Securing Handover in Wireless IP Networks

Yi Ding

Helsinki November 2009

UNIVERSITY OF HELSINKI
Department of Computer Science

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Laitos — Institution — Department
Faculty of Science	Department of Computer Science

Tekijä — Författare — Author

Yi Ding

Työn nimi — Arbetets titel — Title

Securing Handover in Wireless IP Networks

Oppiaine — Läroämne — Subject

Computer Science

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
M.Sc. Thesis	November 2009	87 p

Tiivistelmä — Referat — Abstract

In wireless and mobile networks, handover is a complex process that involves multiple layers of protocol and security executions. With the growing popularity of real time communication services such as Voice of IP, a great challenge faced by handover nowadays comes from the impact of security implementations that can cause performance degradation especially for mobile devices with limited resources.

Given the existing networks with heterogeneous wireless access technologies, one essential research question that needs be addressed is how to achieve a balance between security and performance during the handover. The variations of security policy and agreement among different services and network vendors make the topic challenging even more, due to the involvement of commercial and social factors.

In order to understand the problems and challenges in this field, we study the properties of handover as well as state of the art security schemes to assist handover in wireless IP networks. Based on our analysis, we define a two-phase model to identify the key procedures of handover security in wireless and mobile networks. Through the model we analyze the performance impact from existing security schemes in terms of handover completion time, throughput, and Quality of Services (QoS). As our endeavor of seeking a balance between handover security and performance, we propose the local administrative domain as a security enhanced localized domain to promote the handover performance. To evaluate the performance improvement in local administrative domain, we implement the security protocols adopted by our proposal in the ns-2 simulation environment and analyze the measurement results based on our simulation test.

ACM Computing Classification System (CCS):
C.2.0 [General]: Security and Protection,
C.2.1 [Network Architecture and Design],
C.2.2 [Network Protocols],
C.4 [Performance of Systems]

Avainsanat — Nyckelord — Keywords
handover security, access authentication, local administrative domain, seamless handover

Säilytyspaikka — Förvaringsställe — Where deposited
Kumpula Science Library, serial number C-2009-36

Muita tietoja — övriga uppgifter — Additional information

Contents

1 Introduction

Wireless and mobile technologies have reshaped our way of communication and information access in global scale. The great success of mobile industry stimulates the public awareness of mobility and promotes the markets of Internet services to enrich our daily lives. Nowadays, it is quite common that on our morning commuting to school or office we start browsing the web for the latest news and weather condition, checking personal emails, and even launching a brief chat with our friends via Voice over IP (VoIP) service [VoI09]. With the support of advanced mobile devices including mobile phones, personal digital assistants (PDAs) and wireless enabled laptops, we can enjoy the Internet access almost at any time, anywhere.

For the future broadband wireless networks such as the fourth generation (4G) networks [3GP08b], data communication will rely mainly on packet switching that is based on the Internet Protocol (IP) [Ins81, DH98]. Multiple wireless technologies such as GSM [GSM08], WiFi [Wi-08], and WiMAX [WiM08] utilize IP to transmit data across the Internet. However, since IP was not originally designed to meet the mobility requirements, there are several problems that need to be addressed before we achieve seamless mobility in the forthcoming IP-based wireless networks.

Among the challenges posed to future IP-based wireless networks, one major issue is to maintain the network connectivity for nomadic users when they migrate from one access network to another in a mobile environment. In the context of wireless mobility, the the procedure of switching access networks is referred to as handover [MK04]. To support continuous connectivity and host mobility, Mobile IP [Per02, JPA04] was proposed as a general protocol to support mobility. Regarding the nomadic case, Mobile IP enables a mobile node (MN) in an IP network to change its point of attachment, redirects data traffic to its current location, and keeps the existing connection uninterrupted.

Although the basic mobility issue can be solved by Mobile IP, the fast development of wireless technologies and real-time applications such as VoIP raise more technical challenges to the mobility management. Due to the nature of wireless communication, the coverage of communication region for each access network is limited. During the movement across neighboring regions, handover will take place to transfer the essential user context and service status. Because of the overhead introduced by handover in terms of control signalling, packet redirection, and authentication processing, data exchange over the existing communication session may suffer from unwanted packet delay or loss. For delay-sensitive real time applications, the lengthy delay between the break of previous connection and the establishment of new one may lead to serious performance degradation. Therefore, how to maintain the transparency of connectivity for users who are using real-time applications has attracted numerous research efforts from both industry and academia.

At the same time, security challenges make the handover procedure even more complicated in a wireless and mobile environment. Unlike fixed networks, wireless networks are more vulnerable to malicious attacks such as eavesdropping and data manipulation due to the vulnerability of wireless communication [BH07]. This nature demands the adoption of security in wireless and mobile networks covering authentication, authorization, and privacy protection. However, given that majority of classical security mechanisms for data communication target mainly at personal computers connected via wired networks, traditional security implementations at current stage are insufficient to guarantee secure communication over wireless connection, nor suitable for mobile devices with resource limitations in terms of computing capacity and energy supply. The complexity grows hand in hand with the ever-increasing heterogeneous networking environment. As mobile users move across domains managed by different operators, the security schemes and policies may vary, as well as the technologies to access the network. Given that users at different locations may demand different levels of security and different applications should also be treated differently according their requirements, a fixed level of security can not work for all kinds of scenarios in such a dynamic environment.

In wireless networks, authentication and authorization are necessary to secure the user mobility. The user credentials need to be verified and validated through authentication. The usage of services allowed for a particular user needs to be identified and confirmed by authorization. Because authentication and authorization may bring overhead and unexpected effects to mobility management, especially during the handover phase, we need to identify and analyze their impact in wireless environment. The results will enhance our understanding of the correlation between existing security mechanisms and different kinds of mobility schemes.

In this thesis, we analyze the handover procedure as well as the state of the art security schemes for handover in the context of mobility management. By identifying the major factors and phases in handover security, we propose the local administrative domain, which is a well defined mobility and security enhanced access network to achieve a balance between the handover security and performance.

To evaluate the performance improvement in local administrative domain, we implement the authentication protocols adopted by our proposal in the ns-2 simulation environment [NS 08], and provide preliminary measurement analysis based on our simulation results.

The rest of the thesis is organized as follows: Section 2 provides an overview on wireless networks, IP mobility support, and standardization development. Section 3 discusses the handover procedure to achieve seamless mobility. In Section 4, we study the state of the art security mechanisms for handover management, followed by the analysis of security impacts on handover. Section 5 covers our proposal of local administrative domain, the protocol implementations in simulation environment, and our analysis of testing results. We conclude our work in Section 6.

2 Mobility in Wireless IP Networks

The future networking environment is envisioned to be a unified network interconnected by an IP infrastructure and supported by heterogeneous access technologies [AXM04]. Following this trend, a fast transition is taking place in current mobile industry, which is marked by the ongoing convergence of IP data networking with the widely deployed telecommunication infrastructure to provide integrated voice, video, and data services to mobile subscribers. As a key feature enabled by wireless communication, mobility exerts enormous impact during this transition towards the integrated wireless IP networks.

To provide an overview of mobility solutions in current wireless IP networks, we cover the following topics in this section, including the organization and architecture of a general wireless network, state of the art mobility solutions, and mobility related standardization development by Institute of Electrical and Electronics Engineers (IEEE), Internet Engineering Task Force (IETF), and 3rd Generation Partnership Project (3GPP).

2.1 Wireless networking

The impact of wireless networking is profound. Triggered by the success of cellular networks, the number of mobile phone users is growing at an unprecedented speed over the last decade with the number of cellular subscribers already surpassing the number of main telephone lines [INT07]. This huge growth stimulates the development of mobile devices and wireless access technologies to support better mobile service experience. Owing to reasonable setup cost and flexible service provision, wireless systems are chosen to supplement or replace fixed line systems in many developing countries where it is too expensive or impossible to set up well connected wired telephony infrastructure [Sch03]. In most developed countries, wireless devices such as mobile phones have become critical tools offering great convenience and assistance to daily lives. In addition, wireless local area networks (WLAN) with their increasing popularity provide good enhancement and alternative to wired networks at many locations including homes, campuses, and companies. New applications such as wireless sensor networks and smart homes are emerging gradually from research prototypes to concrete implementations. By connecting all kinds of wireless devices ranging from smart sensors to personal computers, wireless networks enable us fast and convenient access to the Internet and step by step bring the ideal of ubiquitous computing closer to reality.

Concerning the data communication in wireless networks, information exchange is based on transmitting signals in the form of electromagnetic waves over unbounded media such as atmosphere. The means of data transmission make the wireless communication different from its wired counterpart, in which signals are transmitted along fixed and bounded media such as copper twisted pair or optical fiber. This characteristic of wireless transmission offers a unique advantage of supporting user mobility in that it releases the data communication from stringent physical restriction, such as wires connecting to the network. Take WLAN for example, the access point (AP) in a building, which is a device allowing wireless devices to connect to the wireless network, can enable wireless connectivity for all users within its radio coverage range. Users equipped with wireless devices can then move freely within the confined zone and enjoy easy access to the Internet at anytime, without being bothered to find a wire to access the network whenever they move.

Compared with the first milestone in wireless communication marked by the Guglielmo Marconi's transmission of wireless telegraphs across Atlantic Ocean in 1896 [Sta05], wireless transmission schemes and network organizations have evolved a lot. Nowadays, the existing wireless networks can be mainly divided into two groups as infrastructure based and ad hoc based [Sch03]. Figure 1 highlights the basic architecture of two types and key components in a general wireless environment.

Based on Figure 1, we identify the following components that form a general wireless network: mobile node (MN), wireless link, access point (AP), and network infras-

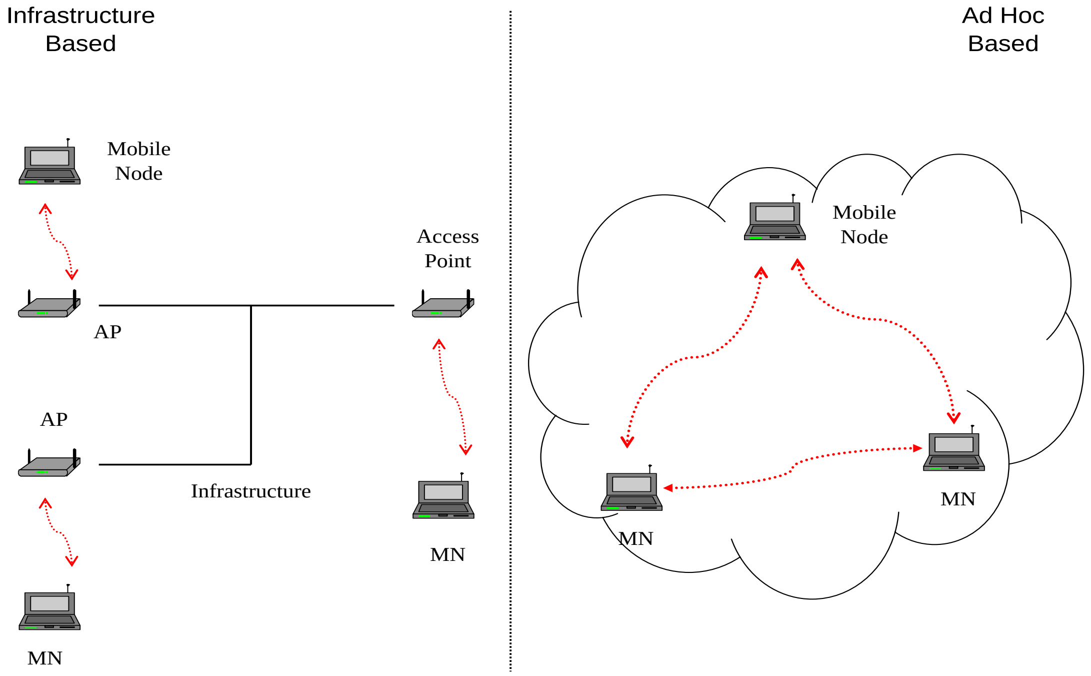


Figure 1: Infrastructure based and ad hoc based wireless networks.

tructure. First, mobile node is a wireless enabled device that lies at the edge of the network and runs software applications. In real practice, a MN could be for example a laptop, a mobile phone, or a PDA. Second, wireless link is the term covering both technology aspects and physical media that connect the MN to the access point. Different wireless link technologies have different transmission rates over different distances. Third, an access point is the coordinate system responsible for sending and receiving data to and from the MNs that associate to the access point. Every AP has its own radio coverage to allow MNs within this range to communicate to it. Cell towers in GSM/GPRS system and access points in IEEE 802.11 WLAN are the examples in these popular wireless systems. Fourth, network infrastructure is the network core that interconnects other systems and networks. Via network infrastructure, mobile nodes can communicate with other hosts belonging to different networks.

