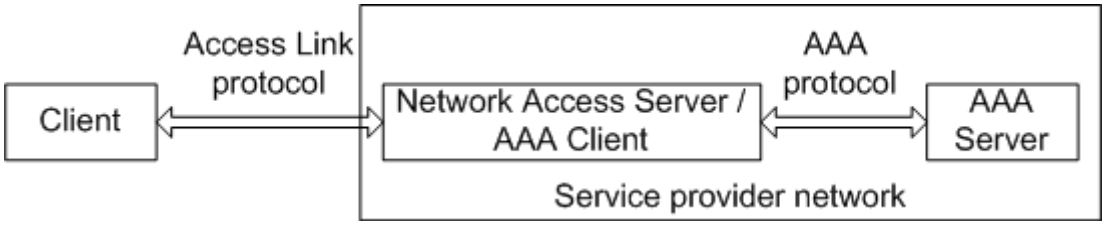


Figure 2-5: Three-party authentication deploying AAA infrastructure

Authorization is the act of determining whether a particular privilege can be granted to the presenter of a particular credential. The privilege can be the right of access to a resource like a network, a database etc. Authentication is frequently confused with authorization, but an authenticated user is not necessarily authorized to use every resource of the network; this is defined by rules and by the specific contract the user has signed with his network operator. Authentication only proves that the user is who he claims he is; authorization states which (usually authenticated) user has access to which resource.

Accounting involves among others the tracking of the usage of the network by the user in units like call minutes or data packets, the conformance to usage policy, data collection for purposes like trend analysis, forensics etc. Accounting is frequently confused with billing; however, billing is one application that can benefit from data accounting. There can be two types of accounting: intra-domain and inter-domain accounting. Intra-domain accounting involves the collection of data on resource usage and its processing within a single administrative domain. Inter-domain accounting, on the other hand, involves the collection of information on resource usage within an administrative domain for use within another administrative domain.

As network sizes grown up, gradually operators realized that in order to handle authentication in large networks, it was more practical to have backend servers that undertake this task and reduce the burden of NASes. Later on, more features were added to these authentication servers which over time were transformed into full AAA servers. In the following sections two of the most dominant AAA protocols will be briefly presented: Remote Access Dial-In User Service (RADIUS) [28] and Diameter [29].

2.3.1 RADIUS

The most widespread AAA protocol today is RADIUS. RADIUS was designed to serve the purpose of allowing a NAS to forward a dial-up user's request and credentials to a backend server following the three-party authentication model. While it was originally designed to support Password Authentication Protocol (PAP) and Challenge-Handshake Authentication Protocol (CHAP) [30], it was also extensible by nature. Therefore, RADIUS was extended to support Extensible Authentication Protocol (EAP) [31] thus supporting more complex EAP-authentication methods [32][33]. Furthermore, RADIUS was later extended to support authorization [34] and accounting [35] procedures as well.

RADIUS is a client-server protocol in which a NAS usually plays the role of the client. The RADIUS client is responsible for passing user requests to the RADIUS server waiting for a response before proceeding to any action. The RADIUS server, on the other hand, is responsible for processing requests and sending the respective response. Figure 2-6 presents an operation example of RADIUS utilizing CHAP in order to authenticate an end user. Here it is clearly depicted that the RADIUS server is the one who decides whether the proof of user authenticity is

valid or not and NAS does only transfer the messages and allows the user to utilize the network resources based on RADIUS server's response.

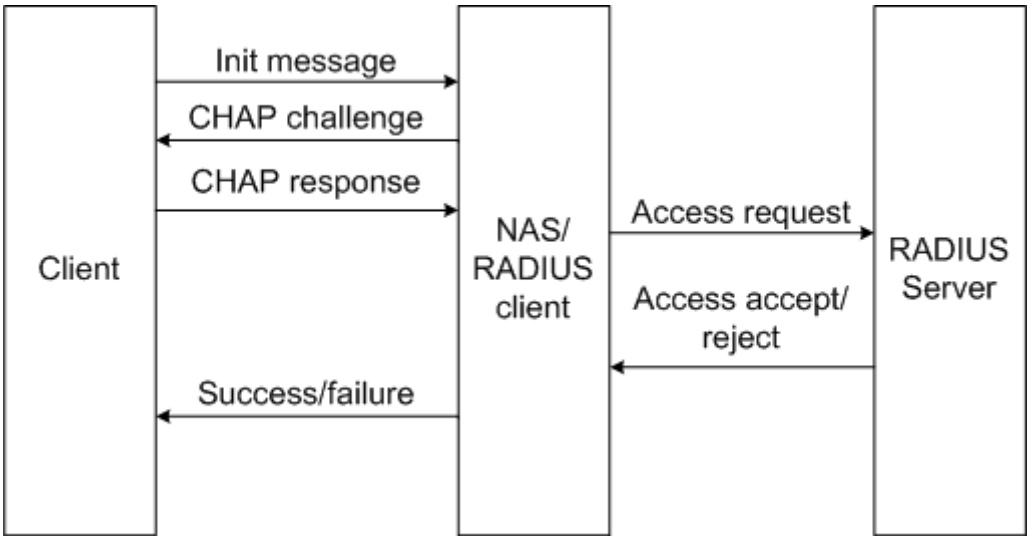


Figure 2-6: RADIUS messaging using CHAP for authentication

RADIUS offers operators central control of user access to their networks; however, there are a number of issues concerning its operation from which the most important are presented here. First of all, security protection in RADIUS is rather primitive. Two main functions are provided: attribute hiding and authentication of certain messages. These functions are performed using MD5 hash function and a shared secret between RADIUS server and client, which is often called the RADIUS shared secret. The use of shared secrets as the basis for providing security functions in RADIUS causes many vulnerabilities in RADIUS deployments. One of these vulnerabilities is that the shared secrets are static and must be configured manually at the NAS resulting in high administrative cost or lower security level if the administrator decides to use the same key for all NASes; furthermore, no prediction for refreshing the key is defined in the base RADIUS specifications. In order to prevent spoofing, the RADIUS server uses the source IP of the packets to lookup the shared secret instead of using the NAS IP. While these two IPs are normally the same, if NAS IP changes then this correspondence is lost and the NAS becomes practically non operable. Another issue with RADIUS is proxy chaining. In real world RADIUS deployments, a number of RADIUS proxies may reside in the path between the NAS and the RADIUS backend server; however, the NAS only shares a secret with the first proxy. This means that the trust between the NAS and the RADIUS server is only transitive; if a rogue proxy exists in the path then security is compromised. Attribute hiding provides selective application layer protection of certain attributes like passwords. This mechanism does not protect in any way whole RADIUS messages or the protocol layers these messages are riding on so IP addresses can be easily spoofed and non protected attributes can be altered.

2.3.2 Diameter

Around 2000 the IETF RADIUS working group (WG) decided it had concluded its work and a new working group called AAA WG started its work on finding a new AAA protocol which would be the successor of RADIUS. Based on requirements defined for support of NAS in [36] and RADIUS weaknesses, a rather complete set of criteria was defined in [37] serving the overall evaluation

of AAA protocols that went beyond just supporting NASes. The criteria that the new AAA protocol should meet were:

- Scalability
- Failover
- Mutual authentication between client and server
- Transmission level security
- Data object confidentiality
- Data object integrity
- Certificate transport
- Reliable AAA transport mechanisms
- Ability to run over IPv4
- Ability to run over IPv6
- Support for proxy and routing brokers
- Audibility and
- Ability to carry service-specific attributes

After the evaluation of a number of proposals, Diameter [29] was selected and the base protocol was standardized in September 2003. Diameter follows a modular approach where the base protocol specification defines most of the basic building elements like a basic set of messages, attributes and their structure, and accounting procedures since they are required by all applications. All protocols and procedures needed to support other services are considered as add-ons over the base protocol and are named “Diameter applications”. Some of the most common Diameter applications are: NAS [38], Mobile IP [39], EAP [40] and SIP application [41]. NAS application describes the details of the interaction of the Diameter servers with NASes for authentication and other procedures. Diameter Mobile IP application facilitates, among other things, identity verification and authorization mechanisms for end hosts using Mobile IPv4. EAP application defines procedures for carrying EAP exchanges over Diameter messages between the NAS and Diameter servers. Diameter SIP application is designed to be used in conjunction with SIP and provides a Diameter client which is co-located with a SIP server, with the ability to request authentication of users and authorization of SIP resources usage from a Diameter server.

As it has already been analyzed in the previous section, RADIUS is a client-server protocol where the client always issues the requests and the responses are always created by the server. Diameter, on the other hand, is a P2P protocol which means that either the client or the server can create a request or a response. Diameter follows the three-party authentication model as well, while being carefully designed to support multidomain environments. The reliance of RADIUS on hop-by-hop security based on shared secrets has created many problems for modern applications of AAA protocols. Diameter tries to accommodate such shortcomings by mandating support of IP Security (IPsec) [42] for both Diameter clients (usually NASes) and servers. Diameter servers must also support TLS while for clients it is optional; one reason for that is to relax the need for a Public Key Infrastructure (PKI). Therefore, IPsec can be used for edge or intra-domain traffic and TLS is the recommended way for protecting inter-domain traffic. While these two protocols offer hop-by-hop security, the base specification of Diameter also encourages the use of Cryptographic Message Syntax (CMS) extensions [43] for end-to-end security protection of Diameter messages. Another feature of Diameter is authorization of functionality. The fact that a peer has been successfully authenticated does not mean that it is authorized to act as a server supporting the applications it is advertising. Thus, before initiating a connection, a Diameter peer must check that its peers are authorized to act in their acclaimed roles. Furthermore, in a multi-domain environment, the home server prior to authorizing a

session must make sure that the route traversed by the request is accepted and has not gone through untrusted realms.

Since Diameter is the successor of RADIUS, a comparison between them is likely to reveal a number of advantages of Diameter over RADIUS:

- Fail-over. Diameter defines a special flag which should be set when two nodes agree on failure support. When fail-over is enabled all the pending requests to an agent are forwarded to another agent when a transport failure with the first agent is detected.
- Server initiated messages. RADIUS provides only optional support for server initiated messages, thus, it is difficult to implement features like unsolicited disconnects, re-authentications and re-authorizations on demand across heterogeneous networks. For Diameter, on the other hand, support for such messages is mandatory.
- Reliable transport. RADIUS utilizes UDP as transport protocol making reliability an issue, especially regarding accounting. Diameter runs over TCP or SCTP offering reliability; this choice, however, makes the deployment of Diameter more difficult.
- Capability negotiation. In RADIUS, the client and server do not have any way of indicating their support of various attributes to each other making capability discovery and negotiation a very difficult task. Diameter includes support for error handling and capability negotiation.
- Security. RADIUS defines an application-layer integrity protection mechanism that is only required for Access Response messages, while authentication is based on shared secrets and trust is established only in a hop-by-hop manner. Diameter defines both transmission level and end-to-end security and requires mandatory support of IPsec and optional TLS support at the clients.
- Inter-domain support. RADIUS does define the roles of agents and proxies clearly resulting in behavior variances between implementations and causing interoperability and security problems. Diameter addresses these limitations by explicitly defining the behavior of agents and proxies and providing support for inter-domain roaming, message routing and transmission level security.
- Peer discovery and configuration. RADIUS implementations typically require the names, addresses and shared secrets of clients and servers be manually configured resulting in heavy administrative burden. Diameter enables dynamic discovery of peers and derivation of session keys.
- Compatibility. While RADIUS and Diameter follow different approaches it is still possible to make them compatible through appropriate translation gateways.

2.4 IP Multimedia Subsystem

As it has already been argued in previous sections, 3G networks aim to merge two of the most successful paradigms in communications: cellular networks and the Internet. The IP Multimedia Subsystem (IMS) [1] is a set of specifications that describe the architecture for implementing IP based telephony and multimedia services in 3G and consequently in NGN. IMS is not a new technology; it is rather an implementation of existing Internet standards that bring the control to the core of the network. The protocols that play cardinal role for the operation of IMS are: SIP as session control protocol, Diameter as AAA protocol followed by appropriate Diameter applications like Diameter SIP application, Real-time Transport Protocol (RTP) and RTP Control Protocol (RTCP) [20] to transport real-time media like video and audio, Common Open Policy Service (COPS) [44] for policies transfer and ITU-T Recommendation H.248 [45] and its packages to control specific types of nodes.

The idea of IMS is to offer multimedia services everywhere and at any time. While multimedia services can be supported in networks used today, the difference with IMS is in a number of issues concerning among others accounting, charging and Quality of Service (QoS). The systems that offer support in these operations are numerous consisting of many different disparate systems rather than one architecture supporting all media types. The primary purpose of IMS is to provide session control at the core of the network, while enabling other support needed to provide those services, regardless of media type. This means that one common control plane is used for video, voice, data, messaging and any other media format needed. What's more important is that this control plane can support new media types without any modification. The main IMS aims can be summarized in the following:

- Combine the latest trends in technology
- Make the mobile Internet paradigm come true
- Create a common platform to develop diverse multimedia services and
- Create a mechanism to boost margins due to extra usage of mobile packet switched networks.

Following these aims, IMS was defined as a framework created for the purpose of delivering IP multimedia services to end users; this framework needs to meet certain requirements and more specifically to support:

- the establishment of IP multimedia sessions
- a mechanism to negotiate QoS
- interworking with the Internet and circuit switched networks
- roaming and inter-domain environments
- strong control imposed by the operator with respect to the services delivered to the end user
- rapid service creation
- access independence

Starting the analysis of the aforementioned requirements, it can be argued that the most important services for users are audio and video communications. While multimedia communications were already standardized in previous 3GPP releases, those multimedia communications take place over the circuit switched network rather than the packet switched network. The main service delivered by IMS is multimedia services over packet switched networks offering the possibility of simultaneous existence of several media types.

A key component of IMS is the ability to negotiate a certain QoS level. The QoS is determined by a number of factors such as the maximum bandwidth that can be allocated to the user based on the user's subscription or the current state of the network. IMS allows operators to define different QoS levels, thus allowing operators to define different groups of customers based on their needs.

IMS is required to have two targets when considering interworking: interworking with the Internet and interworking with circuit switched networks. Support of interworking with the Internet will offer a great number of potential sources and destinations for multimedia sessions to end users. Interworking with circuit switched networks will offer access to networks like the Public Switched Telephone Network (PSTN) or existing cellular networks.

Roaming has been implemented in 2G where users have been able to roam to different networks, especially when visiting a foreign country, subject to roaming agreements signed between their home and the visited network. IMS supports this feature as well, offering the

possibility to the end users to initiate multimedia sessions even when the access to their home network is not available.

Service control can be enforced with policies which fall into two categories: general policies and individual policies. General policies apply to all users in the network and they can be rules like, for instance, enforcing the use of one video codec that is more efficient over another not so efficient codec, in order to conserve bandwidth. The second category includes policies that apply to individual users based on the contract of each user with the network operator. If, for example, the subscription of a user does not include video call, then a video call session initiation will be prevented by the operator even if both the network and the end user's terminal can support video calls.

In IMS, rapid service creation is succeeded by standardizing service capabilities rather than services. In 2G every new service had to be standardized in order to be operable and supported by the operators. This standardization process and interoperability tests caused significant delays in service deployment and even then sometimes there was no guarantee that the service would work when roaming to another network. By standardizing service capabilities only, IMS tries to reduce the time it takes to introduce a new service to the market.

IMS, being an IP based network, supports a number of different access layer technologies offering access independency. While the first release of IMS focused on GPRS access (both in 2G and 3G networks), later releases make possible the access of IMS through WLAN, Asymmetric Digital Subscriber Line (ADSL), Cable Modem etc.

2.5 Summary

Advances in wireless networking technology have made it possible for end users to receive services everywhere. While different network access technologies have been developed and succeeded in penetrating the market, there is a trend towards the convergence of these technologies into an all-IP heterogeneous architecture. This all-IP architecture will borrow advantages from both 3G and WLAN worlds and set the basis for the deployment of novel and better applications in the near future.

In this environment, delivery of multimedia services, being a very demanding application class, is a very interesting and active research field. There are numerous protocols enabling multimedia delivery as well as protocols offering AAA services over IP networks. These protocols mainly operate on the Internet today; however, they can also be used in NGNs since they operate over IP. IMS is such an architecture which promises to offer secure multimedia services over NGNs providing ensured QoS level, charging and central control to network operators.

Chapter 3 - Survey of secure handoff optimization schemes

One of the most challenging problems of NGN is handoff management since a large number of networks with heterogeneous link layer technologies belonging to different operators may offer their services. Such networks will naturally overlap with each other and mobile users will need to frequently handoff among them for a number of reasons, including the quest for higher speeds and/or lower cost. Handoffs between such hybrid networks should be fast enough to support demanding applications, like multimedia content delivery, but also secure since different network providers are involved. This gets even more complicated considering that network providers may not simultaneously be multimedia service providers as it is the case today.

In order to support security operations in a large scale the employment of an AAA protocol is mandated; however, this adds more delay to the handoff process. This chapter presents the performance problems raised in multimedia services during handoff when security services are also required. Next an analysis of the prominent methods proposed so far that optimize the secure handoff process in terms of delay and are suitable for uninterruptible secure multimedia service delivery is provided. Furthermore, these methods are compared to each other based on certain criteria and this comparison reveals, among others, that no method takes into consideration the privacy protection of end users.

3.1 Problem statement

The aforementioned environment will not only provide the basis of new and better applications, but shall impose new technical problems as well. The tradeoff between security and efficiency is one of the most challenging issues in wireless communications and this is not likely to change, at least in the near future. This tradeoff is especially true in environments where the network provider is different from the service provider. In such cases, the end user must be authenticated to both providers in order to use a single service, and to many more if he plans to use more services or perform a handoff. That is, authenticate to the network provider and additionally perform a number of authentications as many as the service providers, or in case of a handoff, authenticate to the new network provider and re-authenticate to the services he already receives. For example, considering SIP registration in UMTS networks, the user must first authenticate in order to access the network and then authenticate (again) to access SIP services. These authentications are accomplished using AAA protocols, like RADIUS [28] or DIAMETER [29], which are in general costly, especially when the home network is many network hops away from the visiting network. This delay is even more crucial and must be carefully considered during handoffs.

Mobility management protocols like Mobile IP [46] and Cellular IP [47] do not consider AAA operations during handoff. In order to cope with long delays, a number of techniques have been proposed to optimize the handoff procedure. One way to achieve this is by minimizing the delays introduced by AAA interactions during the handoff phase. This survey looks into such schemes and compares them in terms of security, efficiency and scalability; a short description and a critical constructive view of each method is provided as well.

To better analyze the problem, in the following we describe a typical scenario of using multimedia services over all-IP wireless heterogeneous networks. Under this context, a user terminal can roam between networks that utilize different access technologies like IEEE 802.11, 802.16 and UMTS. Each of these networks may belong to the same or to a different administrative domain. For example the user (terminal) is able to move from a WLAN to another WLAN, which belongs to the same operator (performing an intra-domain handoff) or from a WLAN to an UMTS network, which belongs to a different administrative domain (performing an inter-domain handoff). In general, inter-domain handoffs tend to be more expensive than intra-domain because of the network delays imposed by the distance, in terms of network hops, between the local and the home domain.

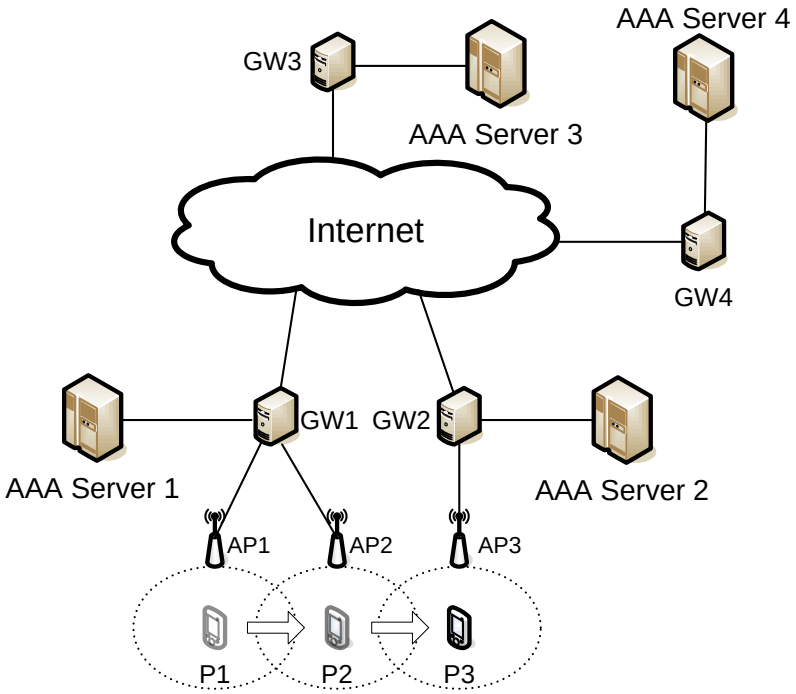


Figure 3-1: General heterogeneous network architecture

Figure 3-1 depicts the general architecture of a network composed of different technologies and administrative domains; for instance, access points AP1 and AP2, which reside in the same administrative domain, could use 802.11, and AP3 could be a cellular operator's access point. Each domain is represented by a gateway (GW) and an AAA server. This, of course, is a simplified representation and every gateway could act either like a true gateway to the Internet, or a multimedia server, or a directory for AAA servers lookup, etc.

Next, we consider a scenario where a terminal using a multimedia service from a server residing out of the local domain is executing an inter-domain secure handoff. Initially the terminal is at position P2, using a multimedia service from GW3 and its home domain is controlled by AAA Server 4. If the user moves to position P3, a handoff is going to occur. What should be assured during this handoff is the continuation of the multimedia service without severe quality degradation. The procedure that has to take place is as follows: the terminal first requests access to the network from GW2, which refers to its local AAA Server 2 which in turn refers to AAA Server 4 to authenticate it. After the terminal is granted access to the network it must access the multimedia service, so through GW2 the terminal requests the service from

GW3, which refers to its local AAA Server 3, which in turn refers to AAA Server 4 to authenticate it.

The aforementioned example is the worst-case scenario where the home domain, the local domain and the multimedia server reside in three different places. In such cases the operations taking place during the handoff procedure result in long delays, especially when the involved servers are distant from each other. However, the previous worse case situation describes a non-optimized scheme, which does not consider the problems related to multimedia services during secure handoffs. The next sections concentrate on schemes that try to solve or mitigate these problems.

3.2 Proposed solutions

In the following we constructively describe all the major secure handoff optimization schemes proposed until now. This is necessary for the qualitative analysis provided later; moreover, for the sake of completeness, we decided to also reference a number of other subordinate schemes which bare similarities with the ones presented hereupon.

3.2.1 OIRPMSA

In [48] the authors are examining the case of a secure handoff using Mobile IP [46] and SIP [49]. Their scheme namely “Optimized Integrated Registration Procedure of Mobile IP and SIP with AAA operations” (OIRPMSA) attempts to reduce the roundtrips needed between the mobile terminal and the home AAA server. Normally, 3 such roundtrips are needed:

1. Mobile IP registration (Figure 3-2).
2. SIP register (Figure 3-3, actions 1-6): This message gets a 401 (Unauthorized) response which, among others, includes challenge information.
3. SIP register (Figure 3-3, actions 7-12): The terminal tries to authenticate using the previous challenge information and if the authentication is successful it gets a 200 (OK) response.

The suggestion of this work is to minimize the delay imposed by the second message by “converting” it to a local roundtrip between the mobile node (MN) and the Local AAA Server (AAAL). Their idea is illustrated in Figure 3-4 and is as follows: when the MN sends the first message (Mobile IP registration) it states that a SIP registration is about to follow. The home AAA server’s (AAAH) response includes some challenge information which is stored in the local AAA server and will be used later. Then the mobile terminal sends a SIP register (action 9) towards the local AAA server which responds with a 401 (Unauthorized) response (action 10) that includes the previous challenge information. Finally, another SIP register message follows that goes all the way to the home AAA server (actions 11-16).

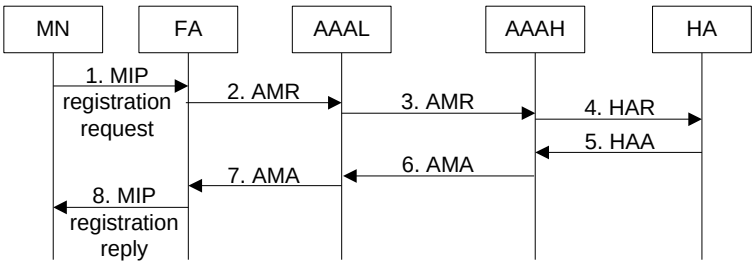


Figure 3-2: Mobile IP registration with AAA operations

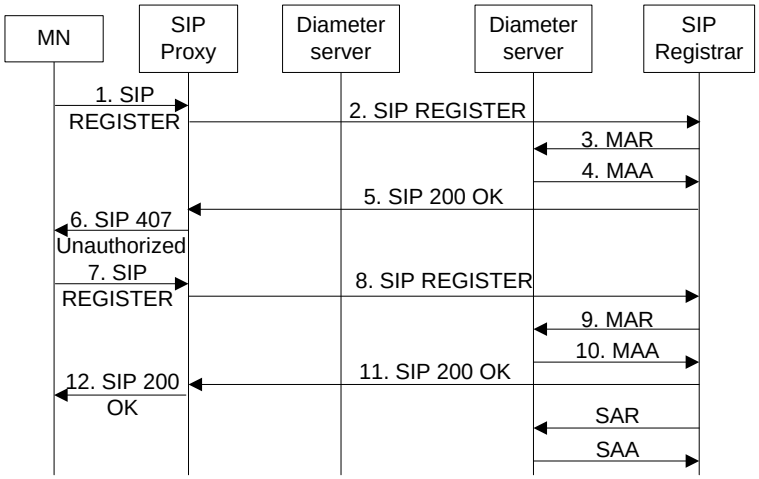


Figure 3-3: SIP registration with AAA operations

One shortcoming of this approach is that it is assumed that the network provider is the same as the service provider. Although this, in many cases, is true today, it is not the general case and of course it is not certain that it will still hold after a few years. Another weakness of this scheme is that the agents used by Mobile IP (FA-Foreign Agent, HA-Home Agent) should be co-located with the SIP proxy and SIP registrar respectively as shown in Figure 3-4.