For infrastructure based wireless networks, the access point is the key component that carries out most of wireless network functionalities within the infrastructure. It controls wireless media access and acts also as a bridge connecting to other wireless and wired networks. Because the communication complexity lies mainly on AP, mobile node can remain comparably simple in an infrastructure based network. Owing to the infrastructure design, several access points can be combined to form a logical network domain that is connected via the existing infrastructure, so that a number of wireless networks can be connected to form a larger network beyond the actual radio coverage of each AP. At the same time, infrastructure based networks lose some of the flexibility that wireless communication can offer. For example, it is not suitable for emergency and disaster scenarios where no infrastructure is left to setup necessary communication channels. Nowadays, cellular and satellite systems are typical examples that adopt the infrastructure based model [Sch03].

For ad hoc based networks, no infrastructure is needed to interconnect each mobile entity. A mobile node in an ad hoc network can directly communicate with others without the controlling and forwarding support from access point. Data transmission in the ad hoc network is possible only if two hosts are within each other's radio range, or there are other hosts along the path that can forward data to the destination. Compared to infrastructure based networks, the complexity of ad hoc networking lies on the mobile nodes that are required to control media access, handle routing, and deal with wireless specific communication problems such as hidden or exposed terminal problems [Sch03]. As an example, Bluetooth technology which is based on IEEE 802.15 wireless personal area network specification [WPA08] operates in ad hoc manner.

A major difference between infrastructure based and ad hoc based networks is that wireless communication in infrastructure networks takes place only between the mobile node and the access point, not directly between two MNs. While in ad hoc networks, a MN can communicate directly with other hosts in an instantaneous manner without other infrastructure entities involved.

Nowadays both infrastructure and ad hoc models are used in various wireless systems to organize the network. Wireless hosts in those systems are connected from the edge of the network to the larger infrastructure via wireless links. Although wireless links are employed in network infrastructure to connect different routers and systems, most of recent developments are occurring at the edge of networks. Therefore, we focus on the wireless communication at the edges of wireless networks following the infrastructure based model. Ad hoc networking is out the scope of this thesis.

For the forthcoming wireless network environment as proposed in Beyond 3G (B3G) and 4G [3GP08b], there is an increasing need to integrate the existing wireless networks under a unified infrastructure based on IP. Taking into account the comple-

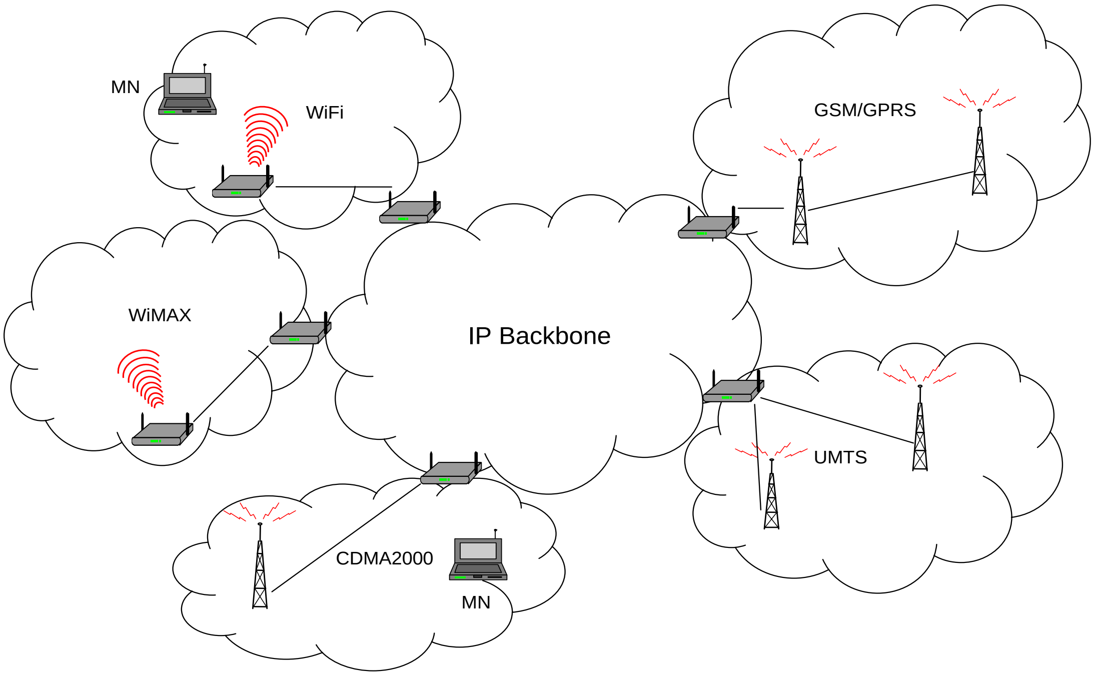


Figure 2: Envisioned architecture for unified wireless IP networks.

mentary characteristics of WLAN and cellular systems with fast short-distance access, and slow long-distance access, respectively, the recent trend in mobile industry is to integrate the existing WLAN networks with other wireless mobile systems including the Mobile Communications/General Packet Radio Service (GSM/GPRS) [GSM08], Universal Mobile Telecommunications System [UMT08], Code Division Multiple Access 2000 [CDM08], and 802.16 [WiM08] under a common IP backbone. As shown in Figure 2 of an envisioned architecture for the future wireless IP based networks, the rational combination of heterogeneous wireless systems under a unified platform will bring us a cost-effective network environment that is capable of providing ubiquitous information access [Sch03]. The current and emerging wireless systems coupled with the potential applications promise us a future of always connected information world.

2.2 Mobility support

Wireless access provides inherent mobility supports, although it is not a prerequisite. To help identify and understand challenges in the mobile wireless networks, we draw a distinction between two key concepts in this thesis: the wireless nature of communication, and the mobile capacity that wireless communication enables. The former term wireless includes the technologies and schemes that users can utilize to access the network and communicate with other parties. The mobile term, also referred as mobility, describes the moving capabilities of devices and individuals.

In current environment, networking nodes can be wireless but not necessarily mobile such as the stationary workstations in an office that has wireless connection. At the same time, mobile devices may require no wireless connection such as a wired laptop that is used at home at first and then carried by its owner to the office. Obviously, the most exciting area belongs to the intersection of wireless and mobile, where users can utilize multiple wireless technologies via mobile devices to obtain all kinds of network services on the move, with always on connectivity at anytime, anywhere.

The urgent need to support mobility in wireless networks brings up multiple proposals to assist mobility management. The IETF Mobile IP [Per02, JPA04] and its enhancements [S^{+}05, J^{+}05] are the most well known mobility management schemes that are studied and deployed. The lately proposed Proxy Mobile IP protocol [LYC08, G^{+}08] provides network-controlled mobility support, in which mobility management functions are carried out by the access network without involving of mobile terminals.

2.2.1 Macro mobility with Mobile IP

Internet Protocol (IP) [Ins81, DH98] is the architectural foundation of the current Internet. As a common base, thousands of applications employ IP to communicate across different types of networks. Bearing in mind the success of Internet owing to the principles of simplicity and scalability, to develop mobility supports at the IP layer can minimize modifications over existing network architecture.

In contrast to mobile telecommunication systems such as GSM, which target at mobile users from the beginning, the original design of IP did not take mobility into account. In IP networks, routing is based on the stationary IP address, which is used as the location identifier for the host. Every IP datagram is forwarded in a hop-by-hop manner from its source to the destination. Regarding the IP routing, it is

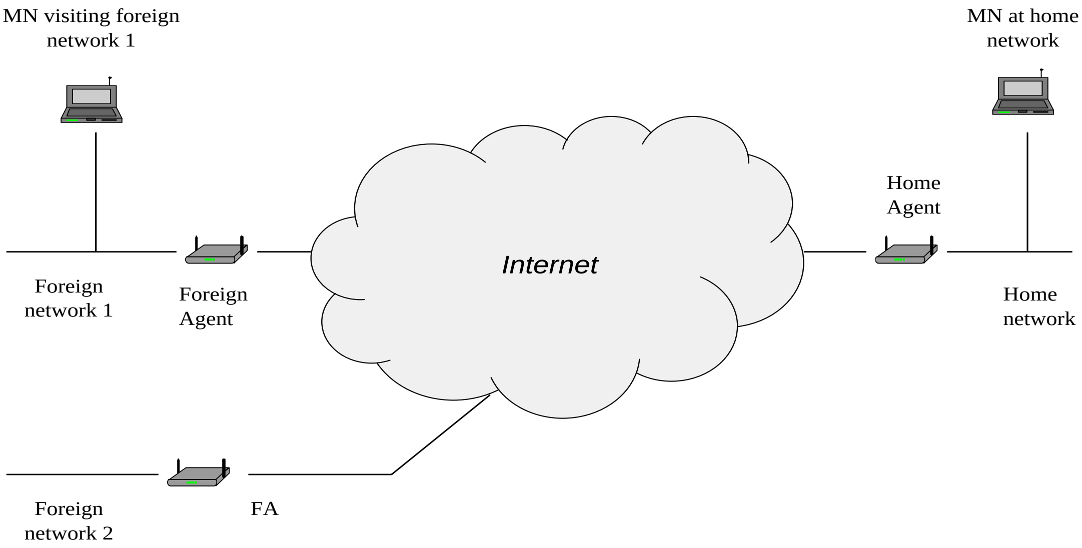


Figure 3: Mobile IP entities and correlations.

assumed that the IP address of a host can uniquely identify its point of attachment. Therefore, each host must be located at the network in accordance with its IP address in order to receive IP datagrams destined to it. Otherwise, datagrams can not reach the destination. This nature of IP routing limits the host mobility. As every time a mobile node moves to another network, the change of IP address could lead to communication interruption because all the datagrams of ongoing session will be forwarded to the home network of the mobile node instead of the current network.

To support mobility in IP networks, the Mobile IP protocol [Per02, JPA04] is proposed with the goal of solving the routing problems during the host's migration from one IP subnet to another. As an open standard, it provides mobility support over a large area, allowing a mobile node to keep its IP address, stay connected, and maintain the ongoing communication sessions while moving between neighboring networks.

As shown in Figure 3, three functional entities are introduced by Mobile IP: mobile node (MN), home agent (HA), and foreign agent (FA) [Per97]. First, MN is an end-system or router that can change its point of attachment to the network using Mobile

IP. MN keeps its home address when the point of attachment changes. As long as the link layer connectivity is provided, MN can continuously communicate with other entities in the network. MN is not necessarily a small device such as mobile phone or laptop, while for example, a router on an aircraft can be an MN. Second, HA is a router residing on MN's home network. It is able to deliver datagrams to MN when MN is away from the home network and maintains the location registry for MN. HA also configures and maintains mobility security association including shared keys. Third, FA is a router on the foreign network that provides mobility services to the MN during its visit. FA is able to forward datagrams coming from HA to the MN currently served by it. Any node in the Internet communicating with an MN is referred to as correspondent node (CN).

Mobile IP employs two IP addresses to enable mobility: a fixed home address (HoA) that serves as unique identity for the MN; a care-of address (CoA) that changes at each new point of attachment in foreign networks. Between the home IP address and MN's current CoA, an association is maintained by the home agent. The CoA is a temporary address for MN in a visiting network in order to route the datagram from HA to MN's current location.

In essence, Mobile IP consists of three functions that cooperate with each other, including the agent discovery, registration, and tunneling. For agent discovery, it is the process of obtaining an IP address for the MN. By periodically receiving agent advertisement messages that are broadcasted from FA, MN is able to detect a new subnet in which it moves. A newly arrived MN can send agent solicitation messages on the link to learn the presence of any prospective FA.

Regarding the MIP registration as depicted in Figure 4, when an MN discovers that it has moved or will move into a foreign network, MN obtains a new CoA from the FA. The care-of address can be obtained by listening agent advertisement or soliciting the FA actively. The MN registers the new CoA with the HA through FA. HA updates the MN's mobility binding by associating the latest care-of address with the home address. A successful Mobile IP registration sets up the routing for transporting datagram to the MN as it moves across different networks.

For tunneling, in order to deliver the datagram to MN when it is outside its home network, Mobile IP uses tunneling and encapsulation to solve the IP routing problems. When a datagram destined to MN arrives at the home work, HA will intercept and encapsulate it. The encapsulated datagram will be tunneled to MN's current CoA. After receiving the datagram, the FA serving the MN will decapsulate the datagram and further forward it to the MN.

When a mobile node moves from one network to another, it needs to change its point of attachment to the access network, which is referred to as handover [MK04]. The handover procedure in terms of network routing will take place in following steps: 1) MN obtains a new CoA when entering a new network; 2) MN registers the new CoA with its HA; 3) HA sets up a new tunnel associated with MN's current point of attachment, and removes the old tunnel associated with the old CoA; 4) Once the tunnel is set up, HA will deliver future datagrams destined to the MN to its new CoA through the tunnel.

In the original design of Mobile IP, when MN replies to the correspondent node with which MN is communicating, the datagram will be routed directly to the destination with the MN's home address as its source address. However, this data path may bring routing problems in that many currently deployed firewalls demand a topologically correct source IP address to deliver datagrams. If the source address of a datagram implies that it is not from the foreign network where MN is visiting, the firewall of the foreign network will drop the datagram instead of forwarding it.

To solve this problem, an extension of Mobile IP referred as reverse tunneling is proposed [Mon01]. As illustrated in Figure 5 on the difference between reverse tunneling and basic Mobile IP in terms of routing path, the scheme in reverse tunneling requires FA to tunnel all the datagrams coming from MN back to HA. Therefore,

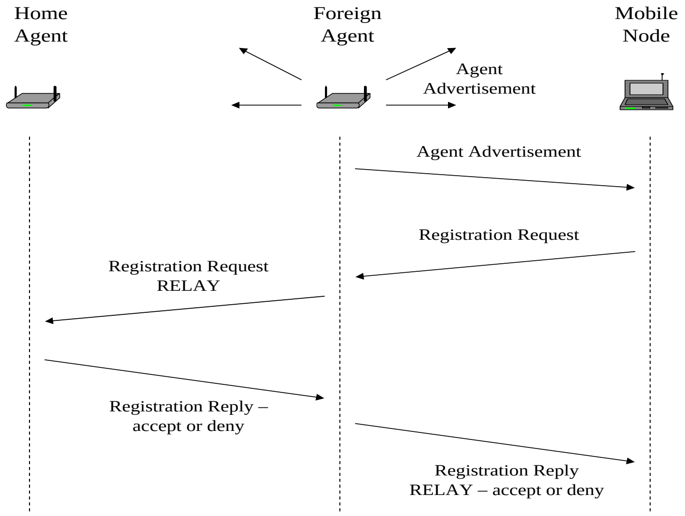

Home Agent
Foreign Agent
Mobile Node
Agent Advertisement
Agent Advertisement
Registration Request
Registration Request RELAY
Registration Reply – accept or deny
Registration Reply RELAY – accept or deny

Figure 4: Agent advertisement and registration in Mobile IP.

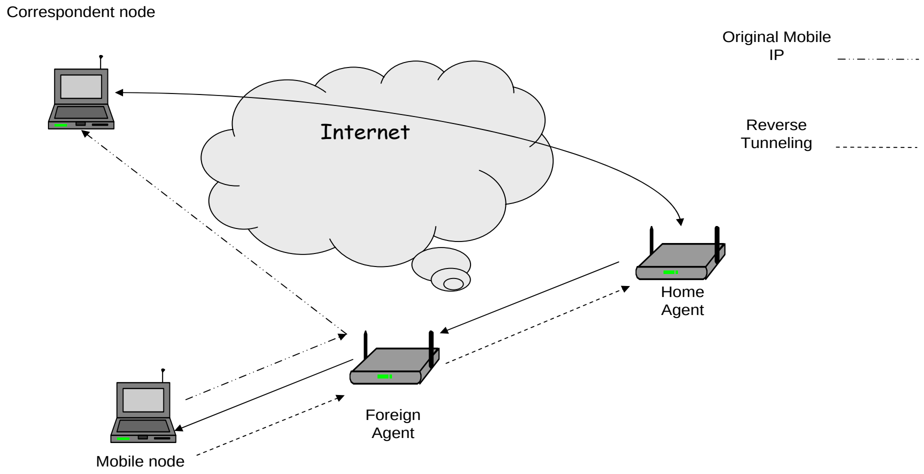


Figure 5: Reverse tunneling for Mobile IP.

the tunnel between FA and HA becomes bi-directional. The original payload from MN will be encapsulated in a datagram with FA's IP as the source address and HA's IP as the destination address. After receiving the datagrams from FA, HA decapsulates them and forwards the original datagram further to the correspondent node. By building a reverse tunnel, the datagram can pass through the firewall at the foreign network.

In general, by introducing tunneling and traffic redirection mechanisms, Mobile IP solves the basic mobility problems in IP networks. The solution is scalable because only the participating components need to be Mobile IP aware. No other routers or hosts with which the MN is communicating need extra modifications. As the mobility mechanism is built at network layer, Mobile IP is independent of link layer technologies. This feature makes it suitable for mobility management across different domains with heterogeneous access technologies.

2.2.2 Micro mobility enhancement

Mobile IP provides a solution to the mobility management in the Internet. However, the protocol is far from perfect. For mobile nodes moving frequently between subnets inside one domain, Mobile IP can generate a significant amount of control traffic across the Internet. The mobility management signalling delay increases as the distance between foreign network and home network increases. For delay-sensitive applications such as real time voice service, the increased delay makes it impossible to maintain the quality of service at an acceptable level when the attachment to the network changes. Moreover, establishing new tunnels to deliver packets to frequent moving nodes leads to additional delays, causing unnecessary packet loss. As the number of mobile users grows, the mobility associated traffic in the core network will affect the data traffic negatively [CKK02]. Due to the problems stated, Mobile IP is unsuitable for mobile entities that move with high frequency across different domains. The challenges of Mobile IP in terms of performance and scalability on the level of global Internet lead to the macro/micro mobility design that divides the mobility administration domain into two parts in terms of geographic coverage.

As illustrated in Figure 6 of macro/micro mobility approach, for macro mobility across wide area networks, Mobile IP is a suitable mechanism to manage the movements of mobile nodes between distant domains. On the other hand, to optimize local handover with less overhead, micro mobility schemes are used to manage the host mobility inside each local domain. Currently multiple proposals exist for micro

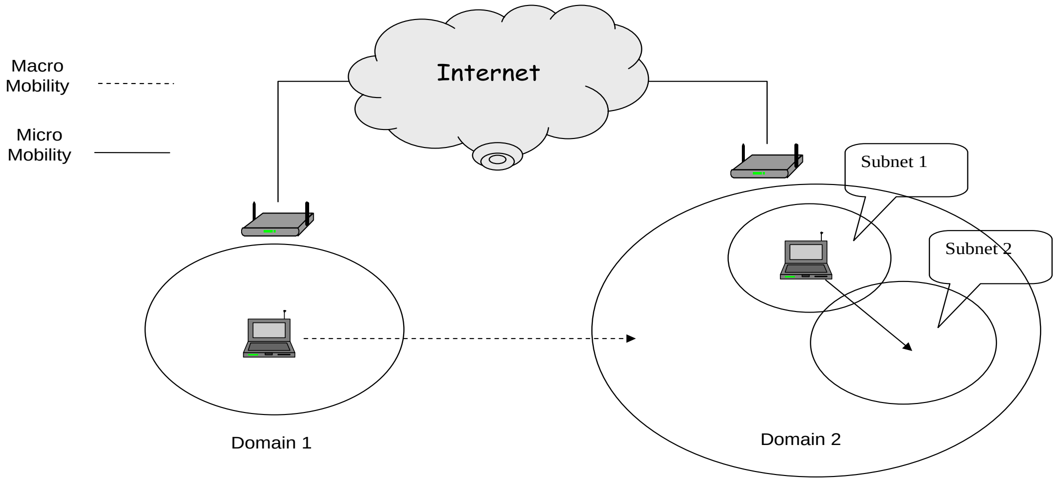


Figure 6: Overview of macro/macro mobility approach.

mobility management that can be broadly classified into two groups: tunnel-based schemes and routing-based schemes [AXM04].

In tunnel-based schemes, local or hierarchical registration and encapsulation are adopted to limit the scope of mobility related signalling. The tunnel-based design rely on a tree-shape structure that is formed by a set of foreign agents. Encapsulated datagram from a home agent is delivered first to the foreign agent which is the root of the tree-shape domain. Each foreign agent on the tree decapsulates and then re-encapsulates the datagram as the traffic is forwarded down the tree of foreign agents towards mobile node current point of attachment. A set of foreign agents in the tree maintain the location database in a distributed form for each visiting mobile node. The location entries in database are created and maintained by registration messages transmitted by mobile nodes. When a mobile node moves between different foreign agents, location updates are conducted at the optimal point of the tree-shape domain.

In routing-based schemes, routers maintain mobile specific routes to forward datagrams. The mobile specific routes are updated with host mobility mechanisms. The routing-based design avoids the overhead from recapsulation and re-encapsulation as used in tunnel-based schemes. The routing-based proposals use routing to forward datagrams toward mobile node's current point of attachment with mobile specific routes. Implicit or explicit signalling is introduced to update the mobile specific routes in routers. To illustrate the design and functionality, we select Hierarchy Mobile IPv6 [S$^+$05] for tunnel-based scheme and Cellular IP [Val99] for routing based scheme. Figure 7 presents the basic architecture of the two proposals.