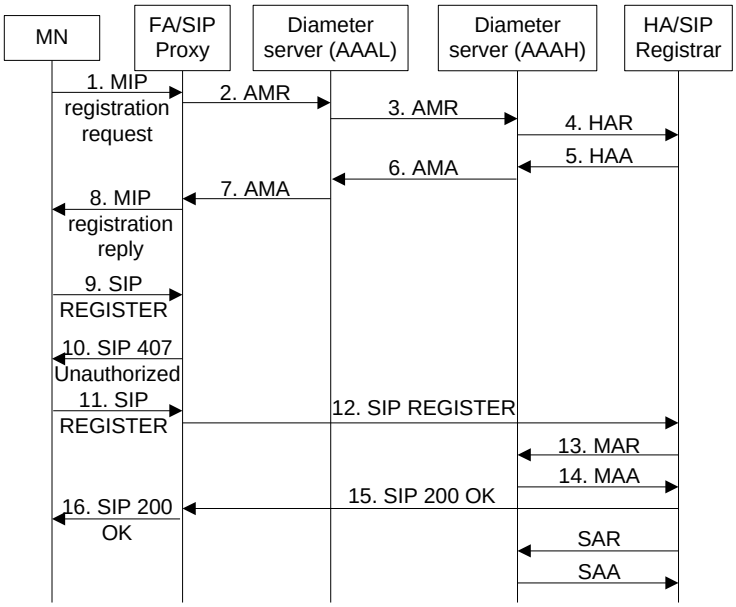


Figure 3-4: OIRPMSA signaling

3.2.2 MPA

Another handoff optimization scheme, presented by Dutta et al. [50][51], is MPA which stands for "Media – independent Pre - Authentication". MPA is a framework that can work over any link layer and can cooperate with any mobility management scheme. To support this claim some of the authors in another work [52] have combined MPA with IEEE 802.21 [53] as mobility management protocol. MPA framework assumes that the following elements exist in every network: Authentication Agent (AA), Configuration Agent (CA) and Access Router (AR). The basic steps taken by MPA are as follows:

1. Pre-authentication (Figure 3-5, action 1): The mobile terminal finds the IP addresses of AA, CA and AR. It performs pre-authentication with the AA, creating security associations with AA, CA and AR.
2. Pre-configuration (Figure 3-5, actions 4-5): When the mobile node is about to change its point of attachment, it performs pre-configuration using the CA to obtain new IP address and other configuration parameters (action 4). Using a tunnel management protocol, the mobile node sets up a tunnel with an access router from the candidate network (action 5).
3. Secure proactive handover (Figure 3-5, actions 6-7): The terminal starts a binding update over the established tunnel by using both the old and the new IP addresses. This means that it has already executed a higher layer handoff before a link layer handoff.
4. Switching (Figure 3-5, actions 8-9): The mobile node completes the binding update and executes the link layer handoff. After that, the mobile node starts communicating from the new point of attachment and deletes or disables the established tunnel.

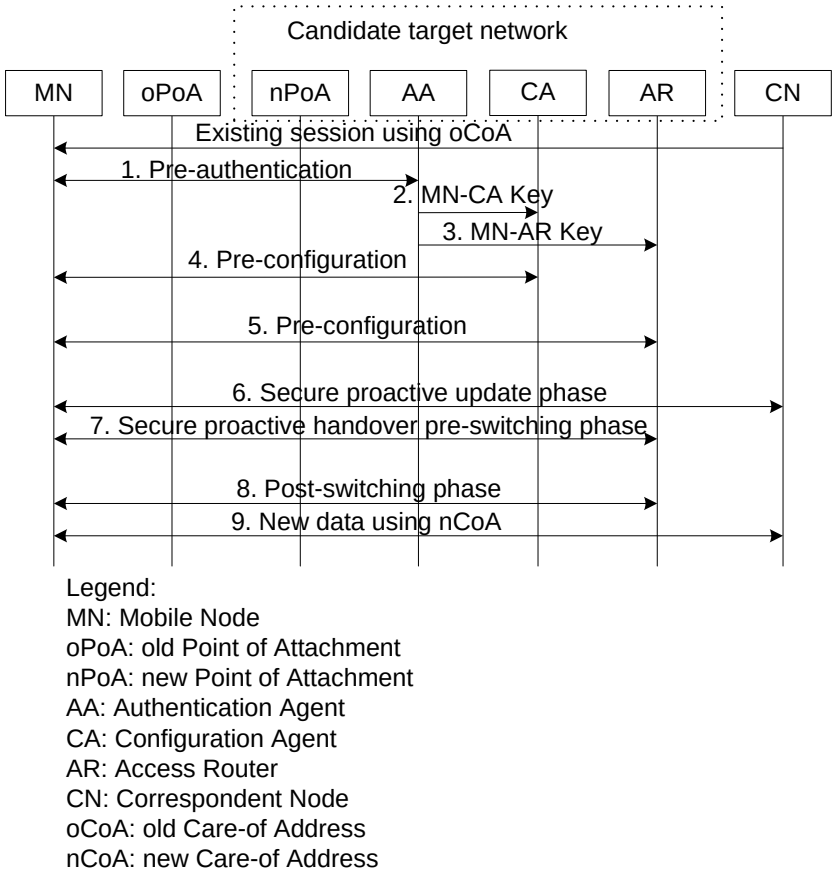


Figure 3-5: MPA signaling flow

In [50] a complete handoff solution is proposed which optimizes a number of parameters that add to handoff delay, like IP address assignment. In the testbed implemented by the authors, a non-MPA handoff took 4 seconds, whereas MPA handoffs to different platforms took from 14 to 600 ms.

3.2.3 Shadow registration

In [54] the Shadow Registration method is proposed in order to optimize secure handoffs. According to this scheme a security association is established between the mobile terminal and every neighboring AAA server before the former enters the domain the server controls. Using Figure 3-6 as reference, when the mobile terminal resides in the central cell, a registration procedure is performed with all 6 neighboring cells. When this happens the necessary AAA operations are processed locally in this new domain without communicating with the terminal's home domain. Specifically, the authors examine two cases where Shadow Registration could be used; the Mobile IP case and the SIP case. In both cases, during the handoff, the requested AAA operations are processed locally and after the completion of the handoff a separate process is executed where security associations are sent to the new neighboring domains of the mobile terminal.

Based on the concept of Shadow Registration, Han et al. [55] have proposed Region-based Shadow Registration (RSR). RSR is trying to solve the problems of heavy traffic and waste of resources introduced by Shadow Registration. Instead of establishing a security association with every neighboring domain, RSR divides the terminal's current cell in regions (a, b and c in Figure

3-6) and performs a Shadow Registration only when the terminal moves to a section with high probability to handoff. When the mobile node resides near the cell core, no Shadow Registration is performed. The outer zone of the cell is divided in three regions and each region is adjacent to two neighboring cells; when the mobile terminal moves to one of these three regions, a Shadow Registration is performed for the two neighboring cells. For example in Figure 3-6, when the mobile terminal moves from the Core to region b, a Shadow Registration procedure is performed with cells 3 and 4. By this way, the two schemes have the same effect in reducing the handoff delay while RSR reduces the traffic between the domains.

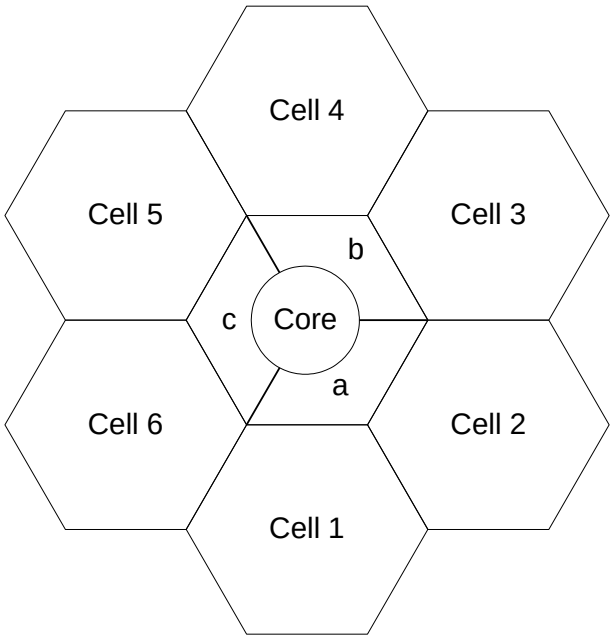


Figure 3-6: Regional cell division

Another similar approach is [56] which uses the Frequent Handoff Region (FHR) concept. Considering this scheme, the network administrators collect information about the location of the access points and the movement of users and construct a weighted directed graph. With the help of FHR Selection algorithm, adjacent access points are grouped in FHRs and the mobile terminal is authenticated in advance towards the access points that belong to the same FHR.

A disadvantage of the above methods stems from the fact that in future heterogeneous networks the areas of coverage in most cases will overlap. In such an environment, when a mobile terminal roams in an area covered by a WLAN access point it is possible that this area is also covered by other WLAN, WMAN and/or UMTS access towers. Under these circumstances it is not obvious which the neighboring domains are, let alone that there can be many of them. This results to excessive signaling (especially in SR) and difficulties in determining which the neighboring cells (in RSR case) are. This maybe not seems to be a problem with FHR, but that would require from the administrators to collect new information every time a new network is deployed in the same area.

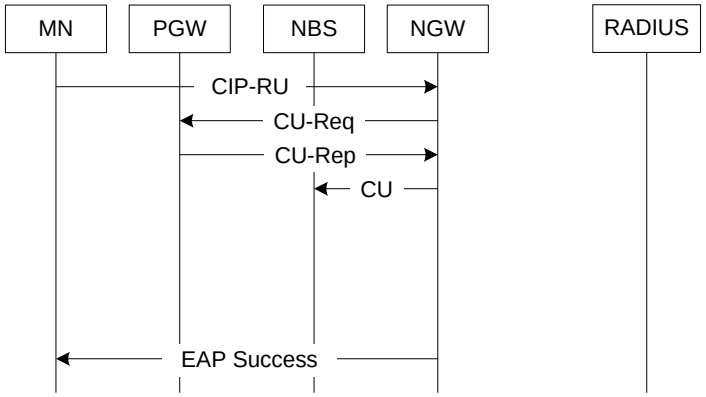
3.2.4 AAA context transfer

The solution proposed in [57] is product of the IST EVOLUTE project that tries to provide secure and seamless multimedia services over heterogeneous all-IP networks using the concept of

context transfer. In RFC 3374 document [58] the context and context transfer terms are defined as:

- Context: The information on the current state of a service required to re-establish the service on a new subnet without having to perform the entire protocol exchange with the mobile host from scratch.
- Context transfer: The movement of context from one router or other network entity to another as a means of re-establishing specific services on a new subnet or collection of subnets.

EVOLUTE uses Mobile IP and SIP for inter-domain mobility management, while for inta-domain mobility uses protocols like Cellular IP and Hierarchical Mobile IP [59]. In order to provide secure access to multimedia services, the EAP-TLS [60] solution is used as the authentication protocol. Figure 3-7 depicts the signaling flow when the context transfer is not used; on the downside, Figure 3-8 shows the same signaling flow when the context transfer is enabled. When the mobile terminal sends a request to handoff to a new gateway (NGW), this NGW gets the context from the previous gateway (PGW) whose IP is indicated in the terminal's request. The terminal is then authenticated to the NGW without contacting its home domain.



Legend:
MN: Mobile Node
PGW: Previous Gateway
NBS: New Base Station
NGW: New Gateway
CIP-RU: Cellular-IP Route Update
Req: Request
Resp: Response

Figure 3-7: EAP-TLS exchange without context transfer

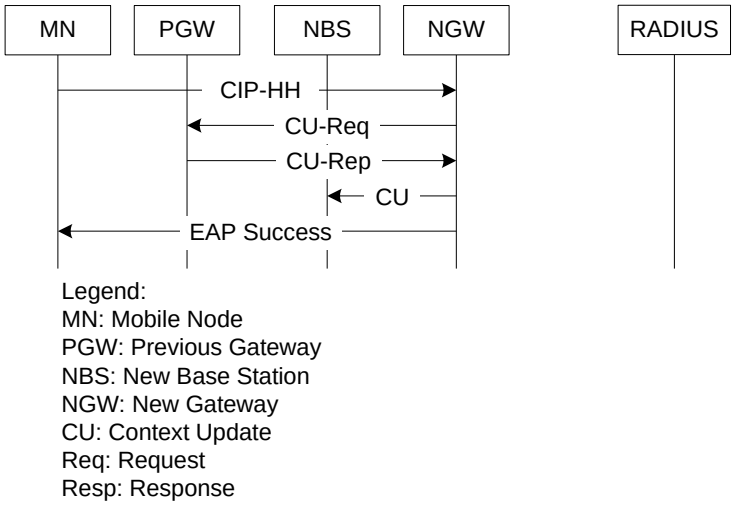


Figure 3-8: EAP-TLS exchange with context transfer

When the method of context transfer is employed, it is assumed that the new network can support the services offered from the previous one. However, in a heterogeneous environment, this might not always be the case and the mobile terminal might have to contact its home domain for renegotiation about the offered services.

3.2.5 Peer-to-Peer security context transfer

The work of Braun and Kim [61] combines the concepts of security context and P2P networks to optimize authentication in heterogeneous wireless networks. According to this approach a security context contains authentication credentials in the form: {random number or nonce or challenge, expected response}. Such security contexts are created at the home domain by the home AAA Server and delivered to AAA Servers (or Brokers) that reside between the home domain and the local domain (and therefore are closer to the mobile terminal). The AAA Brokers take the authentication decision after a corresponding mobile terminal's request based on security contexts; for this reason they are referred as Security Context Controllers (SCCs) as well. SCCs are organized in a peer-to-peer manner and they are able to detect each other using mechanisms originated from P2P networks.

An example demonstrating peer-to-peer organization of SCCs is illustrated in Figure 3-9. At first, the mobile terminal resides in the area covered by SCC 1 which has already acquired the security context from AAAH via SCCx. During this transfer, AAAH and SCCx have stored pointers to the current security context which resides in SCC 1. When SCC 1 gets the security context it broadcasts its acquisition to its neighbors. This way, when the user moves to the area covered by SCC 2, the new SCC acquires the security context from SCC 1 and informs AAAH. If this is not the case, say the user switches off the device in SCC 1 and moves to the area of SCC 3, SCC 3 is not aware of the stored security context in SCC 1 and has to request a new one from AAAH. This request is routed through SCCy and SCCx; when the request meets SCCx, SCCx returns a response that SCC 1 has a security context for the corresponding user. When SCC 3 gets this response it requests the security context from SCC 1 and informs AAAH that is now controlling the security context of the user.

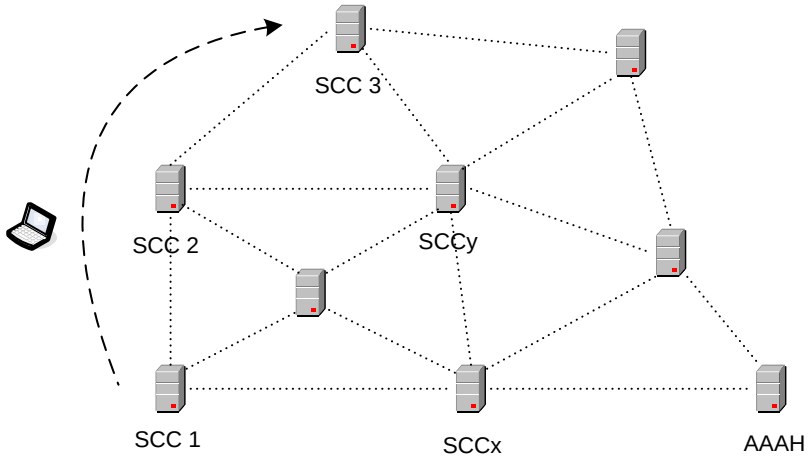


Figure 3-9: P2P organization of SCCs

3.2.6 Optimistic access

In order to minimize the re-authentication delay, an alternative technique is proposed by Aura and Roe in [62]. According to this approach the mobile terminal, instead of executing a so-called strong authentication during the handoff process, it is granted optimistic access to the new network delaying the strong authentication which is held after the handoff is completed.

More specifically as shown in Figure 3-10, when the mobile node (MN) handoffs to the new network a light (fast) authentication takes place. If this authentication is successful the MN is authorized a so-called optimistic access and can communicate through the new network. When the handoff process is complete, the MN must be involved in a new strong authentication to continue using the resources of the new network. After the end of this authentication the Optimistic Access scheme completes its purpose.

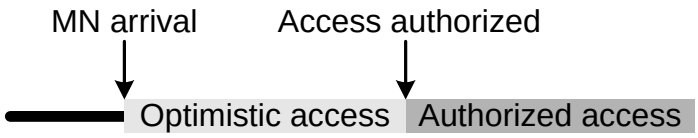


Figure 3-10: Light and strong authentication in optimistic access scheme

The target of the proposed work is to conclude to a protocol that allows optimistic access to well behaving mobile nodes while reducing the risk of possible misuse. A customer that has paid for some other service or is following some rules is considered well-behaving, while unknown users should perform a strong authentication during the handoff process. The protocol operates as following:

1. The old access point sends to the mobile terminal a secret key K and a credential C over a secure channel.
2. The new access point broadcasts challenges periodically and the mobile terminal retrieves one such challenge.
3. The terminal computes and sends towards the new access point a keyed one-way function of the challenge and the secret key K. It also presents the credential C acquired from the old

access point which contains some trust parameters about its previous good behavior. The new access point recovers K from the credential and decides to grant optimistic access or not by evaluating the trust parameters.

When the secure handoff procedure is completed, the strong authentication must take place in a short time. The exact period of optimistic access and authentication method are not covered by the above protocol and are matters of choice of the network administrator.

This protocol makes an obvious trade-off between security and performance. The vulnerability left is the small window of light authentication between the handoff and the strong authentication; still, in order for someone to misuse the resources of the network, the light authentication should be based on a weak protocol or no authentication at all. Another security issue of the above protocol is that all the access points of the network should share a secret key and this can be especially dangerous if a single access point leaks the key. Thus, key management issues concerning optimistic access must be carefully considered and further investigated.

3.2.7 Other schemes

This section references secure handoff optimization schemes that could not be included in the conducted comparison. These methods are left out because they either exhibit many similarities with the mechanisms already described or they do not comprise a general solution supporting secure handoffs between heterogeneous networks for multimedia services delivery.

The first such scheme is *Mobility-adjusted Authentication Protocol* [63] (MAP) which utilizes symmetric cryptography in conjunction with the security context concept relying on special Security Context Nodes (SCNs). The work in [64] reviews fast authentication methods for 802.11 WLAN's for seamless mobility across administrative domains. The authors of [65] use the concept of AAA brokers while their novelty is a formula for finding the best spots within the network architecture to place these brokers. A method which is similar to the Shadow registration concept, and especially to the Frequent Handoff Region variation, is presented in [66]; the difference here is that this method does not require manual configuration and the system is auto configured instead. In [67] the authors are based on Hierarchical Mobile IPv6 (HMIPv6) which is an enhancement to Mobile IPv6 (MIPv6) that supports fast handoffs. Their proposal integrates the Diameter protocol to support authenticated access during roaming. Another approach is the Secure, QoS-enabled Mobility (SeQoMo) [68] architecture which is comprised of components that can be co-located with existing routers, access points, mobile nodes etc. to provide fast handoffs to HMIPv6 based networks. In [69] six approaches are proposed for session state re-establishment in intra-domain scenarios; these approaches are based on the combination of concepts like Fast Handover for Mobile IPv6 (FMIPv6), HMIPv6, AAA and context transfer. Finally, the work presented in [70] shows how seamless handoffs can be realized in UMTS–WLAN integrated networks.

	OSI layer	Security					Efficiency			
		Public vs secret key	Mutual authenti-cation	Privacy	Non – repudia-tion	Assumes trust between domains	Roundtrips		Credential creation	Perform-ance improve-ment (%)[1]
Scheme identifier							During handoff	Total		
OIRPMSA	3, 7	not defined				full	3	3	on-the-fly	18.2% - 33.3%
MPA	3	not defined	√			no	1	6	on-the-fly	85% - 99.65% [2]
Shadow Registra-tion	3 or 7 or both	not defined				full	1	3	on-the-fly	-
AAA context transfer	3 or 7 or both	public key	√		√	full	1	2	pre-computed	78.5%
P2P context transfer	any[3]	not defined				full	min 1, max 2	min 1, max 3	pre-computed	-
Optimistic access	any[4]	secret key		√		full	2	3	on-the-fly	-

Table 3-1: Secure handoff optimization schemes comparison (continued on next page)

3.3 Criteria and comparison

Table 3-1 gives a comparison of the analyzed schemes based on selected criteria. In the rest of this section these criteria are explained and every scheme is compared to each other based on them. Using this approach, a clear view of the advantages and disadvantages of each scheme is provided.

3.3.1 OSI layer

This criterion shows in which OSI layer is the solution to the fast secure handoff problem implemented. We only consider methods operating at either the network or application layer or both of them. When a protocol operates at the network layer, then it offers secure access to a different network even if the new network uses different link layer technology from the old one;

[1] The findings of this column cannot be used to compare the schemes because every scheme optimizes a differently configured network system.
[2] This scheme improves not only AAA related operations but other network parameters as well, like IP address assignment.
[3] It could be used to any layer where authentication is required.
[4] The authors mainly consider 802.11 link layer technology but argue that the same ideas could be applied to other cases.

	Handoff types supported			Changes required	Standards used	Battery consumption	Scalability	4G ready
	Intra- or inter-domain	Pro-active/ re-active	Fast/ smooth/ seamless					
Scheme identifier								
OIRPMSA	both	Re-active	not defined	Diameter Mobile IP and SIP applications, co-location of agents	Mobile IP, Diameter, Diameter SIP application	Depends on the implement-ation	low	√
MPA	both	Pro-active	seamless	Requires network elements: AA, CA, AR[5]	-	Depends on the implement-ation	medium	√
Shadow Registra-tion	both	Pro-active	not defined	AAA protocol, SIP	Mobile IP, SIP	Depends on the implement-ation	low for SR and FHR, medium for RSR[6]	√
AAA context transfer	both	Re-active	seamless	Cellular IP	Hierarchical Mobile IP, Cellular IP, SIP	high	high	√
P2P context transfer	both	hybrid[7]	not defined	AAA protocol	-	Depends on the implement-ation	high	√
Optimistic access	both	Re-active	not defined	2nd layer protocol	-	low	high	

Table 3-1: (continued from previous page) Secure handoff optimization schemes comparison

this way the interconnection between heterogeneous networks is achieved. To put it in another way, it is possible to offer fast secure handoffs not only to multimedia services but to other applications as well. When a protocol operates at the application layer, then it is targeted at one application each time (in our case multimedia services offered by SIP) and this makes it possible to adapt better to the application's needs. Some schemes operate to both layers offering a complete solution to fast secure handoffs.

OIRPMSA combines authentication at network layer with authentication in application layer in order to provide optimized Mobile IP and SIP registration during handoff. *MPA* operates at the network layer and according to the authors it can be utilized in conjunction with any link layer and mobility management protocol. The *Shadow Registration* concept can be used to either layer, while nothing prevents it to operate to both layers in the case of multimedia services;

[5] AA: Authentication Agent, CA: Configuration Agent, AR: Access Router
[6] SR: Shadow Registration, RSR: Region-based Shadow Registration, FHR: Frequent Handoff Region
[7] The first time the handoff is considered reactive but subsequent handoffs are considered proactive.

moreover, two examples are provided, one for Mobile IP and the other for SIP registration. *AAA context transfer* operates either at network layer or at application layer or at both layers as well, and the testbed that the authors demonstrate uses Cellular IP and SIP protocols. Similarly, the *P2P context transfer* solution can be applied to any layer where fast re-authentication is required. The *Optimistic access* scheme is presented as an 802.11 technology solution. However the authors argue that it is also applicable to other technologies and it seems that it can be used to other OSI layers as well since it is rather a fast re-authentication method than a complete secure handoff scheme targeted specifically at one (specific) layer.

3.3.2 Security

In the security group some security related criteria are examined. The first one looks into whether each method uses *public or secret key* protocols to perform the necessary authentications. The next is *mutual authentication* which examines whether the authentication between the mobile terminal and the new access point is mutual or not. The *privacy* criterion checks if the actual identity of the mobile terminal is revealed to the new domain or not. The next criterion is about whether the new domain is able to claim the *non-repudiation* of the mobile terminal's actions. The last security related criterion is about whether it is assumed that there are *pre-established trust relationships* between the home and the visiting domain or not.

OIRPMSA, MPA, Shadow Registration and *P2P context transfer* do not dictate a special protocol neither the type of cryptography to be used, e.g. symmetric or asymmetric. By contrast, *AAA context transfer*, being based on specific technologies, uses the EAP-TLS protocol which is a public key protocol. *Optimistic access*, on the other hand, is based on shared key cryptography and keyed one way hash functions instead of public key signatures.

OIRPMSA, Shadow Registration, P2P context transfer and *Optimistic access* schemes do not support mutual authentication between the mobile terminal and the new access point. The *MPA* scheme, although it does not define an authentication protocol, it mandates that the chosen one should provide mutual authentication. As *AAA context transfer* utilizes EAP-TLS as the authentication protocol, it is straightforward that it can support mutual authentication.

The only protocol which supports privacy is *Optimistic access*. This protocol does not require any mobile terminal or user identity to be included into the exchanged messages. However, it does not specify what data are inserted into the credentials created by the access points and this is a possible breach of privacy.

The only scheme which offers non-repudiation services is *AAA context transfer*. This is based on the authentication protocol used which is EAP-TLS.

All the protocols except *MPA* assume that there exist pre-established trust relationships between the visiting and the home domain. The authors of the *MPA* scheme argue that their protocol works across different administrative domains based on trust relationships between the mobile terminal and each domain.

3.3.3 Efficiency

This group refers to criteria that examine the analyzed schemes in terms of efficiency. The first two are about *roundtrips* occurring *during the handoff* process and the *total* number of handoffs required by the scheme. The next criterion shows whether the *credentials creation* is performed on-the-fly, e.g. when the credentials are requested from the authentication server (this does not imply that the request is done during the handoff process), or are being pre-computed before

the actual request. Also, there is a *performance improvement* criterion which shows the percentage of performance improvement achieved by each scheme. Nevertheless, the methods discussed hereunder cannot be compared based on this criterion because every scheme concentrates on different specific network configuration which attempts to improve.

OIRPMSA performs 3 roundtrips during the handoff process, which is also equal to the total number of roundtrips performed by this scheme. *MPA* needs a total of 6 roundtrips, while only the last of them is executed during the handoff. When *Shadow Registration* is exploited, 3 roundtrips are performed in total, 1 of which is during handoff. The *AAA context transfer* has the minimum total number of roundtrips, requiring only 2, while 1 is entailed during the handoff process. In the case of *P2P context transfer*, the number of roundtrips during handoff is 1 when the previous SCC is known and 2 when is not. The total number of roundtrips is 1 and 3 respectively; the latter applies because the new SCC must inform the home AAA server that is the current SCC. *Optimistic access* needs 3 roundtrips in total, 2 of which during handoff, in order to complete its aim.

In *OIRPMSA, MPA, Shadow Registration* and *Optimistic access* schemes the creation of the credentials is done on-the-fly, whenever there is such a request from the AAA server. In *AAA context transfer* and *P2P context transfer* the credentials which are essential for the mobile terminal's authentication are pre-computed and can be communicated to foreign AAA servers before they are requested.

The authors of *OIRPMSA* provide a theoretical performance analysis on a system composed of Mobile IP, Diameter and Diameter SIP Application [63] protocols. This analysis showed an expected performance improvement between 18.2 % and 33.3 %. In the case of *MPA*, the results of a specific testbed are provided, which employs the following technologies: 802.11 as link layer technology, Protocol for Carrying Authentication for Network Access (PANA) protocol [72] for network access authentication, DHCP as the configuration protocol, SIP Mobility (SIP-M) [73] as the mobility management protocol, RTP/UDP [74] for carrying voice traffic and RAT (Robust Audio Tool) [75] as the media agent. This scheme does not only improve the delay imposed by AAA operations but other delays as well, such as these imposed from the configuration protocol which tend to be higher. This results to an improvement in the order of 85 % to 99.65 % for MPA. For *Shadow Registration,* while a theoretical analysis is given, the performance improvement is heavily related to the distance between the mobile terminal and the home domain, and thus is difficult to be estimated. When the home domain is very close to the visiting domain the improvement is near zero and increases as the distance between the domains is increasing; so a representative numerical value could not be given. A testbed has been implemented in the case of *AAA context transfer* using Cellular IP, SIP and EAP-TLS protocols, resulting in a performance improvement of 78.5% in the case of multimedia service re-establishment, including both Cellular IP and SIP re-registrations, for an inter-domain handoff scenario. For *P2P context transfer* the analogy between performance improvement and domains distance applies as well; although a specific theoretical example shows an improvement of 35 % this value cannot be used as a general improvement indicator. *Optimistic access* is designed for link layer handoff optimization, so no specific value can be given here.

3.3.4 Handoff types supported

This group of criteria refers to what types of handoffs each scheme is able to support. The first criterion examines whether *intra-domain or/and inter-domain* handoffs are afforded. Inter-domain handoffs tend to be most significant because this handoff type is the most expensive one in terms of delay. Next, it is examined whether the handoff each scheme supports is

proactive or reactive; when a handoff is proactive, the operation of the scheme starts before the handoff is actually needed and signaling is exchanged with the new access point before the mobile terminal connects to it; when a handoff is reactive, the scheme is initiated when the handoff is taking place and signaling with the new access point is done when the mobile terminal connects to it. In RFC 3753 [76], where mobility related terminology is listed, the following definitions are given for *fast/smooth/seamless* handover types:

- Fast handover. A handover that aims primarily to minimize handover latency, with no explicit interest in packet loss.
- Smooth handover. A handover that aims primarily to minimize packet loss, with no explicit concern for additional delays in packet forwarding.
- Seamless handover. A handover in which there is no change in service capability, security, or quality. In practice, some degradation in service is to be expected. The definition of a seamless handover in the practical case should be that other protocols, applications, or end users do not detect any change in service capability, security or quality, which would have a bearing on their (normal) operation. As a consequence, what would be a seamless handover for one less demanding application, might not be equally seamless for another more demanding application.

From the above definitions it seems that the most appropriate type of handoff for secure multimedia delivery is the seamless one. Smooth handoffs are more appropriate for file transfers, while fast handovers could be suitable for multimedia delivery with no security restrictions.

The analysis showed that all methods are able to support both types when the distinction is made between intra-domain and inter-domain handoffs. While *Optimistic access* scheme does not explicitly deals with domains, these two types of handoff can be supported with careful selection of the security credentials and trust parameters.

OIRPMSA is a reactive scheme, while *MPA* and *Shadow Registration* are considered proactive because the new access point has received signaling prior to the handoff initiation phase. *AAA context transfer* and *Optimistic access* methods support reactive handoffs. Finally, *P2P context transfer* is using a mix of the two types and is considered a hybrid solution. The first time the user makes a handoff, it is a reactive one, while the subsequent are proactive; when the user cannot find a security context in the path between the mobile terminal and the home domain then this is also considered a reactive handoff.

MPA and *AAA context transfer* are designed with seamless handoffs in mind. The rest of the methods do not define the support of any special handoff type between fast/smooth/seamless types.

3.3.5 Changes required

This section describes the changes required for the deployment of each scheme. It is stressed that the number, the nature and (most important) the cost of modifications to existing systems required by a scheme for its deployment plays a crucial role in its adoption and the transition to it from existing solutions.

OIRPMSA uses Diameter as AAA protocol and introduces the use of reserved flags of Diameter's Mobile IP and SIP extensions. Another modification required by this method is the co-location of Mobile IP's Foreign Agent with SIP Proxy into a FA/SIP Proxy and Mobile IP's Home Agent with SIP Registrar into a HA/SIP Registrar. *MPA* requires the introduction of three functional elements to each network: (a) an Authentication Agent (AA) which is responsible for

pre-authentication, (b) a Configuration Agent (CA) which is used for secure delivery of IP address and other parameters to the mobile terminal (first part of pre-configuration) and (c) an Access Router (AR) which executes the rest of the pre-configuration phase. *Shadow Registration* necessitates the modification of existing messages of the AAA protocol in use; also, when SIP is used, a new SIP message introduced, namely the ANSWER message. *AAA context transfer* scheme requires modifications or adaptations to the Cellular IP protocol which can be summarized as follows: introduction of three new types of messages, modification of one existing message and a need for a context cache at each gateway. The changes mandated by *P2P context transfer* scheme relate with the AAA protocol; some AAA servers should also act as SCCs (Security Context Controllers) and these nodes should be capable of forming a secure Peer-to-Peer network. *Optimistic access* scheme in its current form requires the modification of the link layer protocol in use; if it is to be used for network or application layer re-authentication then the respective protocols should be altered.

3.3.6 Standards used

This section summarizes the existing standards used by each scheme. The utilization of existing standards plays an important role in the commercial deployment of the proposed systems because it solves most problems causing incompatibilities of implementations between different vendors.

OIRPMSA operates based on Mobile IP, Diameter and an extension of it, namely Diameter SIP application. *MPA*, *P2P context transfer* and *Optimistic access* are more generic approaches and are not based on specific standards. The standards used by *Shadow Registration* method are Mobile IP and SIP. *AAA context transfer* uses Hierarchical Mobile IP, Cellular IP and SIP in its deployment.

3.3.7 Battery consumption

This criterion is concerned with the level of power consumption which is very important in wireless networks where the mobile terminals work on batteries and therefore have limited power reserves. The criteria with which battery consumption is related are mainly the number of roundtrips and the type of cryptography used; asymmetric cryptography tends to be very expensive in terms of power consumption for mobile devices in contrast to symmetric cryptography.

OIRPMSA, MPA, Shadow Registration and *P2P context transfer* schemes do not clarify what type of cryptography will be used for the re-authentication of the mobile terminal, thus the cost in battery consumption is highly depended on the chosen implementation. *AAA context transfer* is considered a high demanding scheme because it uses EAP-TLS as authentication protocol, whereas *Optimistic access* with the use of only symmetric cryptography and one way hash functions is regarded a rather low consumption solution.

3.3.8 Scalability

The level of scalability shows how well is the scheme adapted when the number of networks, network elements and subscribers is increasing. It shows how dynamic is the system considering such changes and its possibility to be deployed in a large scale.

OIRPMSA is perceived to be a low scalability solution mainly because it requires the co-location of Mobile IP and SIP servers. More specifically, it imposes the co-location of Mobile IP's

Foreign Agent (FA) with a SIP Proxy into a single FA/SIP Proxy and the co-location of Mobile IP's Home Agent (HA) with a SIP Registrar into a single HA/SIP Registrar node. This could end up to a performance penalty when the number of serving mobile terminals is high. *MPA* is a scheme with moderate scalability; this arises from the resource demanding nature of this method. When a mobile terminal is about to handoff the procedures of pre-authentication and pre-configuration engage a considerable portion of resources, especially when the request rate is high, which may never be used if the handoff will not take place. In the *Shadow Registration* case a distinction is made between the three analyzed variations, namely Shadow Registration (SR), Region-based Shadow Registration (RSR) and Frequent Handoff Region (FHR). SR is considered a low scalability solution because of the excessive signaling during the registration phase, problem which is partially solved by RSR which in turn is considered a medium scalability method. Finally, FHR is a low scalability scheme because it needs the manual collection of information each time new access points are deployed. *AAA context transfer, P2P context transfer* and *Optimistic access* manage to keep the level of signaling rather low, resulting in high to moderate scalability. This must be proved thought considering either a real deployment or a wide scale simulation.

3.3.9 4G ready

According to the 4G vision, future wireless heterogeneous networks will converge into an all-IP platform. This criterion designates whether the schemes in question are ready to support 4G networks.

From the analysis of each scheme it is obvious that all schemes except *Optimistic access* are ready to support the fourth generation of wireless networks. Optimistic access is not included in the 4G capable schemes list because it operates in the link layer; if, however, the ideas presented by its authors were adapted to higher layers then it could be considered a 4G capable solution.

3.4 Summary

It is envisioned that future wireless networks will converge to an all-IP platform offering more bandwidth consuming services at higher speeds. In such an environment the security of multimedia services, being a demanding class of applications, without perceived degradation by the user is a very challenging issue. The realization of this objective includes the cooperation of mobility management schemes with AAA protocols for the secure and uninterrupted multimedia services provision.

In this chapter an overview of the current most representative secure handoff optimization schemes trying to achieve the aforementioned goals was given. Each scheme was briefly presented and some comments were provided where this was considered purposeful. Finally, a comparison of the schemes was conducted and the criteria of this comparison were further analyzed and explained. The purpose of this chapter is to mark each scheme's advantages and disadvantages utilizing not only qualitative but quantitative criteria where this was possible. This way, it can be used as a basis for the evaluation of new proposed schemes and as a reference for the properties a secure handoff scheme should posses.

The final conclusion of this review is that there are a number of solutions that can operate successfully in heterogeneous environments and in different contexts. The choice of the right scheme depends on the specific requirements that should be met and the comparison table can be of great help in that. However, the observation of the comparison table also reveals that

privacy is not protected effectively in any of the presented schemes; although “Optimistic access” scheme does not dictate the transmission of user identities, it does not describe any specific mechanism that can assure the privacy of end users.

Chapter 4 - Privacy preserving secure handoff optimization schemes

The advances in wireless communication technologies towards 4G networks and the wide use of mobile devices have enabled users to communicate with each other and receive a wide range of mobile wireless services through various types of access networks and systems everywhere, anytime. For example, with the rapid proliferation of IEEE 802.11 based networks, it is obvious that mobile users will want to take advantage of the high speeds and low cost that they offer. However, this does not mean that they will be willing to give up the broad coverage of the mobile networks. It is envisioned that in the near future mobile users will be able to use these two types of wireless networks in parallel. In order to have fast, secure handovers a number of methods have been proposed and the previous chapter provides a survey as well as a comparison of these methods. As discussed in [5], while these methods do succeed in minimizing the disruption caused by security related delays, it seems that they do not take into consideration the protection of the end users' privacy at all.

It is true that a lot of work has been done in privacy and more specifically in location privacy; however, there is no previous work in the literature preserving location privacy in methods offering fast secure handovers in all-IP based networks. In this chapter the Context Transfer solution is discussed; the privacy issues arising from the employment of the Context Transfer Protocol (CTP) [77] are highlighted and two schemes are proposed towards solving these problems. In the first one the MN is responsible for the transmission of its own context, while in the second the HD acts as a proxy between the previous and the new administrative domain. These two schemes are further extended based on the observation that the NAI [78] is a suitable type of identity for networks that span across multiple administration domains. Since this applies here temporary NAIs are used as context's identity in order to increase the level of user's privacy. The result of this work is that the decision for user's identity and location disclosure is no longer left to the good will and intensions of the visiting networks and the user is not forced to trust the foreign domains but only his home domain with which he has signed a contract.

4.1 Context Transfer Protocol

One of the most promising methods for seamless handover is the concept of context transfer. This is based on the work done by the SEAMOBY Working Group [79] which led to several RFCs, among them to RFC 4067 [77]. The latter describes the CTP. The idea behind context transfer is that when a MN handovers to a new access router (nAR), the uninterruptible continuation of the established services is not always possible, especially when the nAR is in a different administrative domain. In such a case, prior to the services re-establishment, the MN must authenticate to the new domain and re-authenticate to the services it already receives using an Authentication, Authorization and Accounting (AAA) protocol. To avoid excessive signaling and possible delays, the CTP is exercised as follows: the required information for each service can be stored in a Context Transfer Block (CTB) as illustrated in Figure 4-1. This information can be parameters for the quick re-establishment of services like multimedia or AAA transactions without the need to re-negotiate them. When the MN is receiving more than one service, the resulting CTBs can be bundled into a single Context Transfer (CT) packet and transferred to the

nAR as described hereunder. This way the nAR can handle the handover process more quickly and efficiently, allowing the MN to experience a seamless handover.

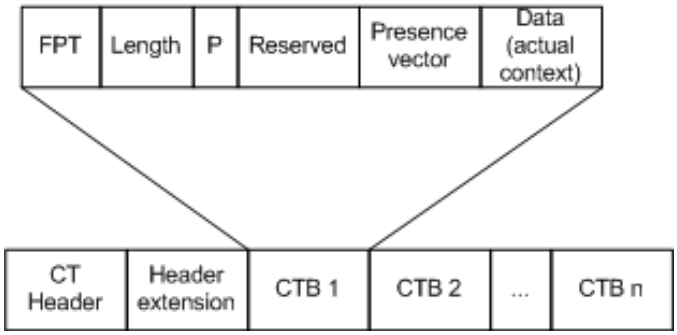


Figure 4-1: Context data blocks bundled into a context transfer packet

The standard way to achieve the desired functionality is to transfer the context between layer 3 entities at the edge of the network (ARs). This can be done in two ways: proactively or reactively. In the proactive scenario, the previous AR (pAR) sends the context to the nAR without the nAR asking for it. In the reactive scenario the nAR requests the context from the pAR. In any case, the handover decision is controlled either by the MN or the network (represented by the pAR, when the initiator of the handover is the previous visiting network, and the nAR, when the initiator is the new visiting network).

In Figure 4-2 an example of a context transfer procedure between layer 3 entities is illustrated. The pAR and nAR belong to two different administrative domains and the MN is moving from position P1 to P2, which are covered by access points AP1 and AP2 respectively, while in use of a demanding service, for example a multimedia session. The context transfer takes place between the two ARs and the only possible role the MN can play to the protocol is to initiate the transfer.

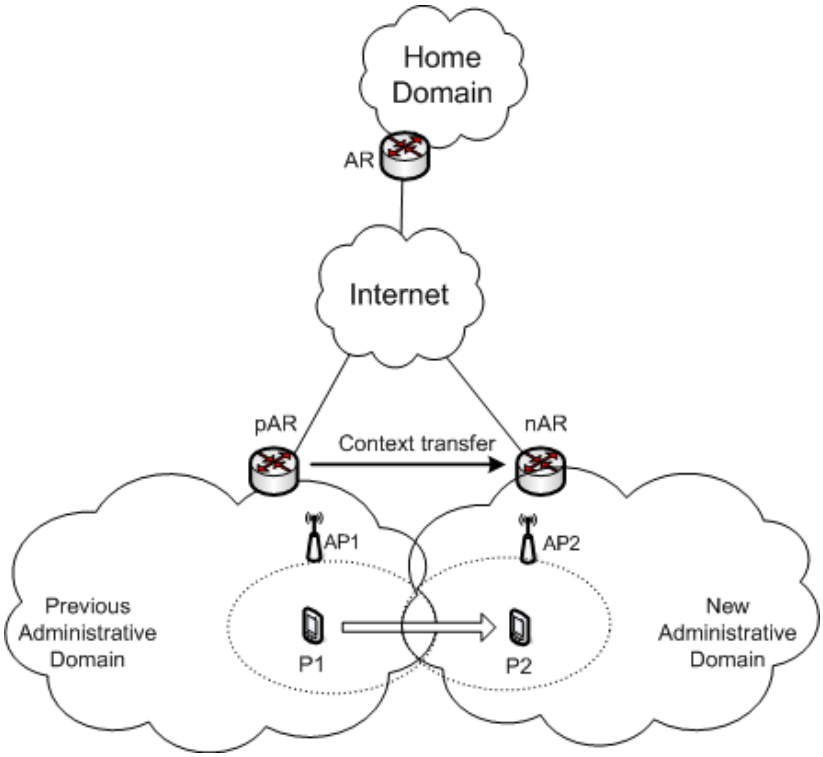


Figure 4-2: The standard way of Context Transfer between ARs

4.2 Network Access Identifier

When dealing with multi-domain models, there should be a way to distinguish not only the users but also the domain they originate from. This is very important for servers that are responsible for services like authentication and accounting in order to route the messages appropriately. In such cases, the NAI is used, which is similar to an e-mail address and is composed of two parts: the user identifier and the domain identifier separated by the "@" symbol, e.g. user_id@domain_id. When the domain_id is the local domain or no domain_id exists in the NAI, then the request is processed locally. When the domain_id refers to another domain (the home domain of the user), the request is routed to the correspondent domain; then the home domain can make an AAA decision based on the user_id.

4.3 The problem: Privacy issues in context transfer protocol

The way the CTP operates, as defined in RFC 4067, arises some privacy issues. These issues concern primarily the end user and more specifically his location and movement between different administrative domains. While a CTP assisted handover allows for seamless service delivery to mobile users, it seems that it comes with a cost in their location privacy.

The first observation has to do with the inner workings of the protocol itself. Every time a handover occurs, the pAR uses the CTP to send various context data blocks to the nAR. That is, for every handover the pAR and the nAR know where the user came from and where he is going. When these two ARs belong to the same administrative domain there are not many things that can be done to prevent the administrative domain from being aware of the movement of a single MN inside its own network. However, when the two ARs belong to different administrative domains there is no reason for the pAR to know which the nAR is and the

opposite. To sum up, with the use of the CTP for seamless handovers, every administrative domain is aware of the previous and the next administrative domain of the MN, without excluding itself. This means that every domain can track a part of the user's movement.

Continuing from the last conclusion, the user's movement can be completely tracked, given that some administrative domains collude. Note that this does not imply that all administrative domains in the path of the user movement are required to collude for such an attack, but every second domain in that path.

Another aspect of the location privacy problem when the CTP is in place is the type of the identifier used by the user/MN during the protocol negotiation to authenticate to the new administrative domain. The utilization of a static identifier like a globally used username of the user simplifies the work of a malicious passive observer. An obvious choice for all-IP networks that belong to different administrative domains is the use of a NAI. However, in the case that the administrative domains collude, they can track the whole movement of the user only by the observation of the use of this static NAI. Furthermore, even when administrative domains do not collude there can be a location privacy breach, since every single domain can recognize an old user that returns to it. It is thus, more than obvious, that systems' logistic files can be anytime processed to disclose information about the whole history of movements of a specific user.

4.4 Scheme I

The first scheme [6][7], protects the location privacy of users roaming between different administrative domains utilizing the CTP. Our solution is twofold and it is proposed that:

- the context should be submitted by the MN, and
- there should be a frequent NAI change.

The basic idea behind this scheme is that the user's sensitive information should only be known to the user himself and his home domain and no-one else, including the visiting domains. This is very important since the user has agreed and signed only one subscription contract; with his home domain. What this solution tries to succeed is to transfer the responsibility and supervision for user's privacy to his home domain; all the other domains only know and trust the home domain of every user that visits them.

4.4.1 Mobile Node Submitted Context

As it is stated in RFC 4067, the context is transferred between layer-3 entities from the old network domain to the new network domain. This way, a part of the MN user's route can be tracked. As already stated this is the case of a single domain tracking the movement of the user; if domains collude, then the full movement of the user can be tracked simply by using the information revealed by the CTP.

One possible solution to avoid such problems is to have the MN submitting its own context to the network it is moving to. The complete abstract protocol steps are as follows:

1. The MN establishes a secure session with the AR of the new domain. This secure session must have the following properties: a) it must be encrypted and b) the AR must be authenticated to the MN.
2. The MN sends the context over the previously established protected channel.
3. The AR authenticates the MN and re-establishes the services based on the context. It is also assumed that the current domain has established some kind of trust relationships beforehand with the home domain. This way the authentication is processed locally based

on an authentication token located in the context, which is digitally signed by the home domain.

The above procedure is the equivalent of a PEAP [80] or an EAP-TTLS [81] authentication and key establishment method using the context as user authentication means. The first phase of the PEAP or EAP-TTLS method is followed as is, e.g. a secure session is established with the use of the digital certificate of the AR. In the second stage the authentication of the user is taking place with the utilization of the credentials contained in the context. The key establishment phase could also be benefited by the context transfer since the context can contain security parameters i.e. cryptographic keys, supported suites, tokens, etc.

The proposed method can be used in either a reactive or proactive scenario. In cases where a high QoS must be preserved, the aforementioned procedure could be executed proactively, that is before the MN actually moves to the new administrative domain. This situation is comparable to the pre-authentication procedure exercised in IEEE 802.11 or 802.16 networks.

An example of a context transmitted by the MN is shown in Figure 4-3. The scenario is the same as in Section 4.1. When the MN moves towards P2 the handover procedure starts. The MN establishes a secure channel with the nAR and through this channel transfers the context. As it can be easily noticed, the ARs do not play any role in the context transfer procedure and there is no communication between them. Also, they are not aware of each other in any way.

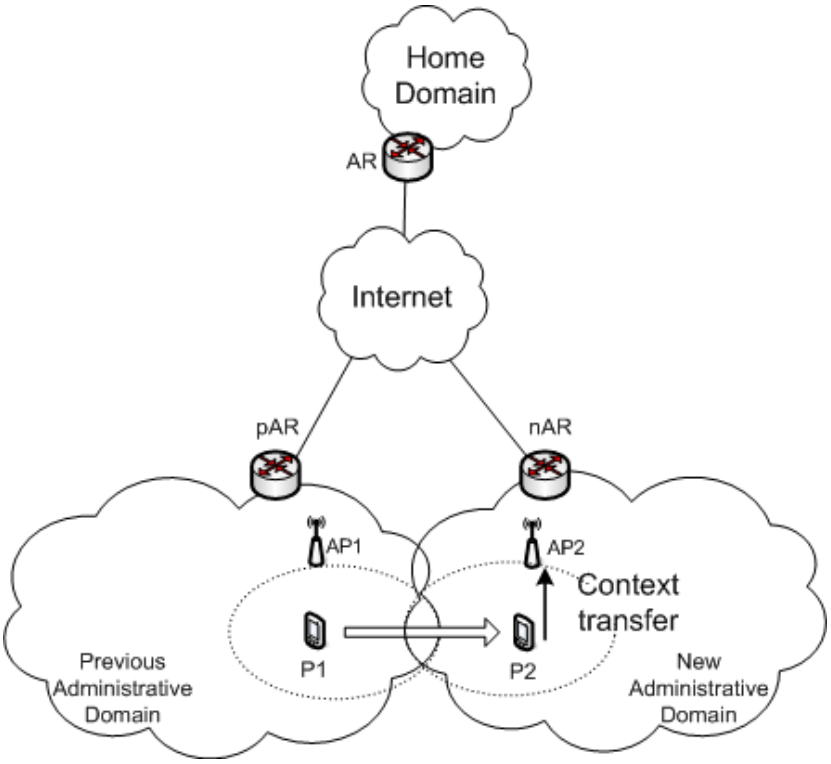


Figure 4-3: MN submitted context

4.4.2 Frequent NAI Change

As it has already been analyzed above, one way to identify the users is the use of NAI. Of course, the NAI can also be utilized in conjunction with the CTP. When the NAI concept is employed in the proposed way (MN submits the context) then the current domain or some colluding domains

still can track the location of the user simply by observing the transmission of NAIs. More specifically, the current domain can always be aware when a single user was present in its network or when a user returns to it; when the domains collude, things get worse since they can observe the exact route of a single user.

The solution is based on the use of temporary NAIs and the frequent change of them:

- The home domain is the only one that has the correspondence between the true identity of the user and the NAI assigned to him.
- When a context is created for the user, it contains a temporary NAI. This temporary NAI uses as user_id a random unused string, which the home domain connects with the true identity of the user, and as domain_id the assigned domain_id. Each temporary user_id is used once for every single domain by one user at a time. When the user handovers to another domain (either new or previously visited) he must use a different user_id. The reuse of a temporary user_id by another user is not forbidden since the home domain is also aware of the date and time each user is using it. Therefore, the only sensitive information about the user that is revealed to foreign domains is the home domain of the user.
- After the completion of the handover of the MN to a new domain, the MN is using a secure channel (like a TTLS session) to contact its home domain and obtain a new temporary NAI. This way, when the user returns to a previous visited domain, the domain cannot recognize him.