Hierarchical Mobile IPv6 (HMIPv6) introduces a hierarchical structure for micro mobility management. A mobility anchor point (MAP) is installed in the local network, which is responsible for this domain. MAP acts as a local home agent within the domain for the visiting mobile node. MAP receives all the datagrams from the Internet on behalf of the mobile node. It encapsulates and forwards them directly to mobile node's current address, the link care-of address (LCOA). As long as the mobile node stays inside the domain of MAP, the globally visible address, the regional care-of address (RCOA) does not change. The MAP internal domain boundaries are defined by the access router (AR) which advertises the MAP information to the attached mobile nodes. MAP assists with the local handover and maintains the RCOA and LCOA mapping. Mobile node registers the RCOA with its home agent using a binding update following the same manner as Mobile IP protocol. When a

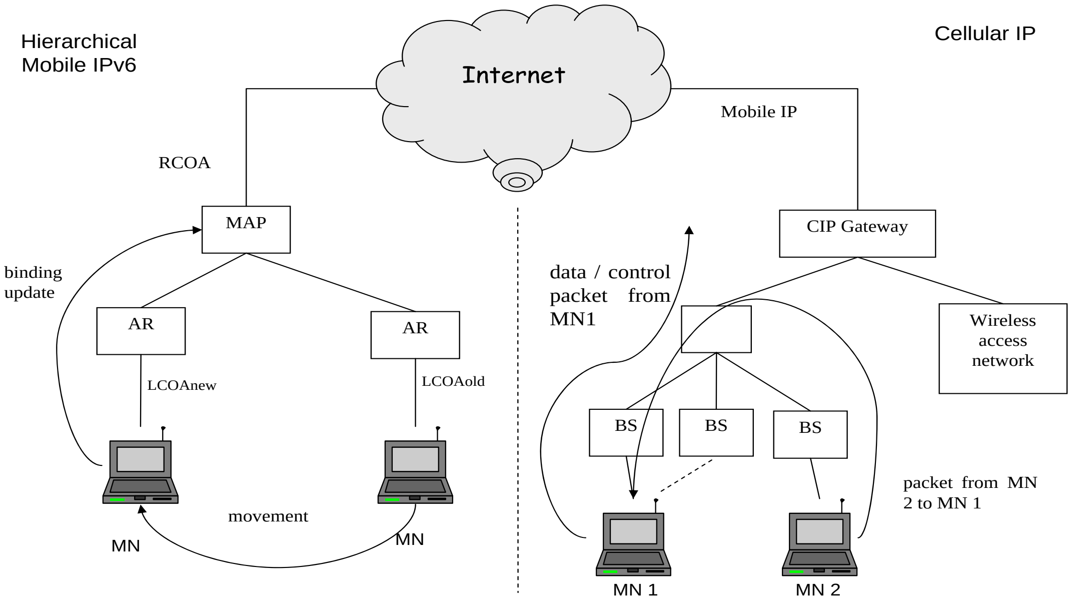


Figure 7: Basic architecture for HMIPv6 and CIP.

mobile node moves locally, it will register its new LCOA only with the MAP in this domain with the RCOA unchanged.

Cellular IP (CIP) provides local handover support without renewed registration procedures. A cellular IP gateway (CIPGW) is installed for each domain which functions as a foreign agent. Inside each CIP domain, all nodes participate in collecting the routing information to access the mobile node. The routing path is built based on the origin of datagrams sent by the mobile node towards the CIPGW. Handover performance is improved by allowing simultaneous forwarding of datagrams destined to the mobile node along multiple paths. A mobile node moving between neighboring cells will be able to receive datagrams via both old and new base stations (BS) if supported by lower protocol layers.

The comparison of various IP layer micro-mobility solutions are discussed in details by F. Chiussi [CKK02] and A. Campbell [C$^+$02] based on different criteria. Although there are differences among micro mobility protocols in terms of routing and control signalling, their operational principle is largely similar [AXM04]. In each proposal, the domain root routers are introduced with the purpose of localizing most of the signalling traffic within the local domain. To avoid global signalling overhead from

Mobile IP, tunnel-based schemes enhance the network scalability by introducing hierarchies, with local registration and update addressed by the gateway router at each hierarchy. The routing-based schemes take advantages of the IP forwarding to maintain the location data base at each intermediate node for every MN in the domain. In general, tunnel-based methods provide the reliability depending on mobility agents at each hierarchy with additional cost and delay. The routing-based methods avoid tunneling overhead, but they suffer from the high cost of propagating mobile specific routes to all routers within the domain.

Given the fast adoption of IP communication in current wireless networks, IP based micro mobility protocols can complement Mobile IP by offering fast and efficient handover control in localized areas. In highly mobile environment where mobile nodes frequently change their point of attachment to the network, micro mobility schemes help alleviate the problems introduced by the basic Mobile IP such as significant network overhead in terms of increased delay, packet loss and signalling. As Mobile IP supports the macro level mobility across IP domains, a rational combination of Mobile IP and micro mobility enhancement protocols can yield a flexible and scalable mobility management framework that supports mobility at the Internet scale.

2.2.3 Network-controlled mobility with Proxy Mobile IP

Network-controlled mobility is a topic gaining popularity in recent IP mobility development. Compared with the host-controlled mobility such as Mobile IP in which mobile node has the primary control of mobility functions, network-controlled mobility provides another solution to manage the mobility in wireless networks.

In network-controlled mobility, network entities are responsible for collecting and measuring network information for mobility management purpose. It is the network side rather than mobile node that handles the mobility management functions such as sending registration and reporting to the home agent the change of point of attachment. To illustrate the principle of network-controlled mobility, we choose the Proxy mobile IP [LYC08, G^{+}08] based on Mobile IP as an example and discuss the benefits of network-controlled mobility.

For mobility management, Proxy Mobile IP (PMIP) defines three terms as depicted in Figure 8: Proxy Mobile IP Domain (PMIP-Domain), Local Mobility Anchor (LMA), and Mobile Access Gateway (MAG). The PMIP-Domain refers to the net-

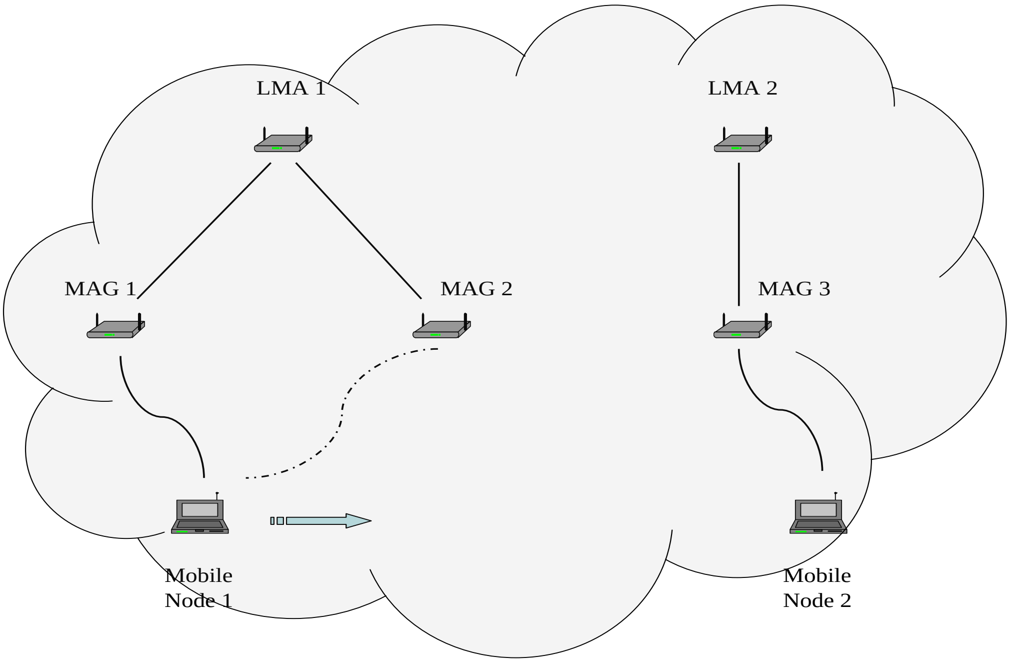


Figure 8: Proxy Mobile IP Domain and entities.

work in which PMIP protocol is used for mobility management. A PMIP-Domain contains a set of LMAs and MAGs, where authorization can be ensured between LMAs and MAGs to send proxy binding updates on behalf of MNs.

Local mobility anchor (LMA) is a router that is responsible for maintaining host routes and forwarding information for mobile nodes under its control. It is the home agent for mobile nodes in a PMIP-Domain. A LMA has the functional capabilities of home agent as defined in MIP protocol specification with the additional capabilities required for PMIP protocol.

Mobile Access Gateway (MAG) is a functional network element on access router in the PMIP-Domain. It manages the mobility signalling for mobile nodes attached to its access link. The MAG tracks the mobile node's IP level mobility between edge links and maintains the bi-directional tunnel together with LMA which is in charge of the mobile node. The MAG terminates the data traffic from LMA to the mobile node which is under its control and forwards data traffic from the mobile node to the corresponding LMA.

As PMIP is based on Mobile IP, its mobility function design follows the same principle as specified in Mobile IP, but mobility related operations are carried out by the network side instead of mobile node. Figure 9 depicts a general procedure for

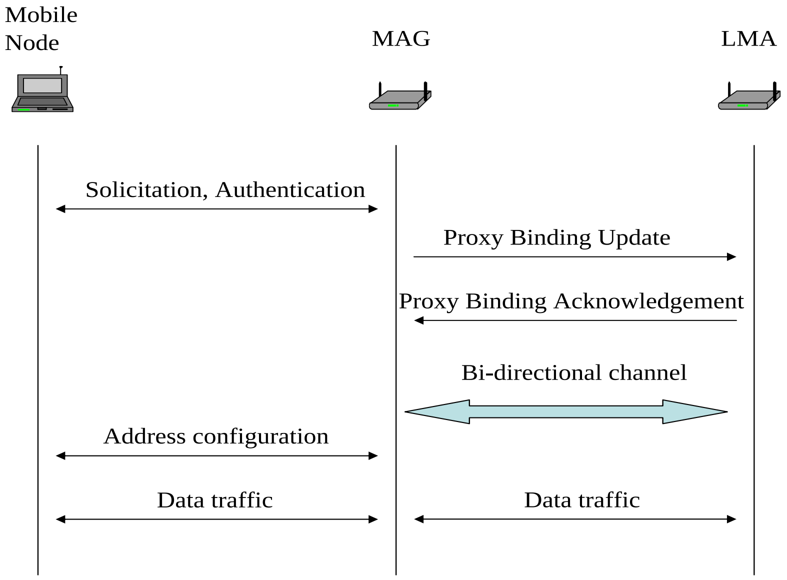


Figure 9: General Proxy Mobile IP network attachment.

network attachment in PMIP.

When a mobile node enters a PMIP-Domain, it first attaches to the access network at link layer and conducts access authentication procedure. If the authentication is successful, mobile node gains the right to access the PMIP-Domain and sends a solicitation message to the MAG. To update the current location of MN, MAG sends a Proxy Binding Update to the LMA which is responsible for the MN. When the Proxy Binding Acknowledgement from LMA is received by MAG, a bi-directional tunnel is established between MAG and LMA for mobile node's data traffic. Following the tunnel set up, mobile node will conduct the address configuration procedure. Once the address configuration is done, mobile node is able to send and receive data from the access network.

For data traffic, LMA is the topological anchor point for all the mobile nodes under its control. LMA receives any datagram sent from other nodes in or outside the PMIP-Domain to mobile nodes of which it is in charge, and appends an outer header on each datagram. LMA then forwards the datagram to MAG through the bi-directional tunnel. On the other end of the tunnel, MAG removes the outer header of the received datagram and forwards it to the mobile node. In PMIP, MAG acts as the default router for mobile node under its control. Any datagram sent from mobile node to correspondent node will be received by MAG first. MAG will forward the datagram to LMA via the bi-directional tunnel. LMA will then route the datagram to the correct destination onwards. Different from the base MIP protocol, PMIP adopts bi-direction tunnel to transfer data between LMA and MAG, which is similar to the reverse tunneling extension in MIP.

As one of the network-controlled IP mobility management protocols, PMIP reuses Mobile IP to gain the maturity of development and deployment in telecommunication industry. This greatly helps drive its development. In recent years, PMIP has become one of the most important IP mobility protocols adopted by various next generation wireless architectures including Mobile WiMAX [WiM06], 3GPP [3GP07a] and 3GPP2 enhancement [3GP06]. Besides PMIP, 3GPP GPRS Tunneling Protocol (GTP) [3GP07b] is another network-controlled mobility scheme in use nowadays.