Even if the correspondence between the true identity of the user and his NAI or any temporary NAI is revealed by accident or other reason, the user's past routes cannot be revealed without the help of his home domain.

The obvious drawback of this method is the increase in the signaling between the domains. However, this is done after the completion of the handover and therefore has no real effect in the QoS perceived by the user during the handover.

In Figure 4-4 a message sequence diagram of the first proposed scheme is presented. The MN has an existing session with the pAR; when it wants to handover to the nAR it first establishes (proactively or reactively) a secure session with it. Then, through this secure session, it transfers the context that will allow the MN to authenticate, establish session keys and re-establish the services it already uses. When the handover procedure is finished and the new session has been established, the MN should contact its home domain in order to obtain some new credentials (for example a new temporary NAI) that will be used in its next handover.

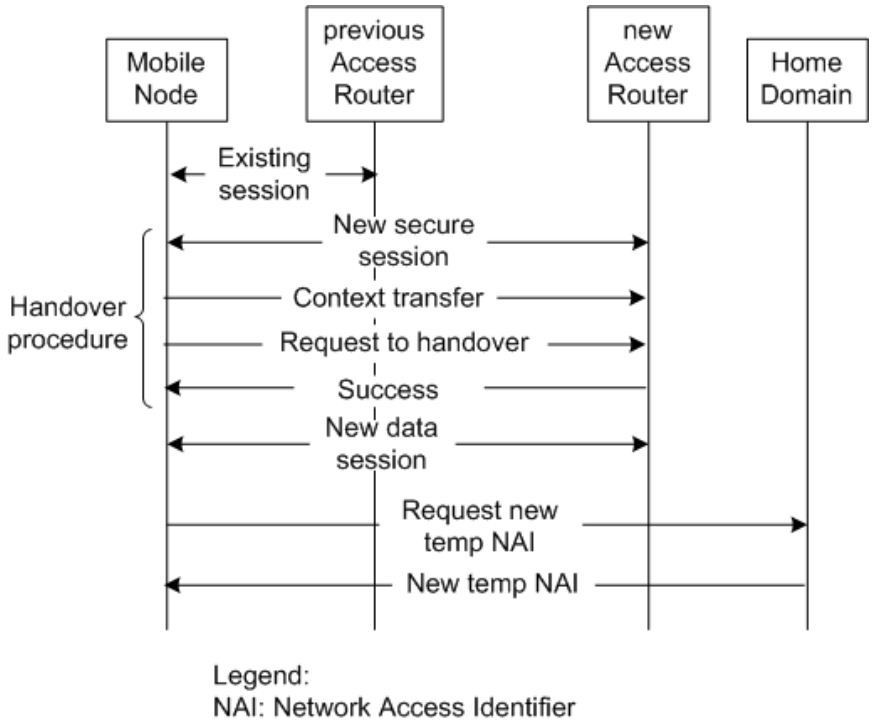


Figure 4-4: Message sequence of scheme I

4.5 Scheme II

The second proposed scheme [7][8] protects the location privacy of users who roam between different administrative domains using the CTP for more demanding services than the abovementioned ones. Again, this solution has two main points:

- the context is transferred through the Home Domain (HD), and
- there is a frequent NAI change as well.

The basic idea shown in the first scheme still holds here; that is the user's sensitive information should only be known to the user himself and his home domain and no-one else, including the visiting domains.

In this scheme the HD acts as a proxy between the pAR and the nAR executing the context transfer prior to the MN's movement to the new domain in order to protect the privacy of the MN's user. Here the frequent NAI change is tightly bundled with the context submission procedure. The complete abstract protocol steps are as follows:

1. The MN realizes that it is about to handover to a new AR that belongs to a different administrative domain from the current one. Thereby, it establishes a secure session with its HD and requests from it to execute a context transfer to the new administrative domain on behalf of the MN. This request contains the current temporary NAI of the MN.
2. The HD requests the context of the MN from the pAR using the MN's current temporary NAI.
3. The HD changes the temporary NAI in the context and forwards the context to the nAR.
4. The HD uses the established secure session with the MN and forwards the new temporary NAI to it.
5. The MN handovers to the nAR using its new temporary NAI.
6. The nAR authenticates the MN and re-establishes other services based on the context. It is also assumed that the current domain has established some kind of trust relationships

beforehand with the HD. This way the authentication is processed locally based on an authentication token located in the context, which is digitally signed by the HD.

The proposed method is clearly a case of a proactive scenario where the context transfer takes place before the MN actually handovers to the new domain.

The procedure of creating and using temporary NAIs is similar to that described in the first scheme. It must be noted here that as long as the MN is located at the area covered by the pAR it uses its current temporary NAI and only when it moves to the nAR it uses its newly assigned temporary NAI.

An example of a context transmitted by the MN is shown in Figure 4-5. The scenario is the same as in Section 4.1. When the MN moves towards P2 the handover procedure starts. The MN establishes a secure channel with the HD and requests from it to transfer the MN's context from the pAR to the nAR. As it is illustrated in Figure 4-5, the HD first retrieves the context from the pAR (step 1), it makes the necessary modifications to it and then forwards it to the nAR (step 2). When the context transfer is completed, the HD sends the MN its new temporary NAI. The protocol is finished when the MN handovers to the nAR. As in the first scheme, the ARs do not play any role in the context transfer procedure and there is no communication between them; therefore, they are not aware of each other in any way.

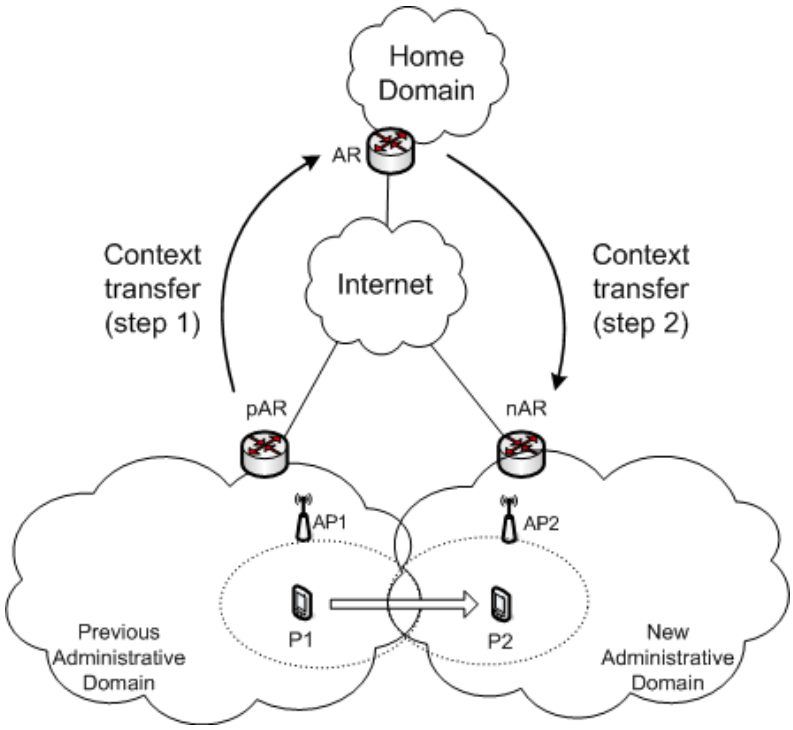


Figure 4-5: HD submitted context

This scheme makes a tradeoff between the privacy of the user and the increased signaling among the administrative domains. Nevertheless, such a cost would be acceptable in cases where the privacy of the user is a priority.

Figure 4-6 illustrates a message sequence diagram of the second scheme. At first the MN has an existing session with the pAR. When the MN decides to handover to the nAR it first establishes a secure session with its HD. Using this secure session, the MN requests from the HD to perform the context transfer acting as a proxy. The HD retrieves the context from the pAR

(step 1), replaces the current temporary NAI with the new one and forwards the new context to the nAR (step 2). Through the previously established secure session the HD also forwards the new temporary NAI to the MN. After these steps the MN can handover to the new domain using the current (active) context.

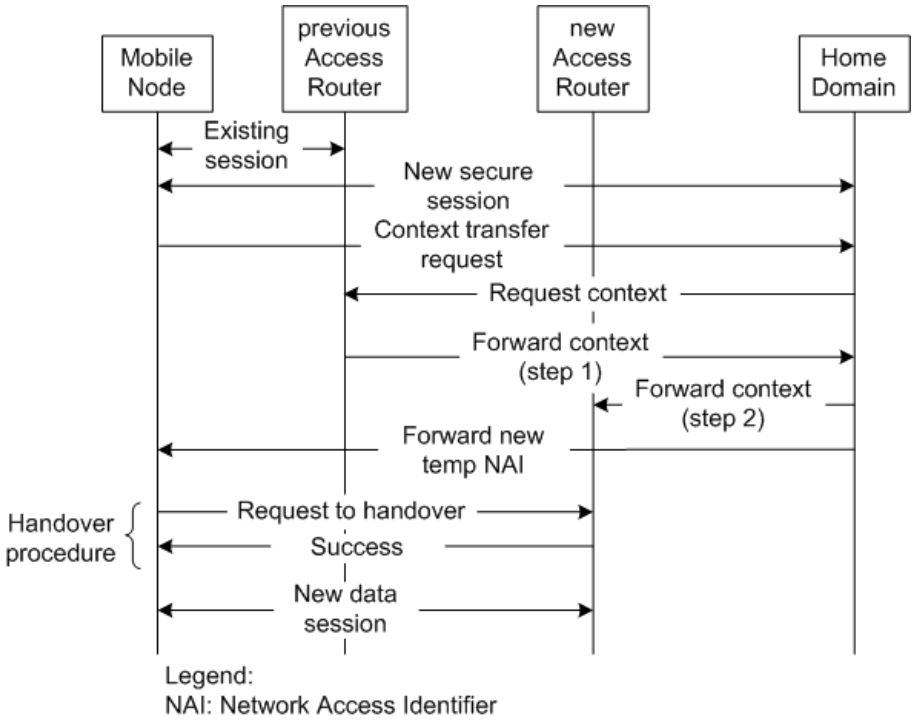


Figure 4-6: Message sequence of scheme II

4.6 Discussion

This section provides some points concerning the deployment of the aforementioned protocols. From the trust requirements point of view, the proposed solutions have some prerequisites that are analogous to those of CTP. More specifically, CTP requires that trust relationships exist among the ARs and between the MN and each of the ARs (pAR and nAR). Here, each AR should have trust relationships with the home domain of the roaming MN; since the MN also has trust relationships with its home domain, new trust relationships between the MN and each AR can be established on-the-fly.

An important factor concerning the wide deployment of a protocol is the number of changes required in the already installed infrastructure. Taken into account the situation as it is today, the two proposed schemes require a reasonable number of such changes which are comparable to those required for the deployment of the CTP. More specifically, in CTP the ARs should be able to transfer the context among them and interpret the contents of the context; the MN should also implement the CTP in order to be able to request the transfer of the context. In the proposed schemes the ARs should only be able to interpret the contents of the context. Also, in the first scheme the MN should be able to handle the context which it possesses according to the proposed protocol, while in the second scheme the HD should be able to play the role of a proxy between the previous and the new domain.

Another point of consideration that applies only to the first scheme is the protection of the context itself. Since in the proposed protocol the context is carried by the MN, actions must be taken so that the context cannot be altered by the user unnoticed. This implies that there should be a kind of digital signature in place ensuring the integrity of the transmitted context. The encryption of the context while stored in the MN is not a strict requirement since the information contained in it is already known to the user. However, having in mind that the MN is a portable device and thus it is easy to get lost or stolen, some care to prevent tampering, unauthorized use, or fraud could be taken. The second scheme does not suffer from such a threat since the HD communicates with other domains through secure channels (e.g. usually IPSec or TLS).

A brief comparison of the two proposed schemes would lead to the conclusion that each one is suitable for different types of applications. The first scheme poses a small amount of load to the HD while at the same time takes longer to handover to a new administrative domain. This makes it more suitable to applications with less strict demands or applications that can tolerate longer delays during the handover procedure. The second scheme requires the exchange of more messages but it is expected to have better performance during the handover. Therefore the second scheme will be more useful towards seamless handovers for demanding applications like multimedia delivery.

One final remark about the context is its expiration. The time interval of expiration should be neither too large, containing expired information, nor too small, causing excessive signaling among the administrative domains. What is obvious for both schemes is that when the MN moves to a new domain the context is renewed since a new temporary NAI is requested. In any case, the expiration interval can be set by the network administrators and the current point of attachment (some AR) of the MN can warn it that its context has expired or is about to expire.

4.7 Summary

In this chapter the privacy issues when using the CTP which is currently employed by the state of the art methods for seamless secure handovers between different administrative domains have been presented. In addition to this, two novel schemes that preserve user's location privacy were proposed. The standard way the protocol behaves arises some privacy issues and the two proposed alternative protocols alleviate these problems. Moreover, as it has been discussed, the proposed use of the context in conjunction with a NAI can further enhance user's privacy.

Chapter 5 - Survey of SIP privacy solutions

Multimedia is an application class with great importance in today's networks no matter whether these are wired or wireless. In fact, it is important that multimedia delivery is based on interoperable protocols so that converged (and possibly heterogeneous) networks can offer uninterrupted services. It is expected that the next generation of wireless networks, namely 4G, will be based on IP, realizing an all-IP architecture. It is obvious at this point that such IP based networks will be fully compliant with wired networks and the Internet with no need for gateways or other translation means. In such an environment the multimedia deliverance will be possible even when users move or change between networks with different access layer technologies.

One of the most important protocols supporting multimedia services is Session Initiation Protocol (SIP) [49]. SIP is an application layer control signaling protocol responsible for the creation, modification and termination of multimedia sessions. One of the facts that show the significance of SIP is that 3GPP consortium [82] chose it to be the multimedia management protocol of 3G networks multimedia subsystem (IP Multimedia Subsystem - IMS). Since SIP is an application layer protocol, it can transparently operate over any type of network; furthermore, it also has the ability to support application layer handovers when a lower layer handover occurs [73].

SIP has been a protocol which has received extensive attention and part of the research has shown that it suffers from security issues [83] some of which have already been solved [83][84]. Here the focus is on privacy and more specifically on the protection of user IDs that normally are publicly available to anyone who has access on the underlying network. While there are some solutions for protecting the privacy of end users, these are not adequate in certain environments.

This chapter starts by analyzing the ID privacy issues of SIP in detail. Next a number of ID privacy levels are defined which will help in the comparison of different methods that can offer privacy services to SIP. These methods are analyzed and compared based on certain criteria which are also presented and analyzed in the remainder of the chapter. Finally, a comparison table is provided summarizing the comparison, followed by a discussion on the comparison findings.

5.1 Problem statement

In this section a generalized SIP architecture which spans across many different administrative domains and the identity privacy issues that arise from it are presented. This analysis is so general that applies to either wired or wireless scenarios or a mix of them.

In Figure 5-1, O'Brien uses a fixed terminal residing in *miniluv* domain and Smith uses a mobile terminal. Smith's Home Domain is *minitrue* but at the moment he roams to a different domain, *minipax*, and wants to contact O'Brien. If Smith's terminal is not aware of its Home SIP Proxy's IP address then a possibility is that other Proxies (like Local outbound Proxy) intervene between Smith and *minitrue.org* as well as between *minitrue.org* and *miniluv.org*. Most of the times these SIP Proxies are unknown to Smith and cannot be considered trusted; moreover, Smith has no means to control which Proxies his messages will travel through. Such messages, which are known as SIP messages, contain among other information SIP URIs. A SIP URI is a URI similar to

an e-mail address which contains a user ID and a domain name separated by the "@" symbol, for example *smith@minitrue.org,* "smith" being the user ID and "minitrue.org" the domain name. In our example, *minipax* is not Smith's Home Domain, but if Smith is to use its services he must have some kind of agreement with it and thus trust it to some extent. However, the credentials used in this domain can be different than those used in his Home Domain which is *minitrue.* A probable requirement here could be that Smith wishes that each set of credentials is available only to the corresponding domain and not to anyone else. Considering ID privacy, SIP cannot protect users' IDs since they are transmitted in the clear while other methods that have been proposed so far and are presented in subsequent sections prove to be inadequate in certain occasions. What is really needed is a solution that selectively makes Smith's ID known only to entitled trusted entities, while hiding it from untrusted ones.

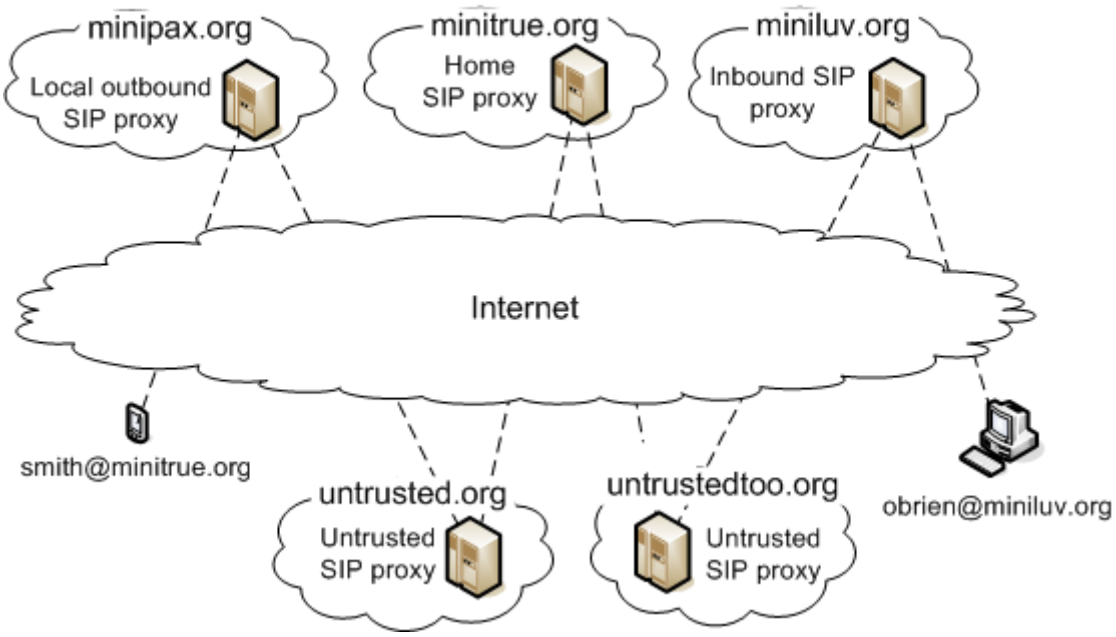


Figure 5-1: Multidomain SIP architecture

Considering the previous example the information that is revealed to third parties is that a user from *minitrue.org* domain has a conversation with O'Brien from *miniluv.org*. In a second example, a more effective, in terms of privacy, scheme could also protect O'Brien's ID so that the only information available to third parties would be that a user from *minitrue.org* has some sort of communication with a user from *miniluv.org*.

In order to demonstrate the SIP ID privacy issue more clearly we examine the headers of a SIP message used for placing a call, e.g. an INVITE sent from Smith to O'Brien (other SIP messages have similar headers):

```
INVITE sip:obrien@miniluv.org SIP/2.0
Via: SIP/2.0/UDP 195.251.161.144:5060; branch=z9hG4bK74b43
Max-Forwards: 70
From: Smith <sip:smith@minitrue.org>; tag=9fxced76sl
To: O'Brien <sip:obrien@miniluv.org>
Call-ID: 3848276298220188511@minitrue.org
CSeq: 1 INVITE
Contact: <sip:smith@minitrue.org>
Content-Type: application/sdp
Content-Length: 151
```

As it can easily be seen in the above message, particular headers reveal private information about the two communicating parties. The headers that reveal information about the caller and the callee are:

- The first line of the message contains a Request-URI which reveals the callee's ID
- <Via> header reveals the caller's host IP address,
- <From> reveals the caller's SIP URI (which is composed from the user's ID followed by his Home Domain name)
- <To> reveals the callee's SIP URI
- <Call-ID> reveals the domain where the caller belongs (in this case *minitrue.org*) and
- <Contact> reveals where the caller can be contacted so that the two parties can establish a peer-to-peer connection (the value of this field can be either a SIP URI or simply an IP address).

There are a number of malicious acts associated with the lack of user ID privacy. The first and more obvious one is that everybody can have access to information regarding who is communicating with whom. If this information is systematically gathered then a certain user can be profiled, based on VoIP calls and other multimedia usage. When SIP URIs are made available then a possible attack is also Spam over IP Telephony or SPIT [90] which is similar to e-mail spam. Another consideration is that the movement of a specific user can be tracked by observing the transmitted IDs. This can happen when a mobile user handovers between different networks and transmits his ID in order to transfer the existing session to the new network. This can also be the case when session mobility is supported and a certain user continues using a session but changes between different devices, either mobile or not.

5.2 Privacy levels

Before describing related work on SIP ID privacy a definition of different levels of ID privacy is provided [9]. The distinction is based on who has access to the real ID of either the caller or the callee or both. These privacy levels are defined based on a number of criteria which are shown below in order of importance:

1. The Domains and the callee are considered more trustworthy than other third parties.
2. All Domains engaged are considered more trustworthy than the callee.
3. The Home Domain is considered more trustworthy than other Domains.
4. The ID of some user must be available to as less entities (other than himself) as possible.

The resulting privacy levels are the following starting from no privacy at all:

Level 1: The ID of some user is available to everyone.

Level 2: The ID of some user is available to himself, the user at the other end of the call and all the Proxies of all domains in the call path.

Level 3: The ID of some user is available to himself, the user at the other end of the call and their Home Domains.

Level 4: The ID of some user is available to himself, his Home Domain and the user at the other end of the call.

Level 5: The ID of some user is available to himself and the user at the other end of the call.

Level 6: The ID of some user is only available to himself and his Home Domain.

Level 7: Only the owner of the ID has access to it.

5.3 Proposed solutions

The issue of privacy protection is not completely ignored in SIP and this is proved by the fact that [49] includes certain mechanisms that can assist a user in protecting his privacy. These mechanisms can be separated to cryptography based ones which are S/MIME [85], SIPS URI/TLS and IPsec, and the non cryptographic solution of "Anonymous" URI. A different approach is the extension of the basic SIP protocol which led to the solution presented in [86] which will be referred here as "Privacy Mechanism for SIP". This is in fact a general purpose privacy mechanism which has also been used in [87] adapted to the specific needs arose there. In the following these solutions are presented in more detail, while more focus is given on how each solution can be utilized to protect end users' IDs.

5.3.1 S/MIME

SIP messages consist of two parts: the header and the body. The body part is nothing more than a MIME body so an obvious solution to protect it is by using the standard way which is S/MIME. Although this may seem out of scope, given that here the focus is on protecting specific SIP headers, S/MIME in the context of SIP can be used to cryptographically protect SIP headers.

S/MIME protects the confidentiality of SIP headers and bodies using digital certificates. In order to protect the privacy of end users S/MIME can encapsulate SIP messages into MIME bodies and encrypt them properly. The encapsulated message can contain the real ID of the caller while the "outer" message contains a <From> header of the form: "sip:anonymous@anonymizer.invalid". When the called party receives the message, he decrypts the body to find the ID of the caller. What must be noted here is that the ID of the callee cannot be anonymized using the same mechanism since the intermediate SIP Proxies do not have access to the plain MIME body and an anonymous <To> field would made them unable to route the message to the intended recipient.

Although S/MIME can seem as a promising solution there are some obvious weaknesses. First of all the receiver of messages must somehow be aware of the identity of the sender a priori, in order to find the appropriate certificate to decrypt the message body. Another privacy weakness is that the receiver knows the ID of the sender, while the receiver's ID is not protected from third parties. Finally, there is no way to hide the IP addresses of the communicating parties, something that also holds for all subsequent methods.

5.3.2 SIPS URI/TLS

It is possible for end users to request that their messages along the whole path to their destination are transported with the use of TLS protocol in order to ensure their privacy protection. This is accomplished with the use of “sips:” instead of “sip:” in a typical SIP URI. While a SIP message having a <To> header of the form: “sip:obrien@miniluv.org” will be visible by anyone, its security enhanced equivalent “sips:obrien@miniluv.org” will request all intermediaries to use TLS in a hop-by-hop manner until the specified domain is reached. After that, the message is handled according to the local security and routing policy.

This approach also presents some worth noting issues. If SIPS URI scheme is selected, then the use of TLS implies the use of TCP as a transport means, while the preferred transport protocol for SIP is UDP. While there is also the solution of DTLS [88], which is the equivalent of TLS using UDP as transport mechanism, it is a scheme that was proposed later than SIP so it is not included in [49]. The main drawback of SIPS URI however is that there is no guaranteed end-to-end protection. While TLS can be used in each hop-by-hop connection, it is not possible to dictate or even be informed somehow that it will be used in every intermediate connection. This can result in two possible attacks; the first one is a downgrade attack, where some intermediate proxy just does not use TLS or replaces “sips:” scheme with “sip:”. In the second attack the caller uses plain “sip:” scheme and some intermediate proxy modifies it to a SIPS URI so that the recipient of the message believes that their communication is TLS protected.

5.3.3 IPsec

For the purposes of SIP, IPsec can be used in a hop-by-hop fashion protecting the data transmitted between two hosts at the network level. The main difference between IPsec and SIPS URI/TLS in the context of SIP is the transparency offered by IPsec to SIP User Agents (UAs). As it is stated in [49], IPsec will be more suitable in cases where the communicating hosts have already established a trust relationship with one another as opposed to SIPS URI scheme.

What holds for end-to-end protection in SIPS URI also applies here; it is not guaranteed. This is because there is neither an available mechanism to impose the use of IPsec in all intermediate hosts nor a way for communicating parties to be aware of whether this actually happened or not.

5.3.4 Anonymous URI

Another approach proposed in [49] for the protection of caller’s ID is the use of an Anonymous URI in the <From> field. This URI has meaningless values and it is of the form: “sip:anonymous@anonymous.invalid”. It must be noted here that this Anonymous URI is inserted into the <From> field by the UA itself which means that the SIP Proxy can never have access to the real URI.

The drawback of this solution is that it cannot support UA authentication since no ID is transmitted. A possible workaround could be a UA device shared among many end users. This device will own a specific pair of username and password for authentication purposes which will be the same for all users; however such a solution creates other important security issues like repudiation of actions.

5.3.5 Privacy mechanism for SIP

The scheme described in [86] is an extension of the basic SIP protocol and defines two ways for the protection of end user's privacy: user and network provided privacy. The end user can choose between these two or utilize both at the same time. When the UA chooses user provided privacy, it populates certain SIP headers with meaningless values, for example <From> field with an Anonymous URI. When network provided privacy is selected an intermediate node is assigned a new logical role for offering anonymization services to UAs while at the same time is responsible for directing messages from and to the anonymous user as a normal SIP Proxy. In order to enable UAs to request such services a new SIP header is introduced, namely "Privacy-hdr", which takes the following values: header, session, user, none and critical. With the use of one or more of these values the users can ask the network to: obscure headers that cannot be altered without the assistance of an intermediate, for example <Via> and <Contact>, provide anonymization services for the session initiated by the message, cancel any default privacy preferences or mark the criticality of the request for privacy. The recommended way for the UA to communicate with the privacy service provider is by using network or transport layer security protocols.

This mechanism has also been adapted to fit certain requirements in [87]. In this version the user sends a SIP message through a trusted set of Proxies revealing his true ID. When the message is about to leave this trusted domain, the last Proxy withholds the true ID of the user. Similarly to the initial scheme the last Proxy must keep state information in order to route back the responses.

A shortcoming of this method is that the node offering privacy services must keep a significant amount of state information in order to complete the proper routing of the messages. Another issue is that this node can potentially be a single point of failure if replication is not used. When user provided privacy alone is chosen then what applies for the "Anonymous URI" solution also applies here. The authors of this method have chosen not to consider any privacy considerations arose by the use of authentication mechanisms like Digest authentication. However, a username used in such a method could possibly reveal private information about the end user.

5.4 Criteria of comparison

In the following sections the solutions presented above will be compared to each other. Here the criteria used for this comparison are analyzed and in later sections the response of each scheme to these criteria will be presented. Finally a table of comparison will be provided summarizing all the information from the analysis that follows.

5.4.1 Cryptography

By this criterion it is examined the use of cryptography for the purposes of each solution. Some schemes are based on cryptography to keep personal information private while others use other means. A direct implication is that schemes that do not use any kind of cryptography will probably be faster and have less administrative requirements, mainly due to lack of key management.

5.4.2 Authentication

Here it is examined whether each solution can support authentication without revealing any private data to non intended parties. More specifically it is checked if the standard

authentication mechanism in SIP, which is Digest authentication, can be utilized without making the real ID of the end user available to third parties.

5.4.3 Public Key Infrastructure (PKI)

With this criterion the proposed solutions are separated based on their PKI requirements. As it will be shown some of them require a full PKI, others a limited PKI while others no PKI at all.

5.4.4 Anonymity vs. pseudonymity

This criterion indicates what kind of ID is used in the place of the real user ID. This can be a static string like “anonymous@anonymous.invalid” or a completely random string in which case we have a completely anonymous scheme. On the other hand, when the replacement ID is produced in some way from the real ID we have a scheme based on pseudonymity. The most notable difference here is that the person receiving a call from a UA using a pseudonym can always return the call using this pseudonym something that is not possible with anonymous schemes. Also, in a poor designed scheme that uses pseudonyms, a user can be tracked down when, for example, is using the same pseudonym repeatedly, even if the correspondence between the real ID and the user ID is kept secret.

5.4.5 Inter-Domain agreements

One of the most common preconditions in schemes offering security services in multidomain environments is that different administrative domains must have pre-existing trust agreements between them. This limits the number of users’ choices only to networks that belong to co-operative domains. In the following comparison it is examined whether each solution needs such pre-existing agreements between domains in order to offer ID privacy to end users.

5.4.6 Multidomain support

Here it is examined whether a solution can support its privacy features when operating in an environment composed of different administrative domains. These domains can belong to different operators and/or service providers. The difference between “Multidomain support” and “Inter-Domain agreements” is that a scheme can support multidomain environments without requiring pre-arranged inter-domain agreements; when a solution requires inter-domain agreements, obviously supports multidomain environments. A scheme can either fully support multidomain environments or not.

5.4.7 Untrusted proxies

When a UA initiates a multimedia session its request can travel through untrusted SIP Proxies until it reaches its Home Proxy which is considered trusted. The main purpose here is to check whether each solution can guarantee UA’s privacy protection even when SIP messages traverse through untrusted proxies.

5.4.8 Domain name protection

Since the focus is on schemes that preserve the privacy of end users the main concern here is on protecting as much private information as possible. With this criterion it is examined if each

method protects among other things the Home Domain's name of each or both the communicating UAs. While domain name is private information its protection is not considered of ultimate importance since its disclosure does not directly reveal the ID of the end user.

5.4.9 IP address protection

What holds for domain names also holds for protecting each end user's IP address. It is private information which is not considered crucial and cannot directly lead to the real ID of the user. However, under some circumstances, it can reveal the current position of the user and in extreme situations, combined with other personal information, even his real ID.

5.4.10 Privacy level

This criterion shows in which of the privacy levels listed in Section 5.2 each method is classified. The classification is based on "how much" privacy each method offers; thus the higher the level, the higher the privacy offered by each method. In order to be more practical, numbers 1 to 7 will be used to indicate which of these levels is reached.

5.4.11 Hop-by-hop vs. end-to-end privacy

As it has already been analyzed, the establishment of a SIP session typically includes a number of intermediate nodes. With this criterion it is checked whether each method can guarantee users' ID privacy in a hop-by-hop or an end-to-end manner; obviously the second is the preferred one since only this way we can be sure that privacy was not compromised along the session path.

5.4.12 Stateful vs. stateless mode

Here it is examined whether each scheme requires SIP Proxies to be stateful or stateless in order to be fully operational. Stateful proxies keep state information for each ongoing session something that speeds up or make possible the offer of specific services, however leads to a need for more storage resources. Stateless proxies on the other hand do not store any information regarding sessions so they have less storage needs and have higher response delays. While each mode has its own advantages over the other, in some occasions some services may be able to run only in one of the two.

5.4.13 Deployment

This criterion indicates the easiness of deployment of a scheme. Here a qualitative measurement will be used based on empirical observation. Three degrees of ease of deployment are defined: easy, medium and difficult.

5.5 Comparison

Here the actual comparison takes place based on the thirteen aforementioned criteria. Table 5-1 shows a summary of the findings of the comparison. In the following sections an extended commenting on each scheme is provided based on the previously defined criteria.

Criteria \ Schemes		S/MIME	SIPS URI/TLS	IPsec	Anonymous URI	Privacy mechanism
Cryptography		√	√	√	×	√
Authentication		×	√	√	×	×
PKI		full	full	×	×	limited
Anonymity vs. pseudonymity		anonymity	anonymity	anonymity	anonymity	anonymity
Inter-Domain agreements		×	√	√	×	√
Multidomain support		√	√	√	√	√
Untrusted proxies		×	×	×	×	×
Domain name protection		×	×	×	√	×
IP address protection		×	×	×	×	×
Privacy level	Caller	5	2	2	7	6
	Callee	1	2	2	1	1
Hop-by-hop vs. end-to-end privacy		end-to-end	hop-by-hop	hop-by-hop	end-to-end	end-to-end
Stateful vs. stateless		both	stateful	both	both	stateful
Deployment		difficult	difficult	difficult	easy	medium

√: supported/required

×: not supported/not required

Table 5-1: Privacy schemes comparison

5.5.1 S/MIME

Cryptography: S/MIME cryptographically protects various SIP headers using public key cryptography and digital certificates of end users.

Authentication: In [85] it is mentioned that encrypting <Authorization> and <WWW-Authenticate> header fields is not considered useful and any encrypted form of these fields will

be ignored. This means ID privacy during authentication is not supported and anyone can have access to all usernames of end users when they authenticate.

PKI: Since S/MIME uses public key cryptography it is straightforward that a sort of PKI is required. In this occasion a full PKI is needed where a digital certificate must be issued for every end user.

Anonymity vs. pseudonymity: In this solution a meaningless value is used in the "outer" <From> field while the real ID is placed into the encrypted MIME body. While the real ID exists in every such message it is encrypted together with other values thus it cannot considered as pseudonym; naturally this solution is based on anonymity.

Inter-Domain agreements: This scheme does not need any pre-existing agreements between administrative domains. Each user must have some kind of trust agreement with the party he is communicating with.

Multidomain support: S/MIME supports multidomain environments since SIP Proxies do not intervene in any way to the part of the message that preserves end user's privacy.

Untrusted proxies: This solution protects user's ID even when the relevant SIP messages travel through untrusted proxies. However, as already mentioned above, it cannot protect the username used for Digest authentication thus S/MIME is considered as a method that is not supporting privacy through untrusted proxies.

Domain name protection: The Home Domain name of the caller is not explicitly revealed, however an eavesdropper can discover which domains communicate with each other. On the other hand the Home Domain name of the callee is not protected.

IP address protection: The IP addresses of the communicating parties are not protected.

Privacy level: This scheme reaches Level 5 concerning caller's ID since caller's real ID is available only to the caller and the callee. Regarding callee's ID S/MIME offers no protection so it reaches Level 1.

Hop-by-hop vs. end-to-end privacy: The privacy protection of this solution is offered in an end-to-end fashion.

Stateful vs. stateless mode: SIP Proxies do not play any active role in privacy protection in this scheme so both modes are supported.

Deployment: The utilization of this solution mandates the deployment of a full PKI; as every typical PKI this includes a number of administrative actions like issuing digital certificates to all end users and revoking them when this is necessary. Another issue with S/MIME is that the callee must know a priori which the caller is in order to be able to choose and acquire the right public key certificate. For those reasons this scheme is considered to have difficult deployment.

5.5.2 SIPS URI/TLS

Cryptography: SIPS/URI utilizes TLS to protect TCP sessions between SIP network elements; obviously cryptography is part of this solution.

Authentication: Digest authentication is supported as is by this solution and the usernames are protected as well.

PKI: A full PKI is needed since TLS is used. According to this scheme digital certificates for all communicating users and intermediate SIP servers must be issued. Since a PKI is a requirement certificate acquisition, management and revocation is also an issue here.

Anonymity vs. pseudonymity: SIP messages are transmitted through secure channels therefore no user ID is revealed; this means that this solution retains user's anonymity.

Inter-Domain agreements: This scheme requires pre-existing agreements between administrative domains so that SIP Proxies belonging to different domains can establish a secure channel with the use of TLS. These agreements can be indirect based on digital certificates i.e. cross-certifications, and an existing PKI. It must be noted here that it is not obligatory for communicating users to have explicit trust agreements between them.

Multidomain support: SIPS/URI supports multidomain environments which have some sort of trust agreements between them, e.g. have been cross-certified beforehand, as already stated above.

Untrusted proxies: This solution should not be used when untrusted SIP Proxies exist in the communication path. If this is the case then it is possible that these untrusted Proxies will not use TLS so no protection is offered to the communicating parties at all.

Domain name protection: When SIPS/URI is used the domain names of each of the communicating parties is protected from eavesdroppers without however being hidden from intermediate Proxies. There are also some cases where everyone can have access to this information like, for example, when only two domains intervene between the two communicating parties so that it is obvious who belongs to which domain.

IP address protection: The IP addresses of the communicating parties are not protected.

Privacy level: For both caller's and callee's IDs the solution of SIPS URI reaches Level 2 since both real IDs are available to all SIP Proxies in the call path.

Hop-by-hop vs. end-to-end privacy: The privacy protection of this solution is offered in a hop-by-hop fashion.

Stateful vs. stateless mode: Since TLS is used, we need a server that is stateful at the transport level which means that storage requirements are higher. At the application level where SIP operates there is no special need to keep state information. Based on these two observations and taking the SIP Proxy machine as a whole we can argue that it operates in stateful mode.

Deployment: The utilization of this solution has as prerequisite the deployment of a full PKI which issues digital certificates to all end users and intermediate SIP Proxies. Also, currently, there are few SIP clients and network servers that implement TLS and SIPS respectively. Taking into account the administrative effort required to setup a full PKI and the changes needed in the existing infrastructure, this scheme is considered to have difficult deployment.

5.5.3 IPsec

Cryptography: IPsec is based on cryptography to protect data exchanged between two communicating parties.

Authentication: Digest authentication is supported and the corresponding authentication usernames are protected by IPsec.

PKI: IPsec usually bases its operation in pre-shared secret values so no PKI is required. However, if IKE [89] is used with certificates then the deployment of a PKI is necessary.

Anonymity vs. pseudonymity: SIP messages are transmitted through secure channels therefore no user ID is revealed; this means that this solution retains user's anonymity.

Inter-Domain agreements: This scheme is based on already established trust relationships between the two communicating parties. Therefore there should be some kind of pre-existing agreement between administrative domains so that Proxies belonging to different domains can establish secure channels with the use of IPsec.

Multidomain support: This solution can also be utilized in environments where multiple administrative domains exist.

Untrusted proxies: What applies to SIPS URI also applies here.

Domain name protection: What applies to SIPS URI also applies here.

IP address protection: What applies to SIPS URI also applies here.

Privacy level: What applies to SIPS URI also applies here.

Hop-by-hop vs. end-to-end privacy: What applies to SIPS URI also applies here.

Stateful vs. stateless mode: When IPsec is used, SIP Proxies can operate in either of these two modes.

Deployment: The utilization of this solution requires every intermediate node in the call path to have a shared secret with every node it communicates with. This makes it a solution with difficult deployment. The number of IKE pre-configured keys needed in a symmetric key system with n network elements communicating with each other is $O(n^2)$. Also, as already mentioned, if IKE is used with certificates then a full PKI is also required.

5.5.4 Anonymous URI

Cryptography: This solution does not utilize any kind of cryptography.

Authentication: Anonymous URI can support Digest authentication but this would mean that either the username must be revealed or an “anonymous” username must be used.

PKI: No PKI is required for this scheme.

Anonymity vs. pseudonymity: Since no caller ID is transmitted this is a solution based on anonymity.

Domain agreements: This scheme does not require any pre-existing agreements between administrative domains.

Multidomain support: Anonymous URI supports multidomain environments without any modification.

Untrusted proxies: Anonymous URI can preserve user’s anonymity even when untrusted proxies reside in the path between the caller and the callee.

Inter-Domain name protection: When Anonymous URI is utilized the domain name of the caller is never transmitted, while anyone has access to the callee’s domain name.

IP address protection: The IP addresses of the communicating parties are not protected.

Privacy level: Regarding caller’s ID, Anonymous URI is at Level 7, because only the caller is aware of his own ID, while for callee’s ID no protection at all is offered resulting at privacy Level 1.

Hop-by-hop vs. end-to-end privacy: This scheme offers end-to-end privacy for caller’s ID.

Stateful vs. stateless mode: This is a solution that can be supported either by stateful or stateless SIP Proxies.

Deployment: Anonymous URI is a method with easy deployment since no modification to the existing infrastructure is required.

5.5.5 Privacy mechanism for SIP

This mechanism has two ways for protecting user's privacy: user and network provided privacy. When user provided privacy is used then what applies for Anonymous URI as analyzed in the previous section, also applies here. The following analysis is valid when network or both user and network provided privacy is used.

Cryptography: This scheme does not base its operation on cryptography. However, the recommended way the UA contacts its Home Domain is over a TLS session; thus we consider here that cryptography is part of this solution.

Authentication: While Digest authentication can be used with this method, the username is not protected at all. In some occasions this can result in privacy violation, for example when the username is the same as the user ID part of SIP URI.

PKI: Considering that TLS will be used, a limited PKI is needed for the management of certificates for SIP Proxies.

Anonymity vs. pseudonymity: This method uses real SIP URIs inside trusted domains while replacing them with Anonymous URIs when SIP messages leave these trusted domains. Thus, it is a method that offers anonymity to its users.

Inter-Domain agreements: In this scheme a privacy service entity is needed which can be, for example, a trusted SIP Proxy. If this Proxy does not belong to the user's Home Domain then a trust agreement is needed between the Home Domain and Proxy's domain so that the end user can trust the latter.

Multidomain support: This method can support multidomain environments but only strictly under the assumption that these domains have established trust agreements with each other. In other words, if a user is located in a place where there is no administrative domain with trust agreement with his Home Domain then he cannot use the features offered by this solution.

Untrusted proxies: This mechanism cannot guarantee the protection of user's privacy when SIP messages are transmitted through untrusted Proxies before reaching his Home Domain.

Domain name protection: When SIP messages leave a trusted domain they are anonymized; however, the responses must follow the same path back in order to be routed properly to the sender. Thus, the name of the caller's domain cannot be kept secret.

IP address protection: The IP addresses of the communicating parties are not protected.

Privacy level: For caller's ID this mechanism reaches Level 6 while for callee's ID it is at Level 1.

Hop-by-hop vs. end-to-end privacy: While in [86] it is suggested that TLS should be used from UA to its Home Domain's Proxy, this solution as a whole is considered an end-to-end privacy preserving one. This is because inside the trusted domains we can be sure that TLS or other protection methods will be used while outside the domain no real ID is transmitted.

Stateful vs. stateless mode: This mechanism requires state information to be kept in certain Proxies, thus it can only be supported by stateful Proxies.

Deployment: Privacy mechanism for SIP is considered a solution which requires medium deployment effort. The UAs and the Proxies must be modified in order to be able to process the

new privacy header; in addition to that Proxies must have the proper logic to withhold user IDs when this is necessary and route responses properly.

5.6 Discussion

This section provides an overview of some interesting points from the observation of Table 5-1; the first one has to do with ID hiding. In some occasions it is desirable from the caller not to reveal his ID to the callee. This ID hiding type is supported by “Anonymous URI” and “Privacy mechanism for SIP”. The problem here is that these methods cannot support this feature while at the same time protecting the Digest username during the authentication process.

Another possible requirement is the ability of each method to maintain its privacy protecting features while operating through untrusted domains even when these domains are placed between the caller and his Home Domain. While S/MIME can protect the user ID, it cannot protect his username during Digest authentication. Furthermore, it cannot offer caller’s ID hiding from the callee.

Another consideration is that only “Anonymous URI” can protect the Home Domain name of the caller; however this method is less practical since it cannot support authentication. Regarding the IP addresses of the communicating parties it is evident that no method can effectively protect them from eavesdroppers. While both domain names and IP addresses are considered private information they should remain publicly available so that the two parties can communicate with each other during as well as after the session establishment.

5.7 Summary

It is envisioned that in the near future SIP will co-exist or even supersede classic telephony systems like PSTN and traditional multimedia delivery methods. Before this becomes reality certain security issues must be solved. While SIP is a simple and easy to deploy protocol, it turns out that some of the security problems related with it are hard to solve. One such problem is privacy since SIP messages cannot be cryptographically protected as a whole.

As it showed throughout this chapter SIP has a number of security and especially privacy protecting mechanisms; however some privacy issues are still open. Here the focus is on the protection of communicating parties IDs and especially on the review of existing solutions that can protect user IDs and comparing them with each other. This comparison has shown a number of deficiencies in these schemes pointing out the necessity of newer and better methods that will provide adequate privacy protection to end users.

Chapter 6 - PrivaSIP: a framework for protecting privacy in SIP

Secure multimedia delivery in modern and future networks is one of the most challenging problems towards the system integration of fourth generation (4G) networks. This integration means that different service and network providers will have to interoperate in order to offer their services to end users. This multidomain environment poses serious threats to the end user who has contract with and trusts only a limited number of operators and service providers; one such threat is end users' privacy.

Probably the most promising protocol for multimedia session management is the Session Initiation Protocol (SIP) which is an application layer protocol and thus can operate on top of different lower layer technologies. SIP is quite popular and a lot of research has been conducted; however, it still has some security issues, one of which is related to privacy and more particularly the protection of user identities (IDs).

In this chapter the ID privacy issue of SIP is presented in detail and a framework called PrivaSIP that can protect either the caller's ID or both the caller's and the callee's IDs in multidomain environments is proposed. Different implementations of this framework are presented based on asymmetric and symmetric cryptography and an analysis of the pros and cons of each one of them is provided. Furthermore performance measurements are presented in order to find out the performance penalty of this framework over standard SIP. Continuing the SIP privacy methods comparison from the previous chapter, a comparison of PrivaSIP with the existing schemes is provided based on the same criteria.

6.1 PrivaSIP framework

Following the privacy issues analysed in the previous chapter, an approach of defeating such problems when using SIP is presented here. While concealing user IDs could be a sort of protection it would make SIP non operable. That is because SIP needs user IDs in order to locate the correspondent users, route the messages appropriately and possibly charge them for the received services. A more convenient solution would be the revealing of user IDs only to absolutely necessary parties so as to route SIP messages appropriately and possibly authenticate the caller before offering their services.

The proposed solution is an identity protection framework named PrivaSIP [9][11]. The main idea behind the PrivaSIP framework is that each ID should be individually encrypted in a way that it can be recovered only by entities that need to do so in order for the SIP protocol to operate correctly. Hereinafter the term "ID" is used to abbreviate either the user ID part of a SIP URI or other types of user IDs like a Digest authentication username. Two different variations of PrivaSIP are defined here: in the first one only the caller's IDs are protected, while in the second both the caller's and the callee's IDs are protected; these two variations will be referred as PrivaSIP-1 and PrivaSIP-2 respectively. It must be noted that PrivaSIP does not aim at protecting the confidentiality of whole messages or providing message integrity; such requirements should be met by utilizing other mechanisms.

In the subsequent sections five different implementations of PrivaSIP framework using different encryption algorithms are presented. The main purpose of doing so is to find out which

category of algorithms or specific algorithm is more efficient when used in the proposed framework. These implementations fall into two categories following the two PrivaSIP variations: in the first one only the caller ID is protected while in the second one both the caller and the callee IDs are protected from third parties. In all these schemes the caller ID is protected and this can be done either with symmetric cryptography using as a key the Digest authentication password shared between the user and his Home Proxy, or with asymmetric cryptography using the public key of the Home Proxy. In the second category of schemes where the callee ID has to be protected as well, the public keys of both the caller's and the callee's Home Domains are used. This category of schemes uses only asymmetric cryptography since the caller usually does not have any shared secrets with the callee's Home Proxy. The specific cryptographic algorithms used in our case are for PrivaSIP-1: the well known public key algorithm by Rivest, Shamir and Adleman (RSA) [93], the Elliptic Curve Integrated Encryption Scheme (ECIES) [94], and the symmetric Advanced Encryption Standard (AES) [95], and for PrivaSIP-2: RSA, and ECIES. Other asymmetric and symmetric algorithms can also be used based on the same principles.

6.2 PrivaSIP-1

The first PrivaSIP variation, namely PrivaSIP-1, protects the caller's user ID and Digest username. Using the previous analysis of a SIP INVITE message the specific header fields that need protection are presented here, so that user IDs are protected as well. The proposed solution is to strip whichever information is not necessary and use encryption for the rest. More specifically:

- we leave <Via> field's value as is, because it only reveals the IP address of the host
- <Contact> field's value is replaced with the IP address of the caller's host. End users' IP addresses usually are not static so eavesdroppers cannot easily relate it with the permanent ID of the user
- the display name in <From> field ("Smith" in our example) is stripped or replaced by the string "Anonymous"
- the user ID part of <From> field (i.e. "smith" in "smith@minitrue.org") is encrypted using either asymmetric or symmetric cryptography. As it is obvious we propose a scheme that rather relies on pseudonymity than anonymity [91]. If the same pseudonym is always used then the user can be "profiled" and his movement (in case of a mobile user) can be easily tracked. For this reason a padding scheme should be used so that the resulting pseudonym is different every time

The resulting message for the first variation of PrivaSIP, where only the caller ID is protected, is shown below. In this example the hexadecimal representation is used for the encrypted part of the URI and its length depends on the cryptographic algorithm.

```
INVITE sip:obrien@miniluv.org SIP/2.0
Via: SIP/2.0/UDP 195.251.161.144:5060; branch=z9hG4bK74b43
Max-Forwards: 70
From: <sip:0AEE5F83...129F32@minitrue.org>; tag=9fxced76sl
To: O'Brien <sip:obrien@miniluv.org>
Call-ID: 3848276298220188511@minitrue.org
CSeq: 1 INVITE
Contact: 195.251.161.144
Content-Type: application/sdp
Content-Length: 151
```

If authentication is not required then the most practical and effective solution would be the employment of “Anonymous” URI in <From> header. However, in a real world environment the most probable case is that the user must be authenticated in order to be charged for the services he receives. If caller ID privacy is also a requirement then the existing schemes presented in the previous chapter are not adequate. As already stated, here only Digest authentication [92] is considered, which is the standard way of authenticating users in SIP environments.

In the following an example will be presented where both the Local outbound SIP Proxy and Home Proxy require Smith to authenticate in order to receive their services. It is assumed that Smith has a different set of credentials for each of the two domains and he is willing to present each of the two IDs he possesses only to the corresponding domain. Naturally, since Smith has credentials from both domains it means that he has some kind of agreement with each one of them, so he is aware of what kind of private information he presents to each domain. The key point here is that the caller has the choice to present private information only to selected domains minimizing the number of entities that posses this information. Moreover, he reveals to each domain only the private information this domain already know about the user’s ID and not IDs that this user may possess from other domains. Caller ID privacy during the authentication process can be assured in a similar way as in the previous example. When the INVITE message is received, the Local outbound Proxy responds with a 407 Proxy Authentication Required message. Smith sends back a new INVITE where he encrypts the username used in <Proxy-Authorization> field with the (public or shared) key of the Local outbound Proxy as shown below. Also, the user ID part of <From> field is encrypted with the key of Home Proxy. It is worth noting that this encryption process does not imply in any way that it supports user authentication; this task is conducted with the utilization of Digest authentication. The different user IDs used here are in accordance with [49] and reveal each ID only to the intended Proxy.

```
INVITE sip:obrien@miniluv.org SIP/2.0
Via: SIP/2.0/UDP 195.251.161.144:5060; branch=z9hG4bK74b43
Max-Forwards: 70
From: <sip:0AEE5F83...129F32@minitrue.org>; tag=9fxced76sl
To: O’Brien <sip:obrien@miniluv.org>
Call-ID: 3848276298220188511@minitrue.org
CSeq: 1 INVITE
Proxy-Authorization: Digest username=”38A8F347...0EA19A98”,
algorithm=MD5, realm=”minitrue.org”, nonce=”1dea4387...00f4e5da”,
qop=”auth”, opaque=”5e7734afdb981200”, response=”ffa1e3...8756ee”,
nc=00000001, cnonce=”abcdefghi”
Contact: 195.251.161.144
Content-Type: application/sdp
Content-Length: 151
```

The Local outbound Proxy decrypts Smith’s username and completes the authentication process and, if it is successful, it forwards the INVITE to Smith’s Home Proxy. The Home Proxy also completes authentication in the same manner. After that, the initial INVITE message is forwarded to the Inbound Proxy which sends it to O’Brien. As we can see no untrusted entities involved in the protocol (including O’Brien) are aware of Smith’s ID. When O’Brien answers the call he uses the same encrypted headers, and his response travels all the way back to *minitrue.org* where the Proxy deciphers <From> header to discover the recipient of the message.