Regarding the design of network-controlled mobility, one major intention is to reduce signal traffic over the air link between mobile nodes and access points. Different from Mobile IP, the sending of registrations and binding updates is handled by MAG in Proxy Mobile IP, rather than letting the MN exchange signalling messages over its air link between MN and AP. Due to the power supply limit on mobile sets, it is better to handle most of mobility related signalling on network side entities rather than the power stringent mobile nodes. By letting the network entities handle mobility signalling, network-controlled mobility supports efficient energy usage for mobile nodes.

Compared with host-controlled mobility, another benefit of network-controlled mobility is the simplified IP stack and system software implementation on mobile nodes. This is a highly desirable solution concerning the advantage of having interoperable and standardized mobility management protocol scalable to large networks while requiring no host stack involvement. With the network side handling mobility functions, there are less requirements and modifications on each mobile node. Given multiple mobility schemes and wireless access technologies developed by different organizations, it is possible to have conflicting requirements on mobile nodes. Assuming the mobility support from the network side, conflicts can be reduced to minimum on mobile nodes, and technology migration process such as IPv4-to-IPv6 could be transparently achieved without the extensive modifications on mobile nodes' software stack.

For completeness, we point out that mobility solutions in wireless networks can be categorized into different layers according to the OSI reference model [OSI94]. Besides the network layer mobility solutions discussed in this section, proposals have been made on other layers to assist mobility, such as Session Initiation Protocol (SIP) [R$^+$02] on application layer, Host Identity Protocol (HIP) [MN06] between network layer and transport layer, and Stream Control Transmission Protocol (SCTP) [Ste07] on transport layer. As we focus on the network layer solutions, details of proposal on other layers are out the scope of our discussion. It should be noted that IP version plays a key role in mobility protocol design. Discussions can be found in related document [JPA04] of IP version impacts on mobility solution, and their contents are beyond the scope of this thesis.

2.3 Standardization development

Given that mobile networking is possible nowadays, it is still challenging to provide complete service continuity during mobile user's movement across different networks. In order to achieve seamless mobility and connectivity in the coming wireless IP networks, organizations including IEEE, IETF, and 3GPP contribute their efforts to the standardization development. With the purpose of understanding the trend in mobility development, we present an overview of recent and ongoing standardization contributions from different organizations aiming at seamless mobility in both homogeneous and heterogeneous wireless access environments.

I. **IEEE**

The standardization focus of IEEE in mobility management includes the IEEE 802.11 [WLA08], IEEE 802.16 [IEE06], and IEEE 802.21 [MIH09]. Up to now, the link layer mobility enhancements for homogeneous environment are provided by IEEE, and its recent proposal of media independent handover (MIH) targets at the integration of heterogeneous access technologies to allow upper layer protocols to take advantage of information from the underlying link.

- IEEE 802.11 WLAN:

  For IEEE 802.11 devices, the migration from the coverage of one access point (AP) to another includes the detection of loss or degradation of the current connection, determining the new AP to connect to, and establishing link layer connectivity with the new access network. The IEEE 802.11k [IEE08a] amends radio resource measurement schemes to facilitate the decision algorithms for detecting the loss or degradation of an ongoing connection. The fast basic service set (BSS) transition amendment of IEEE 802.11r [IEE08b] optimizes the number of information exchanges to establish authentication between a station (STA) and the target AP. IEEE 802.11r suggests employ IEEE 802.11k schemes to reduce scanning time, and introduces a hierarchical key structure to optimize key forwarding and distribution.

- IEEE 802.16 WiMAX:

  The IEEE 802.16 networks provide centralized broadband wireless access. IEEE 802.16e provides a mobility support to reduce handover delays in its amendment [IEE06].

- IEEE 802.21 MIH:

  The main focus of IEEE 802.21 working group is the media independent handover (MIH). With the final standard approved in January 2009, the goal of IEEE 802.21 is to optimize the handover procedure in mobility management between heterogeneous access technologies. The proposal covers both wired and wireless technologies ranging from the media specifications of IEEE 802 to that of 3GPP.

II. **IETF**

IETF invests considerable efforts in the auxiliary enhancements for seamless mobility among heterogeneous access technologies connected by an IP based network infrastructure. The contributions come mainly from different working groups (WGs) within the IETF. The work can be classified as end-to-end approaches, enhanced Mobile IP schemes, and auxiliary mobility enhancements [E$^+$07].

- End-to-end approaches:

  To solve mobility without or with minimum network support, schemes including host identity protocol (HIP) [MN06], stream control transmission protocol (SCTP) [Ste07], and the well known Mobile IP are proposed by IETF.

- Enhanced Mobile IP schemes:

  The IETF MIPv6 signalling and handoff optimization group (MIPSHOP) proposes two enhancement plans based on the Mobile IP: Fast MIPv6 (FMIPv6) and Hierarchical MIPv6 (HMIPv6). FMIPv6 [Koo08] is intended to decrease packets loss by introducing tunnel between the old and new access router. HMIPv6 reduces signalling overhead by introducing hierarchical structure for mobility signalling. The combination of two schemes is further proposed in F-HMIP [J+05] in order to promote the handover performance in mobility management.

- Auxiliary mobility enhancements:

  IETF working groups (WG) gather lots of efforts to provide mobility supports in five categories: 1) context transfer and candidate access router discovery, contributed by SeaMoby WG [Kem05, L+05b, L+05a]; 2) IP network attachment detection, contributed by DNA WG [CD05]; 3) network-based mobility management, contributed by NETLMM WG, which evolves to Proxy MIPv6 [G+08]; 4) establishing secure association with hierarchical key exchange, contributed by MOBIKE WG and HoKEY WG [Ero06, KT06, C+08, ND08, S+08, HO09, OWZ09]; 5) flow based mobility using multiple interfaces, contributed by MONAMI [M+08]. Technical details and achievements of each WG can be found from the corresponding IETF WG documents.

III. **3GPP**

The 3rd Generation Partnership Project (3GPP) [3GP08a] established in December 1998 is collaboration among groups of telecommunication associations, aiming at making a globally applicable mobile phone system specification. Starting from the original Release 98 published in 1998, 3GPP standardization Releases incorporate a large number of individual standard documents covering radio, core network, and service architecture in 3G systems.

Because cellular networks in 3G system inherently encompass supports for seamless homogeneous handover, 3GPP focuses mainly on the make-before-break handover. As proposed in the 3GPP GPRS Tunneling Protocol (GTP), the mobile nodes in 3GPP networks are responsible for reporting radio conditions and locations to the network and network entities collaborate to control and perform seamless handover.

In the forthcoming standardization Release 8, long term evolution (LTE) and system architecture evolution (SAE) are currently under discussion [3GP08b]. The major updates of Release 8 are expected to cover architecture, protocols and radio access technologies. Besides the goal of increasing radio interface bandwidth for wireless mobile networks, LTE/SAE aims at enabling access of 3GPP networks via multiple non-3GPP access networks such as WiFi and WiMAX together with seamless handover support across different access technologies. As proposed in LTE/SAE, 4th generation networks are going to be an ALL-IP based infrastructure with multiple services built upon the IP backbone.

As future networking environment calls for integrated mobility management techniques that can take advantage of IP communication, the IP mobility proposals from different organizations provide promising solutions to enable mobility management. For completeness, we point out that mobility management in wireless networks contains two elements: location management and handover management [AXM04]. The location management enables wireless systems to track the mobile hosts in the network and provides mechanisms for mobile hosts to update their locations. As an example, paging technique is a typical location management scheme that is widely used in current cellular systems. As our thesis concentrates on handover management in wireless networks, mechanisms for location management are out the scope of our discussion.

3 Handover Management

Handover management is a fundamental concern for wireless systems to support seamless connectivity and mobility. Due to its key role in achieving seamless mobility in the future IP based wireless networks, lots of research contributions go into this area. To understand the motivation and design principle, we provide a general discussion of handover management in the context of wireless mobility. IEEE 802.21 Media Independent Handover (MIH) [MIH09] is chosen to illustrate the latest industrial development catering for the future multi-homing and multi-access heterogeneous environment.

3.1 Handover in wireless mobility

Wireless technologies vary widely in terms of bandwidth, latency, network coverage, and mobility. Nowadays, no single wireless system provides an optimal combination with high bandwidth, low latency, wide area coverage, and strong mobility support. Given the current level of technology development, it is a great challenge to provide users with satisfactory mobile experience in existing wireless networks. The key to this challenge lies in a flexible utilization of the available networks achieved by switching between different wireless access technologies whenever necessary [C$^+$06, SK98].

In the context of wireless mobility, the procedure of switching access networks is referred to as handover. By definition, handover is the process by which an active mobile host changes its point of attachment to the network, or when such a change is attempted [MK04]. To enable host mobility in wireless networks with continuous connectivity, handover management is a fundamental issue that needs to be solved.

To discuss handover management we first introduce a mobile specific term, roaming, which is frequently used in current mobile systems. In the context of wireless mobility, roaming is a particular aspect of host mobility. It is an operator based term involving formal agreements between operators that allows a mobile node to get connectivity from a foreign network [MK04]. Roaming includes the functionalities to help mobile hosts exchange their identities to the foreign access network so that inter-operator agreement can be activated to enable services in the visiting network.

In wireless environment, handover management entails most of the functionalities that are necessary for seamless mobility across different networks while preserving the QoS and migration transparency. In general, the handover process includes a series of signalling and context transfer between mobile nodes and network side to exchange user credentials and network information. As an example, Figure 10 demonstrates the signal flow of a simple handover between two access points in GSM

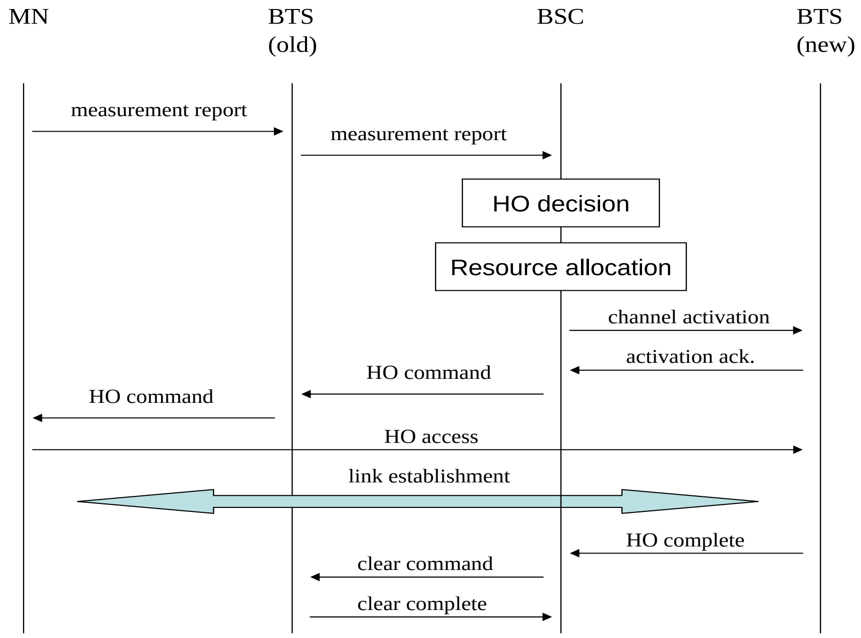


Figure 10: Signal flow of a simple handover in GSM.

system [Rah93].

In GSM system, MN is connected to base transceiver station (BTS), which functions as the access point to the network. Base station controller (BSC) manages the BTS and handles the handover management. When a MN moves into the coverage of a new BTS, it will receive wireless signal from this BTS. MN sends the measurement report to the old BTS, which forwards the report to BSC in charge. Based on the reported value, BSC will decide whether it is necessary to initiate handover. Once the decision is made, BSC will allocate resources for the new channel and send channel activation signal to the new BTS. The new BTS checks whether there are enough resources available, and if resources can be allocated, the new BTS will activate communication channel for the MN and send acknowledgement back to BSC. BSC then issues a handover command which is forwarded to MN via the old BTS. Upon receiving handover command, MN breaks its old link and starts access to the new BTS. The following step is to establish link between MN and new BTS. When the new link is up, handover complete signal will be sent from new BTS to BSC. To release the previously allocated resources, BSC sends clear command to the old BTS. Old BTS will send back a clear complete signal to BSC to indicate the end of a successful handover.

Given the current wireless environment, when mobile nodes move across different networks adopting the same or different access technologies, handover may arise if one of the following conditions is met: 1) When a mobile node is moving out of the coverage of the serving domain and entering a new domain, while the signal strength of previous access point (or base station) falls below a certain threshold value; 2) When a mobile node currently connected to one network chooses to switch to another one for its future service needs; 3) When load balancing is needed to distribute the overall network load among different wireless systems [AXM04].

In general, handover procedure can be classified to different types according to aspects such as control, scope, connectivity, and performance. For control aspect, handovers are considered to fall into one of the two classifications: host-initiated and network-initiated. Concerning the host-initiated handover, the mobile node is responsible to determine its new point of attachment and establish the link connection by following corresponding protocol required by the network side. Handover in Mobile IP is a typical host-initiated process. On the other hand, network-initiated handover let the network side to carry out all necessary tasks such as network measurement and handover decision. Handover adopted in the cellular system is regarded as network-initiated procedure.

Regarding the scope of handover, an important concept is the wireless overlay network, which is referred to as different wireless systems that span from room, campus, metropolitan, and regional size, fitting into a hierarchy of network structure [SK98]. The structure of wireless overlay network is depicted in Figure 11.

In the hierarchy structure of wireless overlay network, lower levels consist of high bandwidth and low latency wireless systems that cover small area, while higher levels are comprised of lower bandwidth and high latency systems that provide wireless coverage over a larger geographic area. Vertical handover, in the overlay network architecture, is a handover between access points (or base stations) that are using different wireless technologies. On the other hand, horizontal handover refers to the switching process between access points (or base stations) that are using the same type of wireless technology. As an example, when a user switches the wireless access from campus area WiMAX to room area WiFi, it is regarded as a vertical handover. When a user moves from one room to another by switching between two WiFi access points, this process is referred to as horizontal handover. Horizontal handover follows the traditional definition of handover in homogeneous

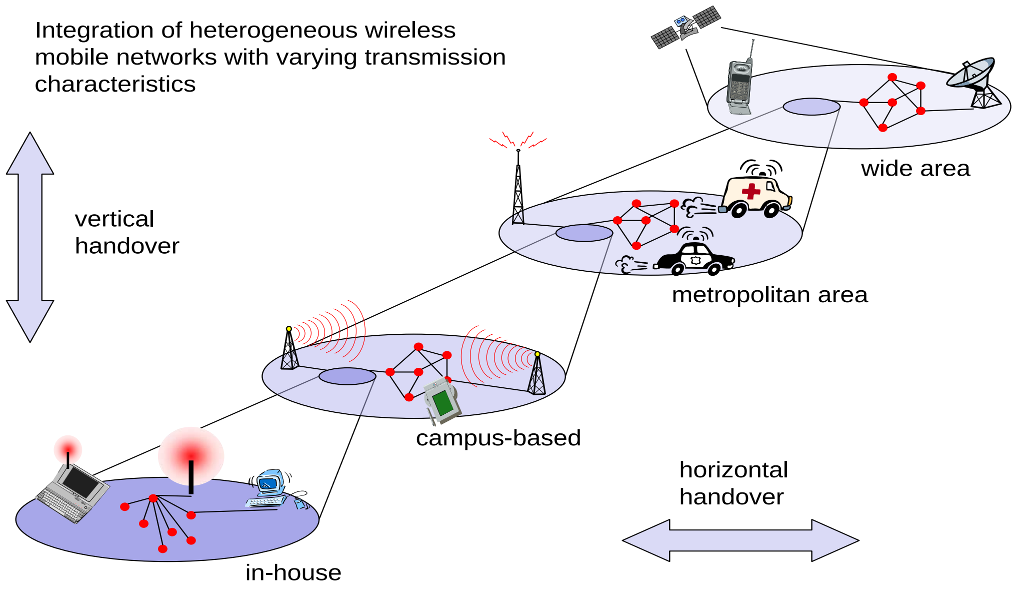


Figure 11: Vertical and horizontal handover in wireless overlay network [SK98].

cellular telecommunication systems [SK98].

Regarding the connectivity aspect, two types of handover are defined: break before make (BBM) and make before break (MBB), which are also referred to as hard handover and soft handover, respectively [MK04]. During a BBM handover, mobile node breaks the previously established connection before it establishes a new connection to the new access network. Therefore mobile node can not maintain simultaneous connectivity with both the old and new access network. Unlike BBM, during a MBB handover mobile node establishes new connection before the old link is taken down. This feature makes the mobile node capable of communicating simultaneously with the old and new access networks. To enable MBB handover, necessary support should be provided by mobile equipments and access networks.

Regarding the handover performance, handover latency and handover loss are two major factors to affect the quality of service (QoS) for mobile services. By definition, handover latency is the difference between the time a mobile node is last able to send and/or receive an IP packet from the previous network, and the time the mobile node is able to send and/or receive an IP packet through the new network. Handover loss is referred to as the packet loss induced by the handover. To promote performance, fast handover is proposed to minimize handover latency without explicit interest in packet loss. At the same time, smooth handover aims at minimizing packet loss without explicit concern of additional delays introduced by handover. Seamless handover is the optimal handover process that is both fast and smooth [MK04].

Currently, the proposed mobility management solutions have ranged from link layer to application layer. As a key part of mobility management, the handover management discussed in this thesis concentrates on link layer and network layer solutions. A recent trend attracting research attentions is the cross-layer approach, which utilizes the information of underlying links to assist the handover process [MIH09]. By using the link layer information such as signal strength report and movement detection, cross-layer approach assists handover preparation in advance, and aims at promoting the overall handover performance.

3.2 Seamless handover

3.2.1 Motivation, challenge, and principle

Owing to the fast growth of mobile Internet markets, the significance of mobility has increased tremendously. As the key to realize seamless mobility in current wireless networks, seamless handover is defined as the attempt to provide persistent connection to mobile users at the given level of quality of service (QoS) during their moving from one access network to another [EN02]. The performance of seamless handover exerts a crucial impact on the overall mobility experience.

For service users, seamless mobility implies that they can exploit the best experience by switching between networks to meet their requirements in terms of cost, QoS, security, and energy consumption. At the same time, service providers can utilize the power of seamless mobility to offer compelling and value-added services to promote their profits as well as the user experience. By allowing users to conduct handover on demand without degrading the performance, network providers can improve their resource utilization and network capacity. The service availability can also be improved by balancing the number of users in different networks. The benefits of seamless mobility are highly desirable and hence motivating the research work on seamless handover.

In general, seamless mobility can be achieved by enabling seamless handover across diverse wireless networks. It requires that the existing communication session should be transferred and resumed in the target network seamlessly. Two performance metrics are regarded as the basic rule for a successfull seamless handover: 1) the handover latency should be no more than a few hundred milliseconds; 2) the QoS of source and target systems should be nearly identical, or the users should not perceive any change according to service experience during and after the handover [LSP08].

Regarding the existing wireless networks, different systems vary from each other in terms of bandwidth, coverage and mobility support. It is complex and challenging to conduct handover seamlessly from one access to another, especially in the forthcoming IP based multi-access heterogeneous environment where vertical handover between different types of access technologies occurs frequently. To enable seamless handover, one major challenge comes from the incompatibility problems across different wireless systems [HY03]. As current wireless systems such as Wi-Fi, WiMAX, GSM, and UMTS, are developed by seperate companies or organizations, the link access technologies, communication protocols, as well as the security policies vary widely from each other. The policy and security implementation in wireless systems demand serious consideration due to the involvement of social and business complexity. To conduct handover seamlessly in such environment, a uniform mechanism based on cross-system design is necessary to enable interoperability.

At the same time, due to the boom of real time IP applications such as voice of IP

(VoIP), the transmission latency requirement catering to the delay-sensitive applications exert a huge impact. As recommended by ITU-T, 50 ms is the threashold value for one way delay regarding most of real time voice and video services [INT08]. This requirement is not easy to meet in that the variation of handover latency could come from multiple sources including error induced retransmission, link re-establishment, access authentication, data base access, and as well as mobility related negotiation procedures. As an example, the vertical handover occurring between two wireless networks with different access technologies can be regarded as a demanding process, as it requires support from mobile devices, mobility protocols, and the network infrastructure. Multi-mode enabled device is necessary for the access to different systems such as Wi-Fi and UMTS. Network should provide adequate coordination to manage the handover between two systems. In general case, cross-platform agreement is necessary to enable authentication and authorization for vertical handover across different networks.

To enjoy mobile services freely powered by seamless mobility in the next generation networks, all the challenges stated for seamless handover need to be addressed. As the research on seamless handover has been going on for several years, previous work renders a good basis for our further research, which can be concluded as five principles of seamless handover [LSP08]:

1) Multi-criteria handover decision - the handover decision should take into account network measurement as well as other handover related information, including QoS information, radio resource availability, link layer triggers, and user preference; 2) Admission control and resource reservation - the resource reservation and admission control at the target network should be made in advance to minimize the handover latency; 3) Context transfer - during the handover preparation phase, critical procedures such as authentication in new network should be assisted by sending security context and QoS context to the target network; 4) Extra information on new connection - source network should provide specific configuration information about the target network, such as available radio channels to mobile terminals; 5) Unified information representation - a unified and unambiguous way of exchanging and interpreting measurement report QoS context should be provided, especially for inter-system handover.

The five principles cover the essential requirements for seamless handover in wireless environment. They provide guidelines for design and implementation and can be used as metrics to evaluate new proposals for seamless handover.

3.2.2 Generic approach for seamless handover

Due to its complexity, seamless handover demands a thorough analysis to be fully understood. Based on previous research work [EN02], a general architectural framework for seamless handover is defined to characterize fundamental elements involved. To enhance the understanding of seamless handover, we discuss this framework together with all the necessary components covering major concepts, an overview of handover procedure, and general approaches to reach seamless handover.

In the mobile environment, a key concept is migration, which occurs when the mobile element (ME) such as a mobile host moves in the network crossing different domains. The source domain (Dom-S) and target domain (Dom-T) are both capable of offering network access for the MEs. The domain representative (Dom-Rep) is serving the MEs that are currently located in the domain. During a migration, the Dom-Reps in Dom-S and Dom-T must execute coordinated actions for a ME leaving Dom-S and entering Dom-T. The coordinated actions are referred to as handover protocol.

Regarding wireless networking, a domain is the region with wireless coverage such as cell in traditional telecommunication systems. The Dom-Rep is the representative of ME to the network side, whose function is similar to the access point in WLAN networks. Dom-Rep offers various services to the ME including delivering network events and data to the ME (down stream), as well as from ME to the network (up stream). Intuitively, a seamless handover indicates that the switching of services from Dom-S to Dom-T should not result in any distinction as to the service received by ME from a single Dom-Rep inside each domain. The transparency of migration is however, highly dependent on services being provided, since each service has a unique set of properties that need to be satisfied during the whole migration. In specific case as to an error-prone unreliable environment, seamless handover also means that the delivery of events should follow the same casual order as if no handover has occurred. This demands the handover protocol to ensure the casual relationship of events in both Dom-S and Dom-T, regardless of the interruption due to migration.

Taking into account the nature of wireless systems in terms of coverage, we assume that both the source and target domains are available during the migration for MEs to connect. The wireless coverage region of two networks can be regarded as overlapped at the place where handover is conducted. To simplify the analysis, our discussion of seamless handover is also based on this assumption.