While the usefulness of PrivaSIP is proven through examples, this does not limit its generality. The same procedure would be followed if, for instance, there were SIP Registrars instead of Proxies and REGISTER messages instead of INVITEs.

In the next sections three implementations will be defined that utilize either asymmetric or symmetric cryptography. The first one is PrivaSIP-1 with RSA or PrivaSIP-1-RSA [9][10][11] for short, and uses the Home Proxy's public key to encrypt the user ID and Digest username when this is necessary. Our next implementation, namely PrivaSIP-1-ECIES [11], is based on the standard encryption scheme for Elliptic Curve Cryptography (ECC) which is ECIES, also known as DHAES [94]. A different approach is used in the last one, which is PrivaSIP-1-AES [11], where a symmetric cryptographic algorithm is utilized, more specifically AES, for the encryption of the caller ID. Since the caller and his Home Proxy share a password which is used for Digest authentication, this password can also be used as a key (or as a key seed or master key) for the encryption of user ID with AES.

6.2.1 Asymmetric cryptography

The first implementation, namely PrivaSIP-1-RSA, utilizes RSA in order to encrypt the respective field values. We have already pointed out the necessity of a padding scheme so that a user cannot be "profiled" even when his IDs are encrypted. For this reason a padding scheme like Optimal Asymmetric Encryption Padding (OAEP) [96] would be suitable for RSA, resulting in a different user pseudonym every time.

PrivaSIP-1-RSA utilizes a digital certificate of the Home Proxy server of the user in order to encrypt his user ID and Digest username so that only this Proxy server has access to it. Thus, a pre-condition for PrivaSIP-1-RSA to work is that there is some sort of PKI, the Home Proxy has a public-private key pair and a corresponding valid digital certificate and the UA possess the public key.

6.2.2 Elliptic curve cryptography

Describing ECIES in high level, to encrypt an amount of data, a new symmetric key is produced each time from the recipient's public key and data are encrypted with a symmetric algorithm using this derived key. For this reason, a padding scheme is not necessary here since the key is different every time so the ID pseudonym will also be different.

PrivaSIP-1-ECIES also requires that the caller's Home Proxy server has been issued a digital certificate so that the user can encrypt his user ID and Digest username. That means that the existence of a PKI is necessary, the Home Proxy must have a public-private key pair and a corresponding valid digital certificate and the UA must somehow possess the public key.

6.2.3 Symmetric cryptography

A suitable padding scheme for PrivaSIP-1-AES would be the one described in [97]; this standard specifies that the padding should be done at the end of the last block of data with random bytes, and the padding boundary should be specified by the last byte. This way, every time the ID pseudonym will have a different value. Other padding schemes like adding zeroes at the end of the data block or adding the same string every time would result in the same pseudonym every time rendering PrivaSIP-1-AES useless in terms of user profiling.

PrivaSIP-1-AES does not use digital certificates so there is no need for a PKI. It does, however, utilize a Digest authentication password shared between the UA and the corresponding SIP

Proxy which is used for the encryption of the user ID and Digest username, thus such credentials must be shared between these two parties.

6.3 PrivaSIP-2

We can further improve PrivaSIP-1 so that it also preserves the called party's ID as well. The second PrivaSIP variation, namely PrivaSIP-2, protects the caller's user ID and Digest username as well as the callee's user ID. Using again the analysis of a SIP INVITE message the specific header fields that need protection are presented here, so that all user IDs are protected. The proposed solution is to strip whichever information is not necessary and use encryption for the rest. More specifically:

- we leave <Via> field's value as is, because it only reveals the IP address of the host
- <Contact> field's value is replaced with the IP address of the caller's host. End users' IP addresses usually are not static so eavesdroppers cannot easily relate it with the permanent ID of the user
- the caller's display name in <From> field ("Smith" in our example) is stripped or replaced by the string "Anonymous"
- the caller's user ID part of <From> field (i.e. "smith" in "smith@minitrue.org") is encrypted using either asymmetric or symmetric cryptography. As it is obvious we propose a scheme that rather relies on pseudonymity than anonymity [91]. If the same pseudonym is always used then the user can be "profiled" and his movement (in case of a mobile user) can be easily tracked. For this reason a padding scheme should be used so that the resulting pseudonym is different every time
- the previous encryption procedure applies to the callee's user ID found in Request-URI and <To> field.

In order to present the inner workings of PrivaSIP-2, the same example will be used. The protection of callee's ID is achieved by a similar mechanism as the caller's ID with the use of asymmetric cryptography; more specifically the ID is encrypted with the public key of callee's Home Domain SIP Proxy. This variation uses only asymmetric cryptography since the caller usually does not have any shared secrets with the callee's Home Proxy. It must be noted here that the caller's ID is also protected as shown in the previous section.

As already discussed, some SIP headers of an INVITE message must be altered to protect the ID of the caller. Apart from these headers, in this scheme <To> field and Request-URI which exposes callee's ID are also protected. The resulting message is shown below:

```
INVITE sip: 73D8A9F7...BC09E1A1@miniluv.org SIP/2.0
Via: SIP/2.0/UDP 195.251.161.144:5060; branch=z9hG4bK74b43
Max-Forwards: 70
From: <sip:0AEE5F83...129F32@minitrue.org>; tag=9fxced76sl
To: <sip:73D8A9F7...BC09E1A1@miniluv.org>
Call-ID: 3848276298220188511@minitrue.org
CSeq: 1 INVITE
Contact: 195.251.161.144
Content-Type: application/sdp
Content-Length: 151
```

What applies for user authentication in PrivaSIP-1 also applies here. The caller can hide both his ID and Digest username, while also the callee's ID is protected from third parties. The

procedure that is followed is the same as presented previously except that when an INVITE is forwarded to the Inbound SIP Proxy, the <To> field is decrypted and subsequently send to O'Brien. When O'Brien responds back he uses the same encrypted headers so that the privacy enhanced SIP message is routed appropriately.

In PrivaSIP-2 two implementations using asymmetric cryptography are defined. Here an approach based on symmetric cryptography (e.g. AES) cannot be applied to protect callee's ID since it is difficult for the caller to have a shared secret with every possible callee's Home Proxy. In the first implementation the RSA algorithm is utilized resulting in PrivaSIP-2-RSA while the second implementation is based on ECIES and will be referred as PrivaSIP-2-ECIES.

6.3.1 Asymmetric cryptography

In the first implementation of PrivaSIP-2, namely PrivaSIP-2-RSA, RSA is utilized in order to encrypt the necessary field values. The necessity of a padding scheme has already been pointed out so that users cannot be "profiled" even when his IDs are encrypted. For this reason a padding scheme like Optimal Asymmetric Encryption Padding (OAEP) [96] would be suitable for RSA, resulting in a different pseudonym every time for the caller as well as for the callee.

PrivaSIP-2-RSA utilizes a digital certificate of the Home Proxy server of the caller in order to encrypt his user ID and Digest username so that only this Proxy server has access to it; it also uses a digital certificate of the Home Proxy server of the callee in order to encrypt his user ID as well. Thus, the pre-conditions of PrivaSIP-2-RSA is that there must be some sort of PKI, the Home Proxy of both the caller and the callee have a public-private key pair and corresponding valid digital certificates and the UA of the caller possesses these certificates.

6.3.2 Elliptic curve cryptography

For the second implementation of PrivaSIP-2, Elliptic Curve Cryptography is used with ECIES resulting in PrivaSIP-2-ECIES; its operation is identical to the previous one based on RSA. This way in both schemes the caller's ID is only recoverable by his Home Domain and the callee's ID is only disclosed to his own Home Proxy. In this implementation a padding scheme is not used since every time a different user pseudonym is produced.

PrivaSIP-2-ECIES utilizes a digital certificate of the Home Proxy server of the caller in order to encrypt his user ID and Digest username so that only this Proxy server has access to it; it also uses a digital certificate of the Home Proxy server of the callee in order to encrypt his user ID as well. Thus, the pre-conditions of PrivaSIP-2-ECIES is that there must be some sort of PKI, the Home Proxy of both the caller and the callee have a public-private key pair and corresponding valid digital certificates and the UA of the caller possesses these certificates.

6.4 Experimental testbed setup

The performance of all the implementations of PrivaSIP for both the client and the server was evaluated [9][11] in a properly designed testbed and the results are depicted in this section. It is well known that security or privacy mechanisms come always at a cost. However, apart from the effectiveness and robustness of the proposed mechanism, the key question in every case is if that cost is affordable. So, the main purpose here is not to evaluate SIP's performance in general but to determine the performance penalty imposed by PrivaSIP compared to standard SIP transactions. In the previous chapter all known schemes that could be used for providing some sort of privacy in SIP were extensively analyzed and compared to each other. However, the

performance of these methods is not compared with that of PrivaSIP here. The chief reason to do so is that each scheme presents different qualities, and each of them is useful under a specific context irrespective of the performance penalty one might impose. For example, when the user ID must be protected and authentication is also a requirement, then PrivaSIP is the only viable solution; when authentication is not a requirement, then Anonymous URI is the right choice. Also, other solutions either do not provide enough or assured privacy (IPsec, SIPS URI/TLS) or do not protect privacy during authentication (Privacy mechanism) or do not support authentication at all (S/MIME, Anonymous URI).

The results that have been tracked and logged are based on two distinct scenarios:

1. Client delay. The time required for a UA to construct an INVITE request was measured; moreover, for comparison purposes, these measurements were recorded when PrivaSIP is in use and when it is not. The measured request creation phase constitutes from the preparation of all SIP headers including the encryption of the respective user IDs when PrivaSIP is utilized. The experiments were conducted using a "low-end client" as well as a "high-end client" so that it could be possible to investigate what is the impact of different implementations of PrivaSIP on different hardware configurations. This scenario runs only on clients and does not involve any network interaction since only the INVITE preparation delay is measured.
2. Server delay. In the second scenario the time required for a SIP Proxy server with different queue sizes to serve a request was measured. The scenario was executed one time for each of PrivaSIP's implementations and once using standard SIP using different queue sizes ranging from 100 to 1000 calls. For each queue size the call rate is automatically adjusted by SIPp [98]. The measured time starts when an INVITE is send and ends when a "180 Ringing" is received by SIPp as shown in Figure 6-1; this means that the user has been authenticated and his call has reached the intended recipient. It must be noted here that the worst case scenarios are executed; all SIP URIs and digest usernames are computed each time they are needed and no party stores call state information. The delays included are:
    - the parsing of the unauthenticated INVITE by Home Proxy (for PrivaSIP the SIP server used is SIP Express Router - SER [99] which decrypts caller's URI),
    - the digest response preparation time by the caller's UA (no encryption takes place here; the encrypted values used are hardcoded in SIPp's scenario file),
    - the parsing of UA's response (for PrivaSIP this involves the decryptions of UA's ID and Digest username),
    - the parsing of INVITE by Inbound Proxy (for PrivaSIP-2 only, this involves the decryptions of callee's URI) and finally
    - the respective network delays.

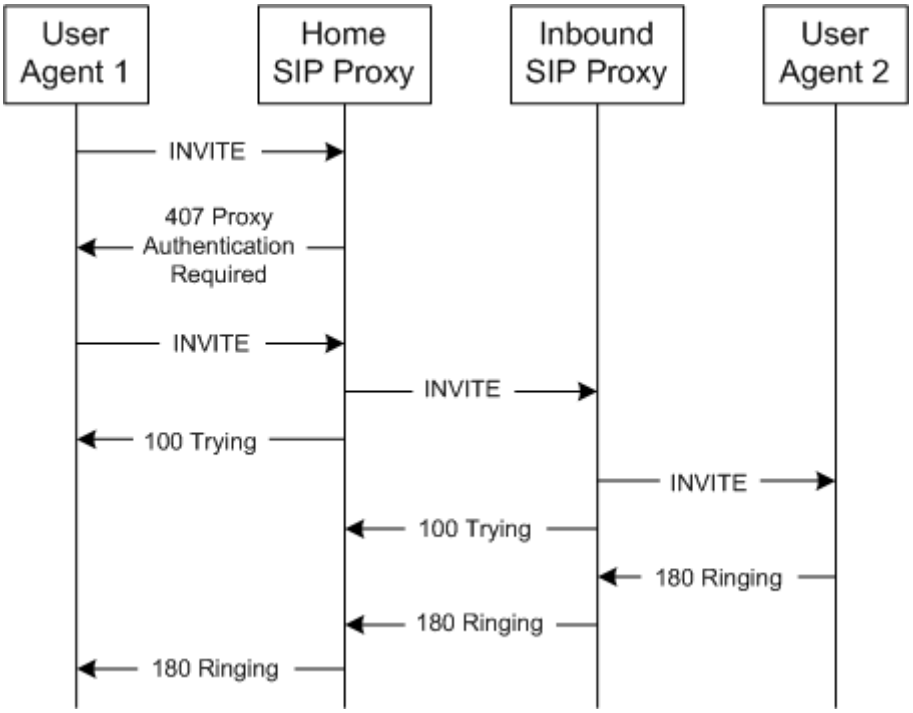


Figure 6-1: SIP call flow

In order to conduct the experiments a testbed was constructed which comprises from the following elements (also summarized in Table 6-1):

- one low-end laptop machine which incorporates an AMD Mobile Athlon 4 CPU at 1.2 GHz and 256 MB of RAM. For the purposes of our experiments, the laptop's CPU was downgraded from 1.2 GHz to 500 MHz with the use of Powersave daemon version 0.10.15, which is part of the machine's Operating System (OS). This enabled us to have similar capabilities as today's handheld and mobile devices. This laptop's network interface was not used since it ran only the client scenario as a "low-end UA". The OS of this machine is SuSE Linux 10.0, kernel version 2.6.13-15-smp, with gcc version 4.0.2 and the software used for measuring client's delay is based on Twinkle SIP softphone version 1.1 [100].
- one desktop PC with an Intel Pentium 4 Hyper-Threading CPU at 2.6 GHz and 512 MB of RAM, which also does not utilize its network card since it is the "high-end User Agent (UA)" for measuring client delay. The OS of this machine is SuSE Linux 10.0, kernel version 2.6.13-15-smp, with gcc version 4.0.2 and the software used for measuring client's delay is based on Twinkle SIP softphone version 1.1.
- one desktop with a dual-core Intel Pentium 4 CPU at 3 GHz and 1 GB of RAM which plays the role of "User Agent 1" in Figure 6-1. This machine connects to the network through a Broadcom NetXtreme Gigabit Ethernet card. Its purpose is to make multiple calls to User Agent 2 through the two Proxies so that we can measure the delay of each request when the Proxies have queue sizes of certain length. This is realized with the use of SIPp 3.0 in client mode which automatically adjusts the call rate so that a stable queue size is maintained. This machine's OS is openSuSE Linux 10.3, kernel version 2.6.22.18-0.2, with gcc version 4.2.1.
- one PC with a dual-core AMD Athlon X2 64 CPU at 1.9 GHz and 2 GB of RAM which plays the role of "Home SIP Proxy" in Figure 6-1. This machine connects to the network through a Realtek RTL8102E Fast Ethernet 100 Mbps network card. The SIP proxy software is based on SER version 0.9.6 supported by MySQL version 5.0.45-community [101] during the

authentication procedure. This machine's OS is openSuSE Linux 11 (32-bit version), kernel version 2.6.25.16-0.1 with gcc version 4.3.

- one desktop PC with a dual-core Intel Pentium 4 CPU at 2.8 GHz and 1 GB of RAM, which connects to the network through a Broadcom NetXtreme Gigabit Ethernet card and is used as the ”Inbound SIP Proxy” in Figure 6-1. The OS of this PC is openSuSE Linux 11, kernel version 2.6.25.16-0.1 with gcc version 4.3. The SIP proxy software is based on SER version 0.9.6.
- one desktop with a dual-core Intel Pentium 4 CPU at 2.6 GHz and 512 MB of RAM which plays the role of “User Agent 2” in Figure 6-1. This machine connects to the network through a Broadcom NetXtreme Gigabit Ethernet card. Its purpose is to receive the calls made by User Agent 1 and send back a “180 Ringing” message which is realized with the use of SIPp 3.0 in server mode. The OS of this PC is openSuSE Linux 11, kernel version 2.6.25.16-0.1 with gcc version 4.3.

Machine	CPU	RAM	OS	Software
Low-end UA	500 MHz (AMD Mobile Athlon)	256 MB	SuSE Linux 10.0, kernel v. 2.6.13-15	gcc 4.0.2, Twinkle 1.1, OpenSSL 0.9.8g, Crypto++ 5.5.2
High-end UA	2.6 GHz (Intel Pentium 4 Hyperthreading)	512 MB	SuSE Linux 10.0, kernel v. 2.6.13-15	gcc 4.0.2, Twinkle 1.1, OpenSSL 0.9.8g, Crypto++ 5.5.2
UA 1	Dual-core 3 GHz (Intel Pentium 4)	1024 MB	openSuSE Linux 10.3, kernel v. 2.6.22.18-0.2	gcc 4.2.1, SIPp 3.0, OpenSSL 0.9.8g, Crypto++ 5.5.2
Home SIP Proxy	Dual-core 1.9 GHz (AMD Athlon X2 64)	2048 MB	openSuSE Linux 11 (32-bit), kernel v. 2.6.25.16-0.1	gcc 4.3, SER 0.9.6, MySQL 5.0.45-community, OpenSSL 0.9.8g, Crypto++ 5.5.2
Inbound SIP proxy	Dual-core 2.8 GHz (Intel Pentium 4)	1024 MB	openSuSE Linux 11, kernel v. 2.6.25.16-0.1	gcc 4.3, SER 0.9.6, OpenSSL 0.9.8g, Crypto++ 5.5.2
UA 2	Dual-core 2.6 GHz (Intel Pentium 4)	512 MB	openSuSE Linux 11, kernel v. 2.6.25.16-0.1	gcc 4.3, SIPp 3.0, OpenSSL 0.9.8g, Crypto++ 5.5.2

Table 6-1: Employed testbed components

Two different 1024 bit RSA digital certificates were issued for the Home Proxy and the Inbound Proxy to be used from PrivaSIP-1-RSA and PrivaSIP-2-RSA, and the corresponding public keys have been transferred to the UAs. For ECIES we have used 160 bit keys, and for AES 128 bit keys. Cryptographic operations related to RSA and AES were executed with the help of the open source OpenSSL library version 0.9.8g [102], while for ECIES we used Crypto++ library version 5.5.2 [103]. The measurements where conducted on the network architecture shown in Figure 6-2. UA 1 and Home SIP Proxy reside in the same 100 Mbps LAN while Inbound SIP Proxy and UA 2 reside in another 100 Mbps LAN. The two sub-networks connect through the Internet over a 2 Mbit ADSL connection with 2048 Mbps maximum downlink and 256 Kbps maximum uplink

speed. The average ping time between the two sub-networks is 22 msec but this value can only be considered as an indication.

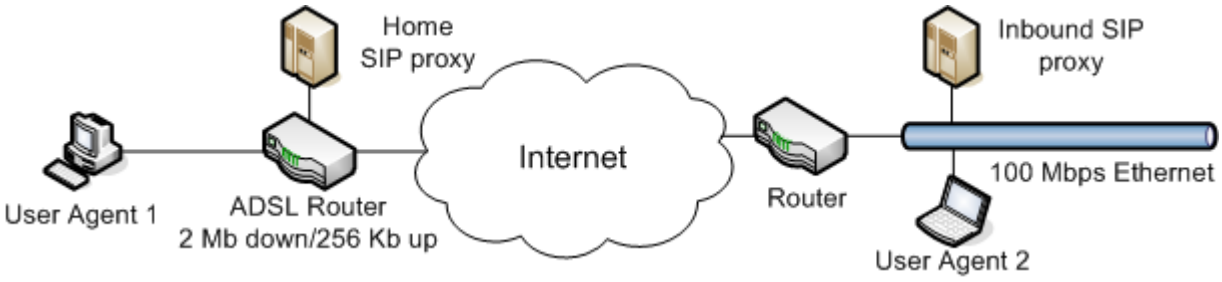


Figure 6-2: Testbed network architecture

The following modifications have been made to the initial versions of the open source software used:

- Twinkle. Twinkle was modified so that it encrypts <From> field and caller's Digest username in PrivaSIP-1 while for PrivaSIP-2 it encrypts both <From> and <To> fields and caller's Digest username. Encryptions involving RSA and AES are performed using OpenSSL while for those involving ECIES Crypto++ was employed.
- SER. Our modified SER decrypt the user IDs, processes the request and forwards the message with the original encrypted user IDs. When Digest authentication is used it also decrypts the username of the UA.
- SIPp. SIPp creates SIP messages based on an XML file that describes a scenario. While encrypted SIP URIs are parsed correctly, we had to modify SIPp in order to parse long usernames (in our case 256 characters). When a 407 Proxy-Authorization request is received, SIPp's response includes the encrypted forms of the user ID and the username used for authentication.

6.5 Experimental results

For the client delay scenario we have taken measurements with twelve different configurations; our five implementations plus standard SIP run on a high-end UA as well as on a low-end-client. For each configuration we have measured the delay of the preparation of a single INVITE message 1,000 times. These configurations are:

1. High-end UA with standard SIP
2. High-end UA with PrivaSIP-1-RSA
3. High-end UA with PrivaSIP-1-ECIES
4. High-end UA with PrivaSIP-1-AES
5. High-end UA with PrivaSIP-2-RSA
6. High-end UA with PrivaSIP-2-ECIES
7. Low-end UA with standard SIP
8. Low -end UA with PrivaSIP-1-RSA
9. Low -end UA with PrivaSIP-1-ECIES
10. Low -end UA with PrivaSIP-1-AES
11. Low -end UA with PrivaSIP-2-RSA
12. Low -end UA with PrivaSIP-2-ECIES

Table 6-2 shows the results for each of the 12 different configurations. Apart from the mean delay, we have included in the table the minimum and maximum delays, the standard deviation of the taken measurements and the 95% confidence interval.

Configuration	Delay (msec)			Standard deviation	Confidence interval (95%)
	Mean	Min	Max		
1	0.16	0.14	1.34	0.07	(0.15, 0.16)
2	0.61	0.55	3.01	0.13	(0.6, 0.62)
3	4.88	4.34	8.42	0.48	(4.85, 4.91)
4	0.18	0.17	0.23	0.01	(0.18, 0.19)
5	0.99	0.89	3.29	0.24	(0.97, 1)
6	9.53	8.88	53.13	1.47	(9.44, 9.62)
7	0.38	0.31	6.11	0.20	(0.37, 0.4)
8	1.6	1.36	8.14	0.26	(1.59, 1.61)
9	24.22	22.26	251.26	7.23	(23.78, 24.67)
10	0.47	0.34	2.24	0.12	(0.46, 0.48)
11	2.66	2.33	10.36	0.48	(2.63, 2.69)
12	46.89	44.17	280.76	7.49	(46.43, 47.36)

Table 6-2: SIP request preparation delay

The observation of the table reveals that when PrivaSIP is in use the INVITE preparation delay is in some cases significantly higher compared to standard SIP and this is obviously due to cryptographic operations involved. The highest delays are observed when ECIES is in use. However, all delays measured are in msecs with an overall maximum of 280.76 msecs, meaning that actually there is no perceived delay by the end user. Also standard deviation of all values remains low, showing that their majority is spread near the mean delay. This observation is further supported by the calculated confidence intervals.

Figure 6-3 shows the impact of hardware configuration on INVITE request preparation delay for the different implementations of PrivaSIP-1, and Figure 6-4 the corresponding delays for PrivaSIP-2. Here we depict the mean preparation delay values presented in Table 6-2 adding the confidence intervals for each mean value as error bars on the graph. The X axis represents the scheme used, while Y axis shows the INVITE preparation delay in msecs.