To conduct handover, Dom-Reps in both Dom-S and Dom-T need to cooperate to resume necessary conditions for ME to be served in the target network. Although the exact procedures depend largely on the specification of each system, the general tasks for handover can be summarized as follows: 1) link access establishment - ME establishes its link connection with new domain Dom-T after fulfilling the necessary authentication and authorization; 2) registration management - either network or ME registers ME's entry at Dom-T and de-register at Dom-S; 3) resource management - network allocates new resources for ME at Dom-T and de-allocate redundent resources at Dom-S; 4) status update - either network or ME updates network-resident state such as route path and address registry at some nodes and/or transfer it from Dom-S to Dom-T; 5) QoS guarantee - to ensure that the functional or non-functional properties of service remain the same during and after handover [EN02].

From a generic viewpoint, main events involved in migration include migration initiation (MI), handover initiation (HI), handover completion (HC), and migration completion (MC) according to timeline. Figure 12 depicts a time-space diagram with stated notations for ME. For the sake of simplicity, only the mobile-initiated handover is discussed since the network-initiated handover follows the same principle. The figure presents an overview of handover events.

Migration initiation (MI) represents that ME initiates the migration, generally with a request to start the handover protocol. Before MI, ME is serviced by Dom-S. Handover initiation (HI) represents that either Dom-S or Dom-T receives the request and starts the handover protocol. Handover completion (HC) indicates the handover protocol in the network is completed with ME's network-resident state being updated. Migration completion (MC) represents the ME's migration to Dom-T is done. At the point of MC, all migration-related events are delivered successfully and ME is serviced by Dom-T.

Concerning seamless handover, two additional events are identified and depicted in Figure 12, referred as old domain exit (oDom-Exit) and new domain entry (nDom-Entry). oDom-Exit represents ME loses access to the network Dom-S. On the other hand, nDom-Entry represents ME gains access to the network Dom-T. Both two events are global events that are observable both at ME and at the network side.

According the events shown in Figure 12, ME is serviced by Dom-S before MI, and after MC, ME should be guaranteed of service from Dom-T. During the interval between MI and MC, ME may be serviced by either Dom-S, Dom-T, or both of them. The condition of network access for ME is marked by oDom-Exit and nDom-Entry.

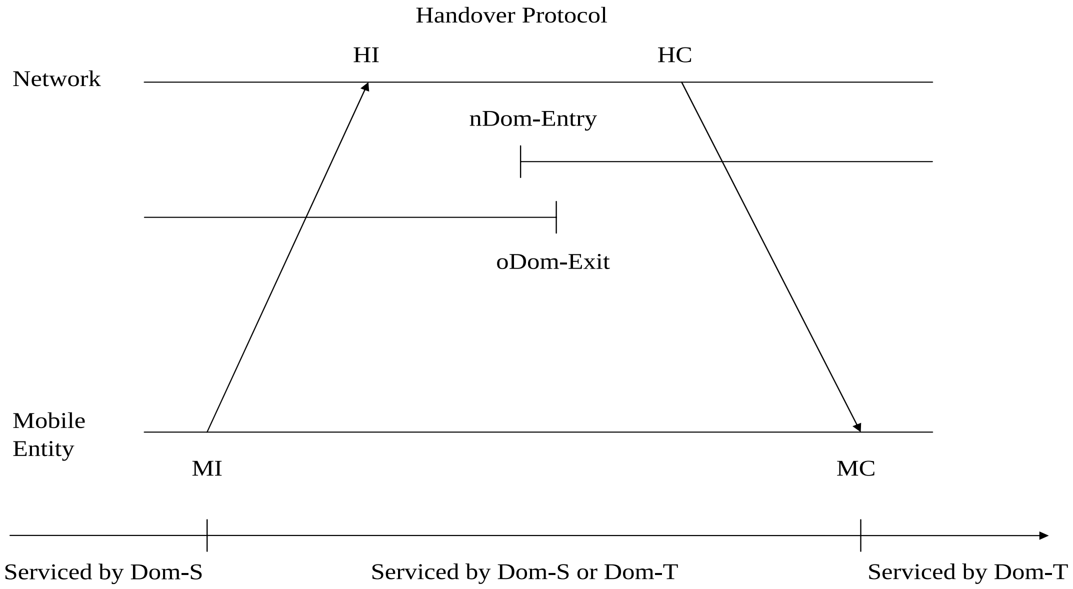


Figure 12: Major events of a handover.

In Figure 12, nDom-Entry occurs before oDom-Exit, which reflects one simple case. In practise, nDom-Entry can occur at any time except after MC, and oDom-Exit can occur at any time except before MI. The events that are not depicted in the figure include the application specific ones occurring in the network which are to be delivered to the ME, and the ones occurring at the ME which are to be delivered to the network. The address assignment conducted at network side is implicit in the figure, as well as the registration request initiated at ME.

Seamless handover, according to Figure 12, is thus concerned with how to provide service in the migration interval between MI and MC. The major goal is to hide from the application/user any difference between the service received during handover and the service received inside a domain. Following our assumption that ME can be served by either Dom-S or Dom-T during the migration interval, the following general approaches for seamless handover are identified: Non-Coordinated Redundant Service, Coordinated Redundant Service, New Domain Service, Old Domain Service [EN02].

For Non-Coordinated Redundant Service, ME is serviced by both Dom-S and Dom-T. The Dom-Reps in both source and target networks should be accessible during the migration interval to make this approach feasible. As network events are delivered by Dom-S and Dom-T, it requires ME to be able to distinguish and discard duplicated event deliveries. At the same time, events from ME sent to both Dom-S and Dom-T need to be properly handled to filter out redundant ones. The Non-Coordinated Redundant Service approach requires necessary system support such as allowing redundant accesses to Dom-S and Dom-T between MI and MC. The network device on ME should support multi-homing feature which allows the ME to send and receive events to and from both Dom-S and Dom-T. This approach is mainly for applications that require uninterrupted service such as real-time applications. One obvious disadvantage is about the resource usage which limits this approach for scenarios where the amount of resources is not a major concern.

Compared with Non-Coordinated Redundant Service, the Coordinated Redundant Service approach avoids the usage of redundant resource during entire migration interval by making Dom-S and Dom-T agree on a switching point, when the service provision is switched from Dom-S to Dom-T, even though both sides can still provide services to ME. The crucial part of this approach lies in the synchronization of Dom-S and Dom-T to update ME's network state. This approach uses fewer resources such as wireless communication bandwidth, and needs no detection or filtering of duplicated events, but at the same time introduces new challenge coming from the synchronization. To make synchronization possible, both domains need to coordinate by following a predefined protocol.

The New Domain Service approach achieves seamless handover by requiring the Dom-Rep of Dom-T to be the sole service point for ME during migration. This indicates that nDom-Entry happens before the MI. As some events may be delivered to Dom-S before the occurrence of HC, in order to ensure correctness of delivery, the Dom-Rep in Dom-S should forward those events to Dom-T. Accordingly, Dom-Rep in Dom-T is responsible for guranteeing the correct delivery of events to ME. The Dom-Rep in new network should merge the stream of events that come from Dom-S with the ones going directly to it. The main advantage of this approach is the simplicity in implementation since the task of merging events is carried out by a single element in network. One disadvantage is due to the extra usage of resources, since for a certain period of time Dom-Reps in both Dom-S and Dom-T are providing services to ME. Another disadvantage is that the event forwarding from Dom-S to Dom-T will prolong the handover period which may affect handover performance.

The Old Domain Service approach allows the Dom-Rep of Dom-S to serve ME during the migration interval, which assumes the oDom-Exit happens after HC. The Dom-Rep of Dom-S determines the occurrence of HC as well as the time when Dom-S stops serving ME and the time when Dom-T should provide service to ME. Essentially, this approach shares the same advantage and disadvantages as the New Domain Service approach. The difference is that in Old Domain Service, the service switching time is postponded. This feature makes it useful for the systems where the handover process or the resource allocation at Dom-T comsumes more time or it is difficult to forward a huge amount of data from Dom-S to Dom-T.

Each of the general approaches discussed is applicable for the scenarios with different requirements. Due to the complexity of seamless handover, there are lots more to be considered regarding to the deployment in the future wireless IP networks. Those general approaches hence provide a high level view over the key events of seamless handover, which help shed light on the implementation in real practice.

3.3 IEEE 802.21 Media Independent Handover

3.3.1 Overview

The performance of current mobile applications depends largely on the capacity of underlying access technologies. Modern wireless systems usually satisfy the service requirements within each system by guaranteeing the level of QoS regarding to transmission delay, loss, and bandwidth [LSP08]. Meanwhile, due to the growth of multi-interface devices and the availability of multiple wireless broadband access technologies, how to enable mobile users to roam seamlessly and securely across different networks has become a key issue for currently incumbent service and network providers. A major challenge to be addressed concerns how to ensure the performance of inter-technology handover in terms of latency and loss. As the operational business requirements vary towards different systems, handover solutions should provide both service providers and network providers the flexibility to implement policies according to their needs. In addition, by taking into account costs, backward compatibility, and competing business interests, the current wireless networking environment will remain diverse for the foreseeable future with multiple access technologies coexisting and cooperating with each other. Supporting seamless roaming is a key to help operators manage and thrive from the heterogeneity [T^+09]. Therefore, a standardized solution is in strong need, which can facilitate seamless handover with the capacity to cope with various mobility management mechanisms in such a heterogeneous multi-access environment.

The IEEE 802.21 Media Independent Handover (MIH) is an industrial driven endeavor to achieve seamless handover across heterogeneous access networks with its specification standard published in January 2009 [MIH09]. The IEEE 802.21 standard provides a media independent framework with associated services to enable inter-technology handover between IEEE 802 family networks (e.g., IEEE 802.16) and non-IEEE 802 networks (e.g., 3GPP networks). Its goal is to improve handover interoperability through the MIH services coordinated under a unified platform.

The motivation behind MIH is coupled with the benefit of standardization and media independency [T$^+$09]. As to improve handover interoperability across different technologies, a common method is to create multiple media specific extensions supporting interoperation. Considering two technologies T1 and T2, to allow T1 interoperate with T2, one extension is needed. The same rule applies to the extension of T2 to T1. Similarily, T1 needs one more extension if it tends to work with T3, and T3 needs one to work with T1. Following this fashion, to ensure N kinds of technologies to interoperate properly with each other, N × (N - 1) media specific extensions are needed. The complexity hence grows on the order of N^2. Obviously, this approach is expensive to implement and does not scale well for environment with multiple access technologies.

To promote efficiency and scalability, IEEE 802.21 standard defines a common media independent framework as a unified platform for handover. This approach allows each access technology only to implement a single extension to interoperate with other ones through the MIH platform. The complexity is on the order of N and scales better than the media specific approach. A set of MIH services is defined by IEEE 802.21 to interact with higher layers of protocol stack. Each technology thus implements one extension to ensure the interoperation with the IEEE 802.21 framework.

Concerning the scope of MIH, IEEE 802.21 defines a series of functions to facilitate handover initiation and preparation through exchanging information, events, and commands. IEEE 802.21 does not attempt to standardize the handover execution mechanism, since link layer mobility is mainly handled by access specific procedures, and traffic re-routing is usually performed by higher layer specifically defined protocols. Therefore, IEEE 802.21 framework is equally applicable to systems that employ Mobile IP at network layer and as to systems that adopt Session Initiation Protocol (SIP) at application layer.

The heart of IEEE 802.21 framework is the Media Independen Handover Function (MIHF). Figure 13 depicts the general architecture of MIH services and their relations with other protocol stacks. As a middle layer between higher layer mobility management protocol stack and lower layer network access protocol stack, MIHF offers a unified interface to upper layers with the exposed service primitives independent of specific technologies and protocols. At the same time, MIHF obtains services from lower layers through a set of media-dependent interfaces. The media-dependent interfaces are specified in the standard according to the respective access technologies including IEEE 802.3, IEEE 802.11, IEEE 802.16, 3GPP, and 3GPP2.