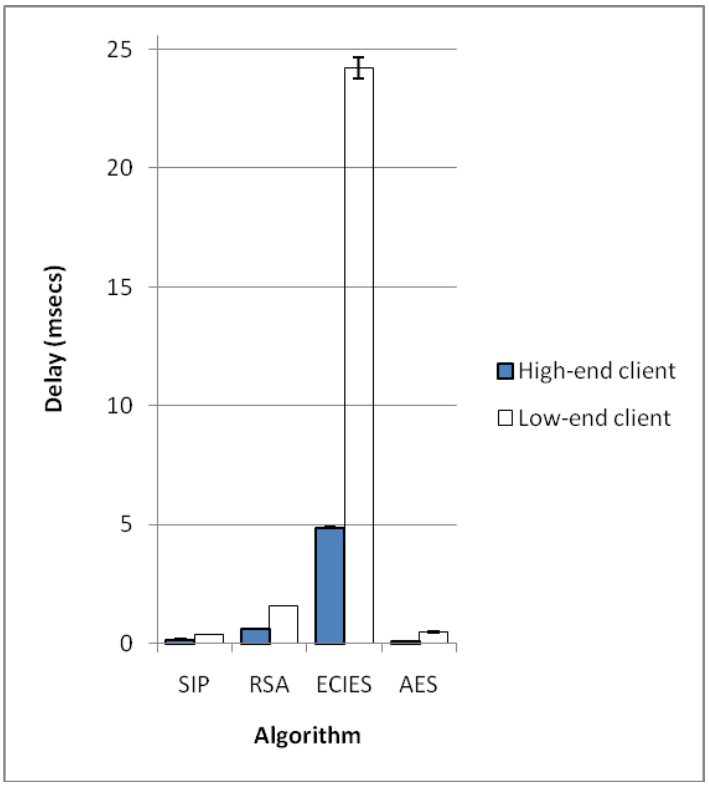


Figure 6-3: Mean INVITE preparation delays for PrivaSIP-1

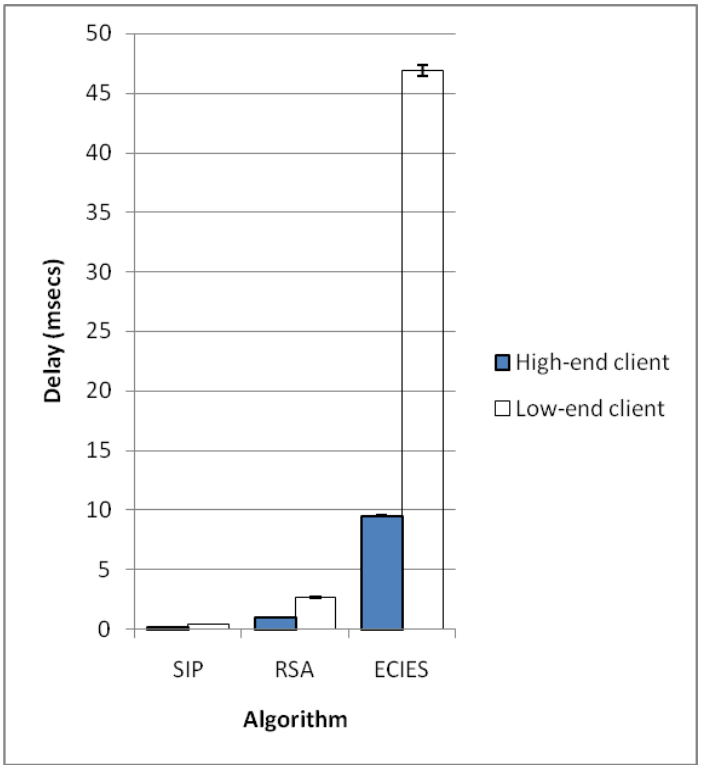


Figure 6-4: Mean INVITE preparation delays for PrivaSIP-2

During the execution of the second scenario we measured the mean server response times for different queue sizes for each implementation. These queue sizes range from 100 to 1,000 calls

and for each one of them we computed the mean response time of 1,000 authenticated calls. For each different implementation examined, server's queue is populated with similar requests, i.e., standard SIP messages for measuring standard SIP's response delays, PrivaSIP-1-RSA messages for measuring PrivaSIP-1-RSA etc. Server's queue population was realized with the SIPp tool, which can create multiple calls with automatically adjusted call rate, so as to keep server's queue at a predefined stable length.

Tables 6-3 to 6-8 show the results for the second scenario. These tables demonstrate the mean server response delays from the moment the user initiates a call until he gets back a "180 Ringing" message; for each implementation we also include the standard deviation of each mean value and the 95% confidence interval. From these results we infer that there is an overhead in PrivaSIP in comparison to standard SIP regarding the response delays. However, these results are based on the assumption that in the first case we only have standard SIP requests while in the second case only PrivaSIP requests. In a more realistic scenario (where probably privacy will be offered with some additional cost) the requests will be mixed at all SIP proxies involved and the performance penalty will be decreased. Furthermore, as it has already been explained, here we consider a worst case scenario regarding the number of cryptographic operations; keeping state information in some SIP Proxies and reusing encrypted URIs will improve the performance of our schemes.

Taking PrivaSIP-2-RSA as an example, in a full roundtrip as shown in Figure 6-1, 6 decryptions take place; 4 in Home Proxy (first INVITE's <From> decryption, second INVITE's <From> and Digest username decryption, 180 Ringing <From> decryption) and 2 in Inbound Proxy (INVITE's <To> decryption, 180 Ringing <To> decryption). These decryptions could be limited to 2 if: (a) the client uses the same encrypted URI for all messages of a session, (b) the server stores a correspondence of the encrypted URI and its decrypted value, and (c) Digest username is the same with <From> user ID. To show the performance improvement that can be achieved we take the delays for server queue sizes of 1,000 calls. The difference between PrivaSIP-2-RSA and standard SIP, that is 1749.49 - 1049.62 = 699.87 msecs, is mainly due to cryptographic operations. So for each cryptographic operation we have a mean delay of 699.87/6 = 116.65 msecs. Following the above optimizations we will have 2 cryptographic operations adding to the delay of standard SIP, i.e., 1049.62 + 2 × 116.65 = 1282.92 msecs which is a lot better than 1749.49 msecs that we measured without any optimization. Of course this is not an accurate value but an estimation, which however, shows how much faster PrivaSIP can be. It is up to the system administrator to decide and make the proper tradeoff between speed and storage needed for keeping state information.

Server queue size (calls)	Mean delay (msec)	Standard deviation	Confidence interval (95%)
100	483.2	757.93	(441.02, 525.39)
200	601.72	1019.41	(543.92, 659.52)
300	693.21	1183.6	(623.74, 762.69)
400	738.24	1243.67	(665.66, 810.81)
500	773.7	1306.31	(696.61, 850.79)
600	919.89	1464.04	(832.61, 1007.16)
700	861.13	1364.84	(779.42, 942.83)
800	934.92	1454.72	(847.26, 1022.58)
900	873.53	1361.27	(791.31, 955.75)
1000	1049.62	1461.18	(959.55, 1139.69)

Table 6-3: Mean server response delays for SIP

Server queue size (calls)	Mean delay (msec)	Standard deviation	Confidence interval (95%)
100	755.77	992.59	(699.13, 812.4)
200	1030.39	1392.59	(941.46, 1119.32)
300	1145.02	1472.4	(1047.01, 1243.03)
400	1205.5	1536.38	(1100.34, 1310.65)
500	1149.63	1434.47	(1051.21, 1248.05)
600	1155.68	1460.5	(1055.96, 1255.4)
700	1213.87	1543.15	(1108.12, 1319.62)
800	1177.49	1515.59	(1072.72, 1282.25)
900	1279.31	1629.86	(1163.13, 1395.49)
1000	1209.43	1514.79	(1106.75, 1312.11)

Table 6-4: Mean server response delays for PrivaSIP-1-RSA

Server queue size (calls)	Mean delay (msec)	Standard deviation	Confidence interval (95%)
100	627.03	877.59	(577.83, 676.24)
200	828.42	1223.53	(756.79, 900.04)
300	993.72	1431.08	(907.41, 1080.03)
400	999.01	1404.05	(912.5, 1085.51)
500	951.42	1348.1	(868.73, 1034.11)
600	973.94	1369.37	(889.11, 1058.77)
700	1000.87	1385.58	(915.21, 1086.54)
800	991.05	1401.81	(904.08, 1078.03)
900	1065.5	1494.8	(972.76, 1158.24)
1000	1019.43	1409.61	(933.09, 1105.77)

Table 6-5: Mean server response delays for PrivaSIP-1-ECIES

Server queue size (calls)	Mean delay (msec)	Standard deviation	Confidence interval (95%)
100	494.43	806.16	(449.32, 539.53)
200	586.44	969.07	(531.12, 641.75)
300	716.1	1182.66	(656.95, 775.24)
400	716.35	1174.95	(647.1, 785.59)
500	731.04	1218.94	(658.81, 803.27)
600	776.28	1295.45	(699.65, 852.9)
700	762.36	1219.36	(689.81, 834.91)
800	834.99	1346.2	(754.29, 915.69)
900	860.8	1356.03	(778.94, 942.67)
1000	936.5	1433.47	(848.75, 1024.26)

Table 6-6: Mean server response delays for PrivaSIP-1-AES

Server queue size (calls)	Mean delay (msec)	Standard deviation	Confidence interval (95%)
100	1244.53	1254.74	(1152.69, 1336.37)
200	1522.85	1592.64	(1382.97, 1662.73)
300	1651.57	1662.91	(1497.93, 1805.21)
400	1634.69	1644.03	(1477.84, 1791.55)
500	1741.45	1744	(1578.31, 1904.59)
600	1576.77	1611.76	(1417.22, 1736.32)
700	1721.93	1701.06	(1562.63, 1881.24)
800	1800.18	1853.96	(1627.34, 1973.02)
900	1858.92	1821.25	(1685.77, 2032.07)
1000	1749.49	1718.94	(1587.58, 1911.39)

Table 6-7: Mean server response delays for PrivaSIP-2-RSA

Server queue size (calls)	Mean delay (msec)	Standard deviation	Confidence interval (95%)
100	886.16	1018.81	(822.8, 949.53)
200	1150.4	1346.81	(1063.47, 1237.33)
300	1363.94	1571.47	(1256.25, 1471.63)
400	1403.3	1575.52	(1294.6, 1512.01)
500	1497.01	1726.1	(1376.34, 1617.68)
600	1523.73	1755.12	(1400.64, 1646.83)
700	1449.43	1623.98	(1336.04, 1562.82)
800	1433.99	1672.32	(1315.63, 1552.34)
900	1392.1	1591.52	(1272.04, 1512.17)
1000	1378.73	1571.97	(1258.98, 1498.47)

Table 6-8: Mean server response delays for PrivaSIP-2-ECIES

Figure 6-5 and Figure 6-6 depict the mean server response delays for all our implementations for different server queue sizes. The X axis represents the size of the queue, while Y axis shows the mean response delay computed for each queue size in msecs. In each point we have also included the corresponding confidence interval as error bars.

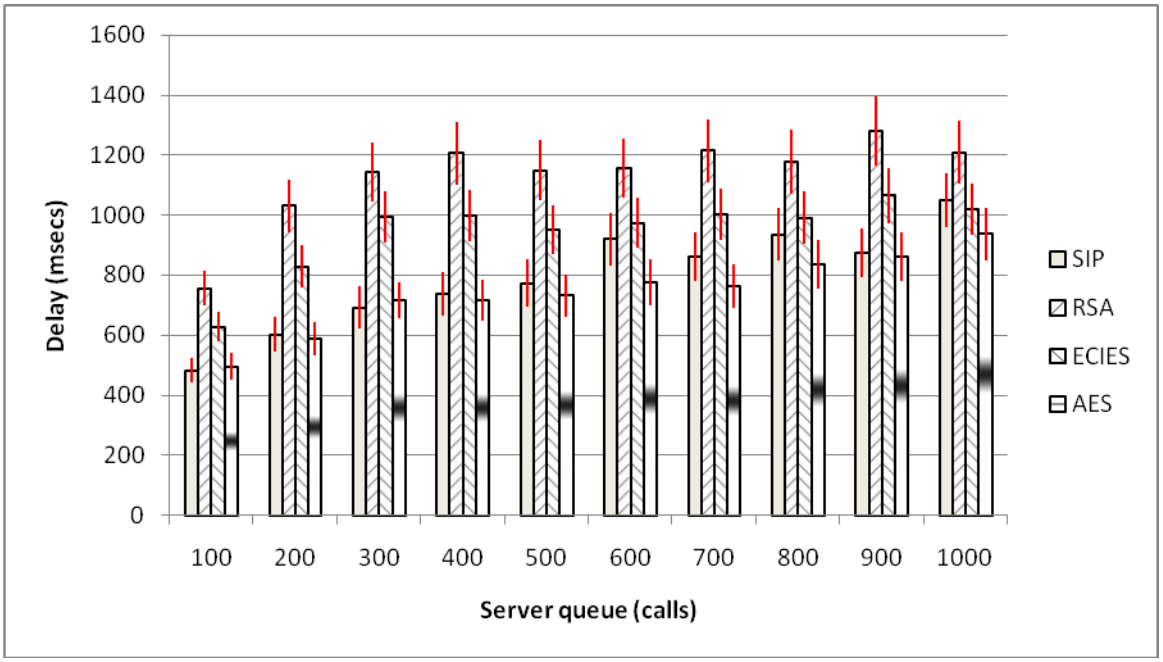


Figure 6-5: Server response delays for PrivaSIP-1

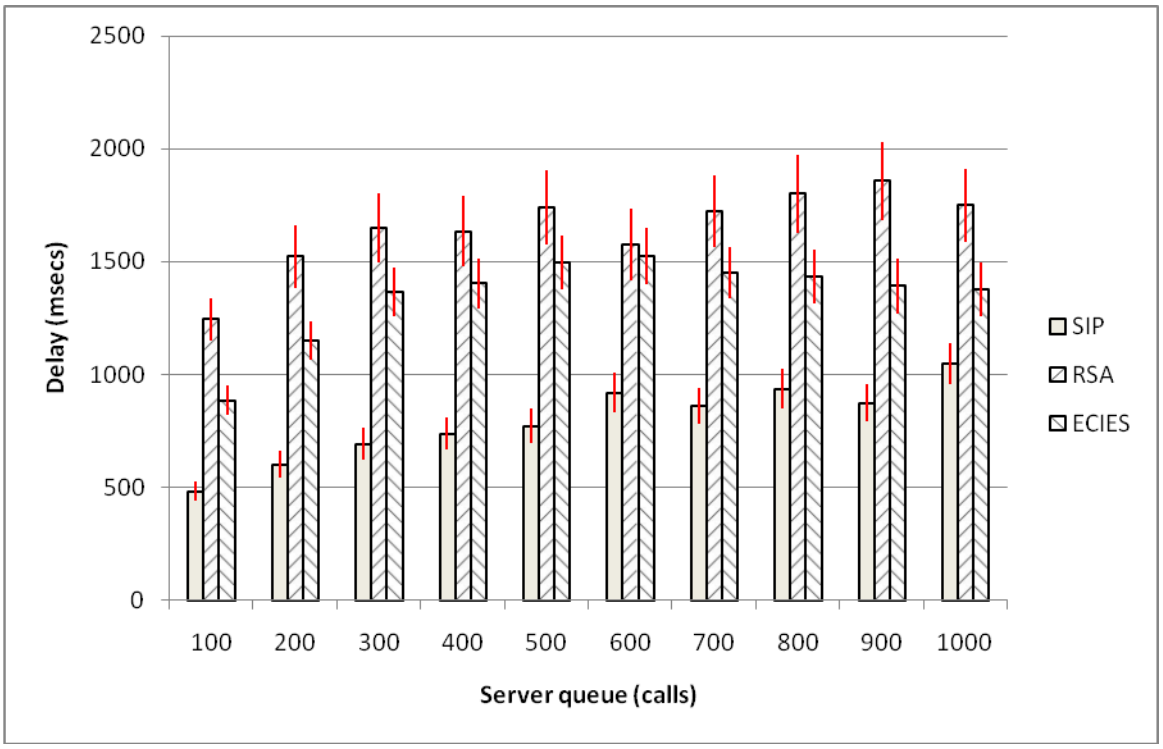


Figure 6-6: Server response delays for PrivaSIP-2

6.6 Comparison with existing schemes

This section provides a comparison of PrivaSIP with related schemes presented in the previous chapter. The same criteria and methodology of the previous chapter will be used in order to have an as accurate comparison as possible. The results of this comparison are summarized in Table 6-9. In this table the findings of the previous chapter were included so as to have a qualitative comparison of all schemes at a glance.

Schemes / Criteria		S/MIME	SIPS URI/TLS	IPsec	Anonymous URI	Privacy mechanism	PrivaSIP-1	PrivaSIP-2
Cryptography		√	√	√	×	√	√	√
Authentication		×	√	√	×	×	√	√
PKI		full	full	×	×	limited	limited[8]	limited
Anonymity vs. pseudonymity		anonymity	anonymity	anonymity	anonymity	anonymity	pseudonymity	pseudonymity
Inter-Domain agreements		×	√	√	×	√	×	×
Multidomain support		√	√	√	√	√	√	√
Untrusted proxies		×	×	×	×	×	√	√
Domain name protection		×	×	×	√	×	×	×
IP address protection		×	×	×	×	×	×	×
Privacy level	Caller	5	2	2	7	6	6	6
	Callee	1	2	2	1	1	1	4
Hop-by-hop vs. end-to-end privacy		end-to-end	hop-by-hop	hop-by-hop	end-to-end	end-to-end	end-to-end	end-to-end
Stateful vs. stateless		both	stateful	both	both	stateful	both	both
Deployment		difficult	difficult	difficult	easy	medium	medium	medium

√: supported/required

×: not supported/not required

Table 6-9: Privacy schemes comparison

[8] It depends on the implementation. For example, in PrivaSIP-1-AES no PKI is needed.

6.6.1 PrivaSIP-1

Cryptography: This solution protects user's privacy based on cryptography.

Authentication: PrivaSIP-1 supports Digest authentication; furthermore during the authentication process the username of the caller is protected.

PKI: No PKI at all or a limited PKI is needed. The term "limited" is used because digital certificates will be issued and managed only for Proxies and not for end users. Moreover, managing certificates for a small number of trusted servers is easier than doing the same for all SIP users. If a method like PrivaSIP-1-AES is used then there is no need to setup a PKI.

Anonymity vs. pseudonymity: The protection of user's ID involves the encryption of this ID and the transmission of its encrypted form. This encrypted form is a pseudonym and the real ID can be recovered by this pseudonym by entitled entities.

Inter-Domain agreements: This scheme does not require any kind of trust agreement to exist between different administrative domains.

Multidomain support: This method can support multidomain environments even when different administrative domains do not have established any kind of trust agreement between them.

Untrusted proxies: This mechanism can protect caller's IDs and Digest authentication passwords even when untrusted proxies exist in the path between the user and his Home Domain.

Domain name protection: This scheme does not protect the name of the caller's Home Domain.

IP address protection: The IP addresses of the communicating parties are not protected.

Privacy level: For caller's ID this mechanism reaches Level 6 while for callee's ID it is at Level 1.

Hop-by-hop vs. end-to-end privacy: PrivaSIP-1 offers end-to-end privacy.

Stateful vs. stateless mode: This mechanism can be supported by either stateful or stateless SIP Proxies.

Deployment: The modification needed by this scheme in UAs and Proxies is the addition of encryption/decryption abilities. Apart from this, a PKI is needed which is however limited to manage certificates issued only to Proxies. Due to the limited nature of PKI this method is consider to require medium deployment effort.

6.6.2 PrivaSIP-2

Cryptography: PrivaSIP-2 is also based on cryptography.

Authentication: This method supports Digest authentication while at the same time protecting the username of the caller.

PKI: Here a limited PKI is needed.

Anonymity vs. pseudonymity: Both users IDs (caller's and callee's ID) are encrypted prior to their transmission and are pseudonyms of the real IDs.

Inter-Domain agreements: Similarly to the previous one, this scheme does not require any trust agreements between different administrative domains.

Multidomain support: Multidomain environments can be supported in this method even when different administrative domains do not have established any kind of trust agreement between them.

Untrusted proxies: PrivaSIP-2 protects both caller's and callee's IDs and Digest authentication passwords even when untrusted proxies exist anywhere in the call path.

Domain name protection: This scheme does not protect domain names.

IP address protection: The IP addresses of the communicating parties are not protected.

Privacy level: For caller's ID this mechanism reaches Level 6 while for callee's ID it reaches Level 4.

Hop-by-hop vs. end-to-end privacy: This scheme offers end-to-end privacy.

Stateful vs. stateless mode: This mechanism can be supported by either stateful or stateless SIP Proxies.

Deployment: What applies in PrivaSIP-1, also applies here; hence this method needs medium deployment effort.

6.7 Discussion

This section provides a discussion on some interesting points from comparing related schemes with PrivaSIP and from the observation of the performance measurements. The first one has to do with ID hiding. In some occasions it is desirable from the caller not to reveal his ID to the callee. This ID hiding type is supported by PrivaSIP schemes and by two other schemes as well, namely "Anonymous URI" and "Privacy mechanism for SIP". The difference here is that only the two PrivaSIP variations can support this feature while simultaneously protecting the Digest username during the authentication process.

Perhaps the most important advantage of PrivaSIP methods is their ability to maintain their privacy protecting features while operating through untrusted domains even when these domains are placed between the caller and his Home Domain. While S/MIME can also protect the user's ID, it cannot protect his username during Digest authentication. Furthermore, it cannot offer caller's ID hiding from the callee.

Another consideration is that only "Anonymous URI" can protect the Home Domain name of the caller; however this method is less practical since it cannot support authentication. Regarding the IP addresses of the communicating parties it is evident that no method can effectively protect them from eavesdroppers; a possible exception to this would be the employment of IPsec in tunnel mode but this would create performance issues. While both domain names and IP addresses are considered private information they should remain publicly available so that the two parties can communicate with each other during as well as after the session establishment.

One final remark concerning PrivaSIP is the acquisition of Proxy certificates. It is assumed that the UAs have in their possession the digital certificates of the Proxies they need. This is a logical assumption concerning the Home Proxy certificate of each user; however the same cannot be straightforwardly asserted for other Proxies. Thus, when PrivaSIP-2 is utilized the caller's UA should first acquire and check the certificate of the callee's Home Proxy and then proceed to the protection of the messages. This however happens usually once and stands for multiple sessions, i.e., until the certificate of the corresponding foreign SIP Proxy that contains the public key expires. When symmetric cryptography is utilized it is assumed that the user shares a Digest password with his Home Proxy which is of course a viable assumption.

As it has already been stated, each PrivaSIP variation serves different purposes. From the conducted experiments some conclusions as to which implementation is more efficient and easy

to deploy in different contexts can be drawn. For PrivaSIP-1 it seems that the implementation with AES is the best choice since it is the most efficient one and the easiest in deployment; it only reuses the already shared Digest password between the caller and his Home Proxy and does not mandate any sort of PKI. For PrivaSIP-2 the most efficient implementation is the one that uses ECIES with the difficulties in deployment being the same with RSA since a PKI is needed to manage certificates for the Proxies. While ECIES presents higher delays on the client side, it also presents lower delays than RSA on the server side; in our case this is a desired effect since the imposed client delays are not perceived by the end user.

6.8 Summary

Before SIP can be utilized in a large scale certain security issues must be solved. While SIP is a simple and easy to deploy protocol, it turns out that some of the security problems related with it are hard to solve. One such problem is privacy since SIP messages cannot be cryptographically protected as a whole. Here, a novel framework has been proposed that can alleviate this privacy issue.

The PrivaSIP framework was presented here and a number of different implementations were evaluated in terms of time delay; also a qualitative comparison with existing schemes was provided. It turns out that PrivaSIP can protect user IDs more effectively in terms of privacy protection than existing schemes and in cases where existing methods fail to satisfy users' privacy needs. This is especially true when a fair balancing between privacy and performance is terminus. The most significant advantage of PrivaSIP framework is that it can assure user ID protection even when SIP messages are transmitted through untrusted SIP domains, while the respective performance results show that this can be achieved with no perceived delay by the end user. The quantitative analysis through testbed experimentation showed that for the client side the delay is negligible, while the cost on SIP Proxies turns out to be highly dependent on the chosen implementation ranging from comparable to standard SIP to quite expensive in terms of time delay. To clear out things, a discussion on which implementations are appropriate under certain circumstances based on efficiency and easiness of deployment has been provided.

Chapter 7 - Conclusions and future work

In the previous chapters the emerging environment of NGNs was described and the privacy issues related with it; it was also shown that when demanding applications are in place, like multimedia delivery, their special characteristics put an extra burden to security designers. The main difficulty here is that security must be preserved but only with some reasonable cost in terms of performance delay, especially when security operations are executed during the time period the mobile node handoffs to a new cell and/or network.

This chapter provides the conclusions drawn from the research conducted during the study for this thesis. It also discusses future work and open issues that were not covered in or spring from this thesis.

7.1 Conclusions

A number of different wireless technologies exist today; while these technologies will probably continue to exist, it is foreseen that they will converge into a common IP-based platform and end users will be able to roam between these different multidomain networks. This transition from closed, proprietary systems like 2G towards the open architecture of the Internet will create several new security related threats and arise trust and privacy issues. The solution to these problems is the reuse of security mechanisms already used in the Internet, appropriately adapted to the new environment. These mechanisms have the advantage of having an open nature in terms of specification design and being extensively tested in practice.

In this new heterogeneous environment the performance of demanding applications like multimedia delivery is a very challenging issue since service disruptions and discontinuations are not acceptable by the end users. The situation becomes even more difficult when security operations are required which can sometimes add significant time delays. This problem necessitated the need for secure handoff optimization schemes which can effectively deal with the abovementioned imposed delays.

A more special case of security in multimedia services over all-IP wireless heterogeneous networks is end users' privacy protection. The reason is that, as it currently stands, end users usually have contract with only one network and service provider. In NGNs where multiple network and service providers will co-exist and offer their services to end users, personal information will follow a path that is untrusted and possibly unknown. This poses serious threats for the users and their sensitive data, let alone the legal aspect of the problem.

The following subjects were examined in this thesis and the conclusions drawn in each are described below:

- **Handoff performance**. Handling mobility in heterogeneous networks is a difficult task especially when multimedia services are delivered to the end user. Chapter 3 reviews such solutions which take security into consideration and try to minimize delays during handoff; this survey showed that there are a large number of very different mechanisms some of which show significant performance improvement. A further remark here is that the methods showing the largest improvement over minimizing handoff time delays are those which use an "overall solution" approach. What this means is that their operation and optimization techniques span across multiple layers and not on one layer, e.g. application or network layer only. For instance there are methods that include DHCP operations taking

place in parallel to security operations on higher layers. Methods that do not follow this approach, present performance improvements only on the specified layer, while all other delays remain intact. Another observation, which led to the next subject of research, is that no secure handoff optimization method protects the privacy of end users; an exception here is the "Optimistic access" method which while operates at the link layer, it was included in the survey with the perspective that it can be applied to other layers as well.

- **Privacy during roaming.** The Next Generation of Networks will span across multiple administrative domains. This creates serious security threats and in chapters 3 and 4 it was shown that privacy in fast, secure handoff methods is a serious and still open issue. In chapter 4 we have proposed two novel schemes for the protection of end users' privacy, each one suitable for different types of applications. The first one takes longer to handoff but poses a small amount of signaling load to the core network, making it more suitable for applications that can tolerate longer delays or for lower priority application traffic. The second one requires the exchange of a larger number of messages but has better performance during handoff, which makes it more useful as a part of an overall solution for seamless handoffs when delivering multimedia services. The privacy in these schemes is protected by not transmitting the real ID of the user to foreign domains and using random temporary NAIs as user IDs. Our solutions are based on the Context Transfer Protocol, but other secure handoff optimization schemes can also be privacy enhanced; this leads to adequate privacy preserving solutions, however, new mechanisms should be designed with privacy, among other security qualities, in mind in order to provide a more complete and rational approach.
- **SIP privacy issues.** Protecting privacy in multimedia services means that higher layers where the actual multimedia signaling occurs should be protected as well. SIP was selected as being the best candidate to comprise the primary multimedia signaling protocol for the Next Generation of Networks. The difficulty of providing privacy in SIP lies in the fact that the encryption of whole SIP messages is not an option. The evaluation of privacy protection in SIP revealed that no extensive research has been done focusing on privacy. Moreover, the existing solutions either provide inadequate protection, or their performance renders them unusable for multimedia delivery, or they cannot be easily deployed in the forthcoming multidomain environment. These observations necessitate the need for further research and proposal of new more adequate and efficient solutions. The key here is the tradeoff between security and performance; although in this study the focus is on multimedia delivery and performance has the first priority, an adequate level of security is required in order to increase the users' trust towards this kind of services and related operators, and envision high market penetration.
- **SIP privacy protection**. The review of SIP privacy protecting methods revealed that existing methods are inadequate under certain circumstances. Privacy protection in SIP for multimedia delivery over all-IP wireless heterogeneous networks needs effective and efficient mechanisms towards the goal of secure and seamless multimedia delivery; the proposed framework, namely PrivaSIP, is one such mechanism. As we showed in chapter 6, PrivaSIP achieves greater privacy levels than existing methods, while at the same time being easier in deployment in existing systems. Moreover, the conducted experiments showed that the performance penalty imposed is low both for the server as well as for mobile clients. For these experiments a number of different configurations were used with different cryptographic algorithms, including the most well known algorithms of each category: AES for symmetric, RSA for asymmetric, and ECIES for elliptic curve cryptography; our proposal is modular enough so that other cryptographic algorithms can be utilized as well. To sum up,

PrivaSIP is a privacy protecting framework for SIP that can satisfy the demanding requirements of forthcoming wireless systems.

7.2 Future work

The convergence of different types of networks into one unified platform creates numerous possibilities for new and existing applications; however, it also creates new threats and security issues. In this thesis we have reviewed, analyzed and proposed possible solutions for issues related with multimedia delivery in all-IP wireless heterogeneous networks and appropriately evaluated their applicability and performance. During this research, new security issues and ideas for improving the proposed solutions came up which are summarized here.

In chapter 3, a review of secure handoff optimization schemes was provided. In this review we have included a method, named “Optimistic access”, which is not purely appropriate for all-IP heterogeneous networks since it operates at the link layer. However, the authors of the proposal agree that it can be adapted to different layers and applications that need an efficient mechanism of re-authentication. Consequently, the adaptation of this mechanism to the forthcoming environment of NGNs would be an interesting and challenging subject of study.

Our proposal for two privacy enhanced security context transfer methods was presented in detail and qualitatively evaluated. The next step here could be a quantitative evaluation based on an experimental testbed or a simulation tool; the second choice seems more promising since it would give the opportunity for large scale experiments.

Another improvement over the proposed methods in this thesis would be the refinement of PrivaSIP. As it has been stated, PrivaSIP protects the ID of end users; an expansion to our framework would be the protection of other private information as well, like IP addresses and domain names. This, however, is a difficult task that needs thorough examination since the availability of IP addresses and domain names to intermediate SIP Proxies, is necessary for the operation of SIP.

As we have already stated, our privacy enhanced security context transfer methods and the PrivaSIP framework can be combined together to form a more complete solution for providing privacy in multimedia delivery over wireless heterogeneous networks. It is obvious, however, that these mechanisms are loosely connected to each other so that each one can also be used individually. A tighter co-operation between these methods together with the inclusion of other time costly operations residing at different layers could provide further improvements towards the goal of secure and uninterrupted roaming among different networks and administrative domains when receiving multimedia services. While this would provide a truly overall solution for Next Generation Networks, it is an open issue which needs further investigation.

ACRONYMS AND ABBREVIATIONS

2G	Second Generation (of mobile systems)
3G	Third Generation (of mobile systems)
4G	Fourth Generation (of mobile systems)
AA	Authentication Agent
AAA	Authentication, Authorization, Accounting
AAAH	Home AAA (Server)
AAAL	Local AAA (Server)
ADSL	Asymmetric Digital Subscriber Line
AES	Advanced Encryption Standard
AN	Access Network
AP	Access Point
AR	Access Router
CA	Configuration Agent
CDMA	Code Division Multiple Access
CHAP	Challenge-Handshake Authentication Protocol
CN	Core Network (in 3G)
CN	Correspondent Node
COPS	Common Open Policy Service
CT	Context Transfer
CTB	Context Transfer Block
CTP	Context Transfer Protocol
EAP	Extensible Authentication Protocol
ECIES	Elliptic Curve Integrated Encryption Scheme
FA	Foreign Agent
FHR	Frequent Handoff Region
GSM	Global System for Mobile communications
GW	Gateway
HA	Home Agent
HD	Home Domain
HMIP	Hierarchical Mobile IP
HTTP	HyperText Transport Protocol
IETF	Internet Engineering Task Force

IMS	IP Multimedia Subsystem
IP	Internet Protocol
IPsec	IP Security
IPTV	Internet Protocol Television
ITU-T	Telecommunication Standardization Sector of the International Telecommunication Union
LAN	Local Area Network
MAN	Metropolitan Area Network
MAP	Mobility-adjusted Authentication Protocol
MCU	Multipoint Control Unit
MEGACO	Media Gateway Control Protocol
MIP	Mobile IP
MN	Mobile Node
MPA	Media – independent Pre - Authentication
NAI	Network Access Identifier
nAR	new Access Router
NAS	Network Access Server
NAT	Network Address Translation
NGN	Next Generation Networks
NGW	New Gateway
nCoA	new Care-of Address
nPoA	new Point of Attachment
OAEP	Optimal Asymmetric Encryption Padding
oCoA	old Care-of Address
OIRPMSA	Optimized Integrated Registration Procedure of Mobile IP and SIP with AAA operations
oPoA	old Point of Attachment
P2P	Peer-to-Peer
PAP	Password Authentication Protocol
pAR	previous Access Router
PGW	Previous Gateway
PKI	Public Key Infrastructure
PSTN	Public Switched Telephone Network
QoS	Quality of Service

RADIUS	Remote Authentication Dial In User Service
RSR	Region-based Shadow Registration
RTP	Real-time Transport Protocol
RTCP	RTP Control Protocol
RTSP	Real-Time Streaming Protocol
S/MIME	Secure / Multipurpose Internet Mail Extensions
SCC	Security Context Controller
SDP	Session Description Protocol
SIP	Session Initiation Protocol
SMS	Short Message Service
SR	Shadow Registration
TLS	Transport Layer Security
UMTS	Universal Mobile Telecommunications System
VoIP	Voice-over-IP
WG	Working Group
WLAN	Wireless LAN
WMAN	Wireless MAN

BIBLIOGRAPHY

[1] Technical Specification Group Services and System Aspects, “IP Multimedia Subsystem (IMS), Stage 2”, V8.6.0, TS 23.228, 3rd Generation Partnership Project, September 2008.

[2] A. Biryukov, A. Shamir, and D. Wagner, “Real Time Cryptanalysis of A5/1 on a PC”, In Proceedings of the *7th international Workshop on Fast Software Encryption*, B. Schneier, Ed. Lecture Notes In Computer Science, vol. 1978, pp. 1-18, 2000, Springer-Verlag, London.

[3] S. Fluhrer, I. Mantin, and A. Shamir, "Weaknesses in the key scheduling algorithm of RC4," Lecture Notes in Computer Science, Vol. 2259, pp.1–24, 2001, Springer-Verlag.

[4] International Telecommunications Union – Telecommunications Standardization Sector (ITU-T), “ITU-T Recommendation G.114, One-way transmission time”, 2003.

[5] G. Karopoulos, G. Kambourakis, and S. Gritzalis, “Survey of Secure Handoff Optimization Schemes for Multimedia Services Over All-IP Wireless Heterogeneous Networks”, *IEEE Communications Surveys and Tutorials*, Vol. 9, No. 3, pp. 18-28, 2007, IEEE Press.

[6] G. Karopoulos, G. Kambourakis, and S. Gritzalis, “Privacy Preserving Context Transfer in All-IP Networks”, *4th International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security (MMM-ACNS-2007)*, CCIS 1, pp. 390-395, 2007 I. Kotenko et al. (Eds.), Springer.

[7] G. Karopoulos, G. Kambourakis, and S. Gritzalis, “Two Privacy Enhanced Context Transfer Schemes”, *3rd ACM International Workshop on QoS and Security for Wireless and Mobile Networks (Q2SWiNet '07),* Chania: Crete, Oct. 2007, ACM Press.

[8] G. Karopoulos, G. Kambourakis, S. Gritzalis, “Privacy Protection in Context Transfer Protocol”, *16th Euromicro International Conference on Parallel, Distributed and Network based Processing (PDP 2008) – Special Session on Security in Networked and Distributed Systems*, D. El Baz, J. Bourgeois, F. Spies (Eds.), pp. 590-596, February 2008, Toulouse, France, IEEE Computer Society Press

[9] G. Karopoulos, G. Kambourakis, S. Gritzalis, “PrivaSIP: Ad-hoc Identity Privacy in SIP”, submitted for publication in *Computer Standards & Interfaces*, Elsevier, manuscript number CSI-D-08-00329.

[10] G. Karopoulos, G. Kambourakis, S. Gritzalis, "Caller Identity Privacy in SIP heterogeneous realms: A practical solution", *3rd Workshop on Multimedia Applications over Wireless Networks (MediaWin 2008) - in conjunction with ISCC 2008 13th IEEE Symposium on Computers and Communications,* A. Zanella et al. (Eds.), July 2008, Marakkech, Morocco, IEEE Computer Society Press

[11] G. Karopoulos, G. Kambourakis, S. Gritzalis, “PrivaSIP: A Framework for Identity Privacy in SIP”, submitted for publication in *Journal of Network and Computer Applications*, Elsevier.

[12] Young Kyun, Kim; Prasad, Ramjee. *4G Roadmap and Emerging Communication Technologies*. Artech House, pp 12-13. ISBN 1-58053-931-9, 2006.

[13] ITU-T, http://www.itu.int/ITU-T/

[14] NGN definition by ITU-T, http://www.itu.int/ITU-T/studygroups/com13/ngn2004/working_definition.html

[15] http://www.webopedia.com/TERM/s/single_signon.html

[16] ETSI TR 101 957 V1.1.1, “Requirements and Architectures for Interworking between HIPERLAN/2 and 3rd Generation Cellular systems”, August 2001.

[17] K. Ahmavaara, H. Haverinen, and R. Pichna, “Interworking Architecture between 3GPP and WLAN Systems”, *IEEE Communications Magazine*, no. 11, pp. 74–81, November 2003.

[18] J. Ala-Laurila, J. Mikkonen, and J. Rinnemaa, “Wireless LAN Access Network Architecture for Mobile Operators”, *IEEE Communications Magazine*, vol. 11, pp. 82–89, November 2001.

[19] International Telecommunications Union, "Recommendation H.323", available on line http://www.itu.int/rec/T-REC-H.323/e

[20] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications”, *RFC 3550*, July 2003.

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

[22] Cuervo, F., Greene, N., Rayhan, A., Huitema, C., Rosen, B. and J. Segers, "Megaco Protocol Version 1.0", *RFC 3015*, November 2000.

[23] M. Handley, V. Jacobson, and C. Perkins, "SDP: Session Description Protocol", *RFC 4566*, July 2006.

[24] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, “Hypertext Transfer Protocol -- HTTP/1.1”, *RFC 2616*, June 1999.

[25] J. Klensin, “Simple Mail Transfer Protocol”, *RFC 2821*, April 2001.

[26] T. Dierks, and E. Rescorla, “The Transport Layer Security (TLS) Protocol, Version 1.2”, *RFC 5246*, August 2008.

[27] C. de Laat, G. Gross, L. Gommans, J. Vollbrecht, and D. Spence, “Generic AAA Architecture”, *RFC 2903*, August 2000.

[28] C. Rigney, S. Willens, A. Rubens, and W. Simpson, “Remote Authentication Dial In User Service (RADIUS)”, *RFC 2865*, June 2000.

[29] P. Calhoun, J. Loughney, E. Guttman, G. Zorn, and J. Arkko, “Diameter Base Protocol”, *RFC 3588*, September 2003.

[30] B. Lloyd, and W. Simpson, “PPP Authentication Protocols”, *RFC 1334*, October 1992.

[31] B. Aboba, L. Blunk, J. Vollbrecht, J. Carlson, and H. Levkowetz, “Extensible Authentication Protocol (EAP)”, *RFC 3748*, June 2004.

[32] C. Rigney, W. Willats, P. Calhoun, “RADIUS Extensions”, *RFC 2869*, June 2000.

[33] B. Aboba, P. Calhoun, “RADIUS (Remote Authentication Dial In User Service) Support For Extensible Authentication Protocol (EAP)”, *RFC 3579*, September 2003.

[34] M. Chiba, G. Dommety, M. Eklund, D. Mitton, B. Aboba, “Dynamic Authorization Extensions to Remote Authentication Dial In User Service (RADIUS)”, *RFC 3576*, July 2003.

[35] C. Rigney, “RADIUS Accounting”, *RFC 2866*, June 2000.

[36] M. Beadles, D. Mitton, “Criteria for Evaluating Network Access Server Protocols”, *RFC 3169*, September 2001.

[37] B. Aboba et. Al., “Criteria for Evaluating AAA Protocols for Network Access”, *RFC 2989*, November 2000.

[38] P. Calhoun, G. Zorn, D. Spence, D. Mitton, “Diameter Network Access Server Application”, *RFC 4005*, August 2005.

[39] P. Calhoun, T. Johansson, C. Perkins, T. Hiller, P. McCann, “Diameter Mobile IPv4 Application”, *RFC 4004,* August 2005.

[40] P. Eronen, T. Hiller, G. Zorn, “Diameter Extensible Authentication Protocol (EAP) Application”, *RFC 4072*, August 2005.

[41] M. Garcia-Martin, M. Belinchon, M. Pallares-Lopez, C. Canales-Valenzuela, K. Tammi, “Diameter Session Initiation Protocol (SIP) Application”, *RFC 4740*, November 2006.

[42] S. Kent, R. Atkinson, “Security Architecture for the Internet Protocol”, *RFC 4301*, December 2005.

[43] Pat R. Calhoun, Stephen Farrell, William Bulley, “Diameter CMS Security Application”, *IETF Internet Draft*, expired September 2002.

[44] D. Durham, J. Boyle, R. Cohen, S. Herzog, R. Rajan, A. Sastry, “The COPS (Common Open Policy Service) Protocol”, *RFC 2748*, January 2000.

[45] ITU-T, “Gateway control protocol: Version 2”, Recommendation H.248, International Telecommunication Union, May 2002.

[46] C. Perkins, “IP Mobility Support for IPv4”, *RFC 3344*, August 2002.

[47] A. Campbell, J. Gomez, C-Y. Wan, Z. Turanyi, and A. Valko, “Cellular IP”, *IETF Internet Draft*, expired April 2000.

[48] Peng Xu, Jian-Xin Liao, Xiao-Ping Wen, Xiao-Min Zhu, "Optimized Integrated Registration Procedure of Mobile IP and SIP with AAA Operations," *20th International Conference on Advanced Information Networking and Applications - Volume 1 (AINA'06)*, pp. 926-931, 2006.

[49] 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.

[50] Dutta, A.; Zhang, T.; Ohba, Y.; Taniuchi, K.; Schulzrinne, H., "MPA assisted optimized proactive handoff scheme", Proceedings of the *Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services 2005 (MobiQuitous 2005)*, pp. 155-165, 17-21 July 2005.

[51] A. Dutta, V. Fajardo, Y. Ohba, K. Taniuchi, and H. Schulzrinne, “A Framework of Media-Independent Pre-Authentication (MPA) for Inter-domain Handover Optimization”, *IETF Internet Draft*, expired January 2008.

[52] A. Dutta, S. Das, D. Famolari, Y. Ohba, K. Taniuchi, V. Fajardo, T. Kodama, and H. Schulzrinne, “Secured Seamless Convergence Across Heterogeneous Access Networks”, *White Paper*, Columbia University, April 2006.

[53] IEEE P802.21/D00.05: Draft IEEE Standard for LAN/MAN: Media Independent Handover Services, January 2006.

[54] T. Kwon, M. Gerla, S. Das, "Mobility Management for VoIP: Mobile IP vs. SIP," *IEEE Wireless Communications Magazine*, Vol. 9, No. 5, pp. 66-75, October 2002.

[55] Sang-Bum Han, Heyi-Sook Suh, Keun-Ho Lee, Chong-Sun Hwang, "Efficient Mobility Management for Multimedia Service in Wireless IP Networks," Proceedings of the *Fourth Annual ACIS International Conference on Computer and Information Science (ICIS'05)*, pp. 447-452, 2005.

[56] S. Pack and Y. Choi, "Fast Inter-AP Handoff using Predictive-Authentication Scheme in a Public Wireless LAN," *Networks 2002 (Joint ICN 2002 and ICWLHN 2002)*, August 2002.

[57] M. Georgiades, N. Akhtar, C. Politis, and R. Tafazolli, “AAA Context Transfer for Seamless and Secure Multimedia Services over All-IP Infrastructures,” *5th European Wireless Conference*, Spain, February 2004.

[58] J. Kempf, “Problem Description: Reasons For Performing Context Transfers Between Nodes in an IP Access Network”, *RFC 3374*, September 2002.

[59] H. Soliman, C. Castelluccia, K. El-Malki, and Ludovic Bellier, “Hierarchical Mobile IPv6 mobility management (HMIPv6)”, *IETF Internet Draft*, draft-ietf-mobileip-hmipv6-08.txt, work in progress, June 2003.

[60] B. Aboba, and D. Simon, “PPP EAP TLS Authentication Protocol”, *RFC 2716*, October 1999.

[61] T. Braun, H. Kim, "Efficient Authentication and Authorization of Mobile Users Based on Peer-to-Peer Network Mechanisms," *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS '05)*, January 2005

[62] T. Aura, M. Roe, "Reducing Reauthentication Delay in Wireless Networks," First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SecureComm 2005), September 2005.

[63] H. Kim, K.G Shin, and W. Dabbous, “Improving Cross-domain Authentication over Wireless Local Area Networks,” *First International Conference on Security and Privacy for Emerging Areas in Communications Networks, (SecureComm 2005)*, pp. 127- 138, Sept. 2005

[64] M. S. Bargh, R. J. Hulsebosch, E. H. Eertink, A. Prasad, H. Wang, and P. Schoo, “Fast authentication methods for handovers between IEEE 802.11 wireless LANs”, in Proceedings of the *2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots (WMASH '04)*, ACM Press, pp. 51-60, 2004.

[65] H. Kim, W. Ameur and H. Afifi, “Toward Efficient Mobile Authentication in Wireless Inter-domain”, in Proceedings of *Workshop on Applications and Services in Wireless Networks (ASWN)*, IEEE, p. 47 – 56, July 2003.

[66] A. Mishra, M. Shin, and W. A. Arbaugh, “Pro-active Key Distribution using Neighbor Graphs”, *IEEE Wireless Communications Magazine*, Feb. 2004.

[67] P. Engelstad, T. Haslestad, and R. Paint, “Authenticated Access for IPv6 Supported Mobility”, in Proceedings of *IEEE Symposium on Computers and Communications (ISCC'2003),* Turkey, vol.1, pp. 569- 575, 2003.

[68] X. Fu, T. Chen, A. Festag, H. Karl, G. Schafer, and C. Fan, “Secure, QoS-enabled Mobility Support for IP-based Networks”, In Proc. *IP Based Cellular Network Conference (IPCN)*, Paris, France, December 2003.

[69] Chen, T., Schafer, G., Wolisz, A., Sortais, M., "A performance study of session state re-establishment schemes in IP-based micro-mobility scenarios," in Proceedings of the *IEEE Computer Society's 12th Annual International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS 2004)*, pp. 159- 166, Oct. 2004.

[70] Hyeyeon Kwon, Kwang-Ryul Jung, Aesoon Park, and Jae-Cheol Ryou, “Consideration of UMTS-WLAN Seamless Handover”, in Proceedings of the *Seventh IEEE International Symposium on Multimedia (ISM '05)*, pp. 649-656, 2005.

[71] M. Garcia-Martin, M. Belinchon, M. Pallares-Lopez, C. Canales, and K. Tammi, “Diameter Session Initiation Protocol (SIP) Application”, *RFC 4740,* November 2006.

[72] D. Forsberg, Y. Ohba, B. Patil, H. Tschofenig, and A. Yegin, “Protocol for Carrying Authentication for Network Access (PANA)”, *RFC 5191*, May 2008.

[73] H. Schulzrinne, E. Wedlund, "Application-layer mobility using SIP," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 4, no 3, pp. 47-57, ACM Press, 2000.

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

[75] http://www-mice.cs.ucl.ac.uk/multimedia/software/rat/

[76] J. Manner, and M. Kojo, "Mobility Related Terminology", *RFC 3753*, June 2004.

[77] Loughney, J., Ed., Nahkjiri, M., Perkins, C., and Koodli, R. "Context Transfer Protocol", *RFC 4067*, July 2005.

[78] Aboba, B., Beadles, M., Arkko, J., and Eronen, P. "The Network Access Identifier", *RFC 4282*, December 2005.

[79] Context Transfer, Handoff Candidate Discovery, and Dormant Mode Host Alerting (seamoby), concluded IETF Working Group, http://www.ietf.org/html.charters/OLD/seamoby-charter.html.

[80] Palekar, A., Simon, D., Salowey, J., Zhou, H., Zorn, G. and Josefsson, S. "Protected EAP Protocol (PEAP) Version 2", *IETF Internet Draft, draft-josefsson-pppext-eap-tls-eap-10*, expired, October 2004.

[81] Funk, P. and Blake-Wilson, S. "EAP Tunneled TLS Authentication Protocol (EAP-TTLS)", *IETF Internet Draft, draft-ietf-pppext-eap-ttls-01*, expired, February 2002.

[82] 3rd Generation Partnership Project (3GPP) Consortium, http://www.3gpp.org

[83] D. Geneiatakis, G. Kambourakis, T. Dagiuklas, C. Lambrinoudakis, S. Gritzalis, "SIP security mechanisms: a state-of-the-art review", in Proceedings of the *Fifth International Network Conference (INC 2005)*, Samos, Greece, July 2005.

[84] D. Geneiatakis, G. Kambourakis, C. Lambrinoudakis, T. Dagiuklas, S. Gritzalis, "A framework for protecting a SIP-based infrastructure against malformed message attacks", *Computer Networks*, Vol. 51, Issue 10, pp. 2580-2593, July 2007, Elsevier.

[85] B. Ramsdell, "Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.1 Message Specification", *RFC 3851*, July 2004.

[86] J. Peterson, "A Privacy Mechanism for the Session Initiation Protocol (SIP)", *RFC 3323*, November 2002.

[87] C. Jennings, J. Peterson, and M. Watson, "Private Extensions to the Session Initiation Protocol (SIP) for Asserted Identity within Trusted Networks", *RFC 3325*, November 2002.

[88] E. Rescorla and N. Modadugu, "Datagram Transport Layer Security", *RFC 4347*, April 2006.

[89] C. Kaufman, "Internet Key Exchange (IKEv2) Protocol", *RFC 4306*, December 2005.

[90] J. Rosenberg, and C. Jennings, "The Session Initiation Protocol (SIP) and Spam", *RFC 5039*, January 2008.

[91] A. Pfitzmann and M. Hansen, "Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management - A Consolidated Proposal for Terminology", Version v0.31, Feb. 15 2008, available at http://dud.inf.tu-dresden.de/Anon_Terminology.shtml

[92] J. Franks, P. Hallam-Baker, J. Hostetler, S. Lawrence, P. Leach, A. Luotonen, and L. Stewart, "HTTP Authentication: Basic and Digest Access Authentication", *RFC 2617*, June 1999.

[93] R. Rivest, A. Shamir, L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems", *Communications of the ACM* 21 (2), pp.120-126, 1978.

[94] M. Abdalla, M. Bellare, and P. Rogaway, "DHAES: An encryption scheme based on the Diffie - Hellman problem". *Submission to IEEE P1363a*, 1998.

[95] J. Daemen and V. Rijmen, *The Design of Rijndael: AES - The Advanced Encryption Standard*. Springer-Verlag, 2002.

[96] M. Bellare, P. Rogaway, "Optimal Asymmetric Encryption - How to encrypt with RSA", Extended abstract in *Advances in Cryptology - Eurocrypt '94* Proceedings, LNCS, Vol. 950, Springer-Verlag, 1995.

[97] ISO 10126:1991, "Banking - Procedures for message encipherment (wholesale) - Part 1: General principles, Part 2: DEA algorithm", International Organization for Standardization, 1991.

[98] SIPp, open source performance testing tool for SIP, available at http://sipp.sourceforge.net

[99] SIP Express Router (SER), free, open source SIP server, available at http://www.iptel.org/ser

[100] SIP softphone, open source, available at http://www.twinklephone.com

[101] MySQL, open source Database, available at http://www.mysql.com

[102] OpenSSL, open source SSL/TLS library, available at http://www.openssl.org

[103] Crypto++, free C++ class library of cryptographic schemes, available at http://www.cryptopp.com