Given the link layer intelligence offered, IEEE 802.21 standard is unique among the set of IEEE standards, as the framework not only allows the interworking within IEEE 802 systems, but the interworking between IEEE 802 and non-IEEE 802 systems, with handover supports covering both wireless and wired systems.

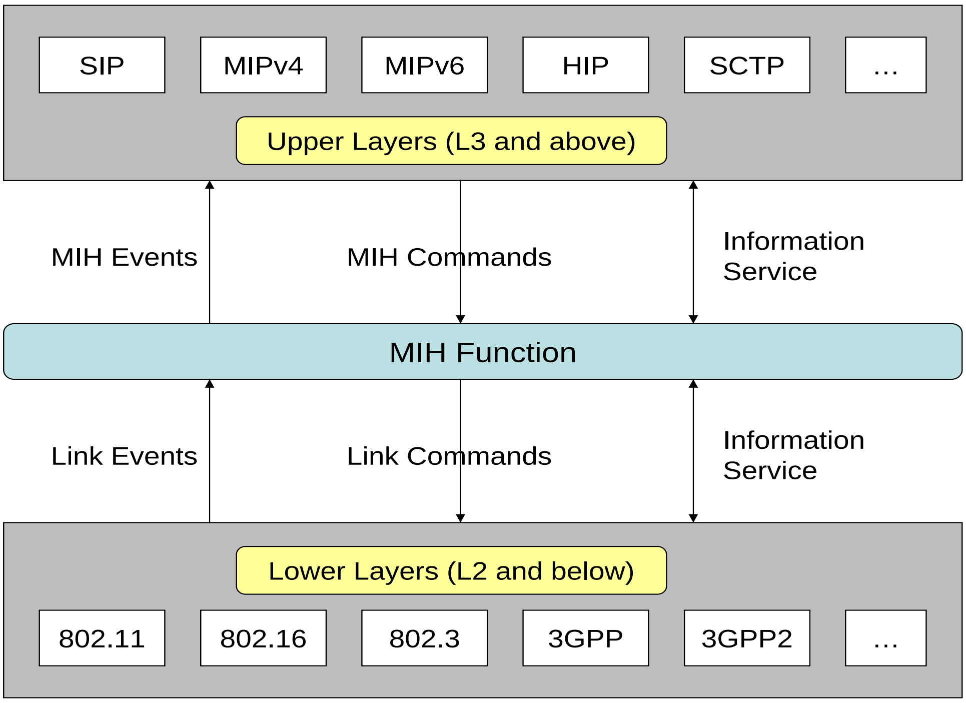


Figure 13: MIH Function in IEEE 802.21 framework.

3.3.2 Framework and services

As stated in IEEE 802.21 standard, seamless handover is a handover associated with a link switch between points of attachment, where the mobile node either experiences no degradation in service quality, security, and capabilities, or experiences some degradation in service parameters that is mutually acceptable to the mobile subscriber and to the network that serves the newly connected interface [MIH09]. To achieve this goal in current multi-access environment, IEEE 802.21 provides a series of services that can support cooperative use of information available at the mobile node and within the network infrastructure. Essentially, the IEEE 802.21 standard consists of following elements: MIH users, MIH function (MIHF), and Service access points (SAPs). Figure 14 illustrates a general reference model defined by IEEE 802.21.

According to the standard, a MIH user is a functional entity that employs MIH services. It is regarded as an abstraction of consumer of MIH services. A typical example of MIH user is a mobility management application that can use the MIH services to optimize handovers.

The MIHF is a logical entity that includes three types of services: 1) media-independent event service (MIES) which detects property changes in link layer and

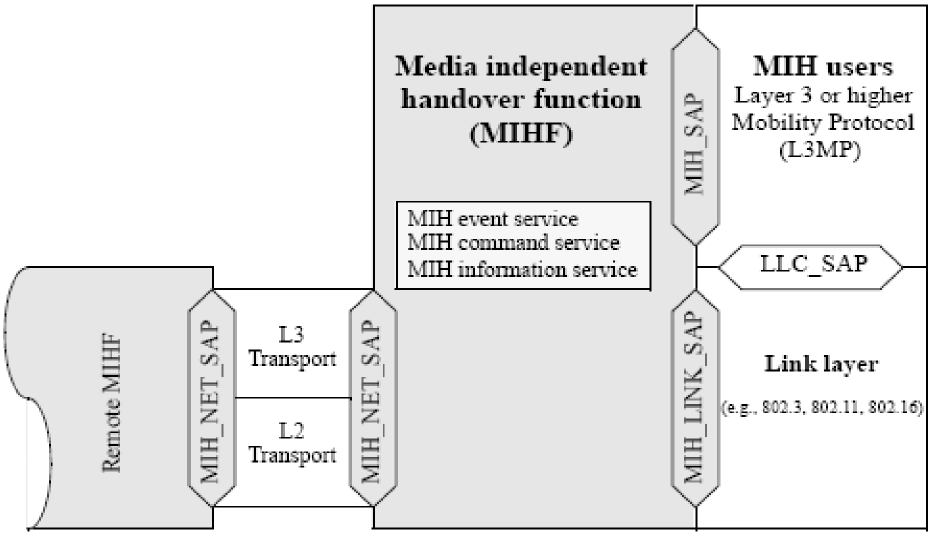


Figure 14: General reference model of IEEE 802.21 [MIH09].

reports events that originate from local and remote interfaces; 2) media-independent command service (MICS) which provides a set of commands for MIH users to control link state; 3) media-independent information service (MIIS) which provides MIH related information about the surrounding networks such as their locations and properties. MIH services can be either local or remote with local operation conducted within local protocol stack while remote operation occurring between two MIHF entities.

The MIH SAPs define media-independent and media-dependent interfaces in terms of primitives according to IEEE 802.21 specification. The standard does not mandate any specific programming language to represent the primitives. The implementers of MIHF are required to define specific APIs with their chosen programming languages. In general, MIH SAPs include the following: 1) MIH_SAP, which is a media-independent SAP providing a unified interface to higher layers to control and monitor different links regardless of specific technology; 2) MIH_LINK_SAP, which is a media-dependent SAP providing an interface to MIHF to control and monitor a specific link; 3) MIH_NET_SAP, which is a media-dependent SAP providing transport services on the local node and supporting the message exchanging with remote MIHF.

To facilitate handover initialization and preparation, IEEE 802.21 MIH relys on three major services: MIES, MICS, and MIIS. To control and configure these three type of services, IEEE 802.21 standard defines a management service, which consists of MIH capability discovery, MIH registration, and MIH event subscription. By using the primitives of management service, MIHF is capable of discovering remote MIHF entities, and obtain services from remote entities. The MIH SAPs function as interfaces to connect different entities. Through a unified SAP to higher layers, MIH enables upper layer applications to have a common view across different access links. Media-dependent SAPs (LINK_SAPs) facilitate MIHF to obtain media specific information that can be propagated to MIH users via a uniform media-independent interface (MIH_SAP). IEEE 802.21 defines a media independent services framework for deployment concern together with the MIH protocol. The MIH protocol enables the communication between peer MIHF entities via the delievery of MIH protocol messages. MIH defines the message formats including a message header and message parameters appended. The messages correlated with the MIH primitives are used to trigger communications. As shown in Figure 15 of the MIH services framework and communication between a local and a remote MIHF entities, it is assumed that events, commands, and information service queries are initiated from the local en-

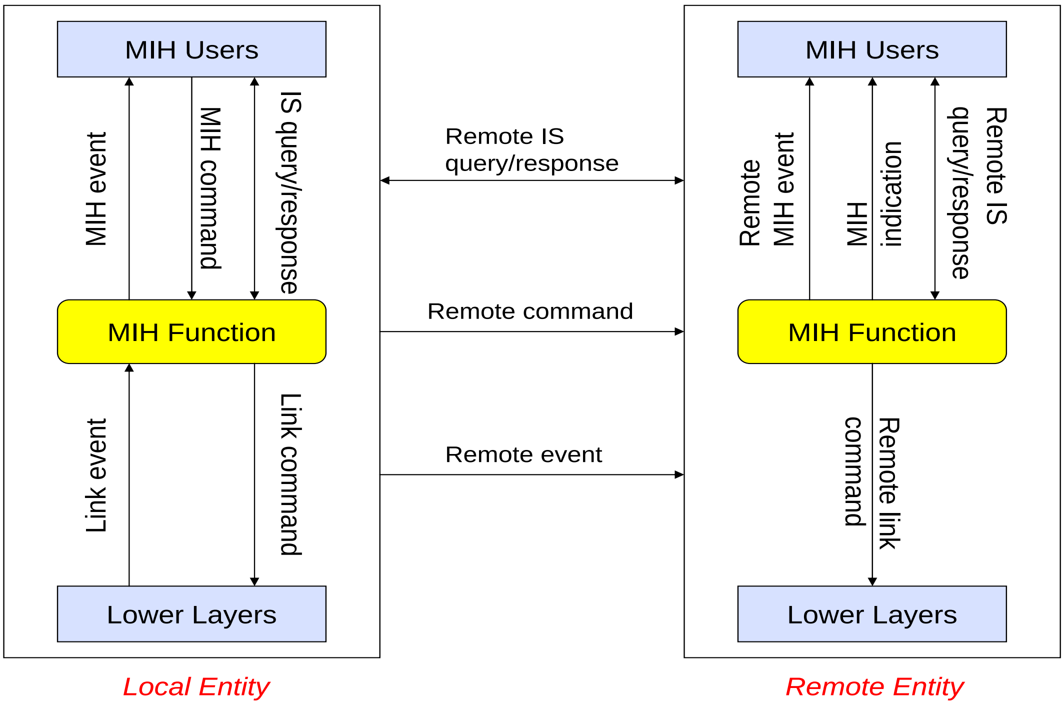


Figure 15: Communication between local and remote MIHF entities [T+09].

tity and forwarded to remote entity as remote events, commands, and information service query-responses. To understand the core of MIH, three main services are analyzed as follows:

Media Independent Event Servicess (MIES) - include events that represent changes in link characteristics such as link state and link quality. Events are used to indicate changes of state and behavior on physical, link, and logical link layers via primitives such as LINK_Up and LINK_Down. Two main categories of events include the ones that originate from lower link layers to be forwarded upwards (Link event) and the ones that originate from MIHF. MIH users can subscribe to receive notification once the corresponding events occur. Events can be local or remote. Local events are contained within single node which can be subscribed by local MIHF. On the other hand, remote events are subscribed by remote nodes and to be delivered over the network following MIH protocol messages. Event notifications are sent to MIHF or upper layer entities that either locate on a local node or on a remote one. For instance, the LINK_Up event that is generated from the link layer is propagated to MIH user on the same node is regarded as local event. If a remote MIH user has subscribed to this event, the local LINK_Up event is delivered to this remote MIH user, as depicted in Figure 15.

Media Independent Command Servicess (MICS) - provide commands to control link state. MIH users and MIHF can invoke commands either locally or remotely. Local commands from the MIH users are forwarded to MIHF and from MIHF further to lower layers. Remote commands are delivered by MIH protocol messages and can be propagated from the MIHF in local protocol stack to the MIHF in a peer protocol stack. As an example, an MIH user can control the reconfiguration of a link with Link_Get_parameters. The receiver of command can be located within the protocol stack that originates the command, or within a remote entity. Remote commands can go to lower layers as Link command or go up to MIH users as MIH indication as shown in Figure 15. For instance, a Link_Action command generated from MIHF is sent to link layer of the same node when it is a local command. The command propagated from the local MIHF to the link layer of a remote node through peer MIHF is a remote one.

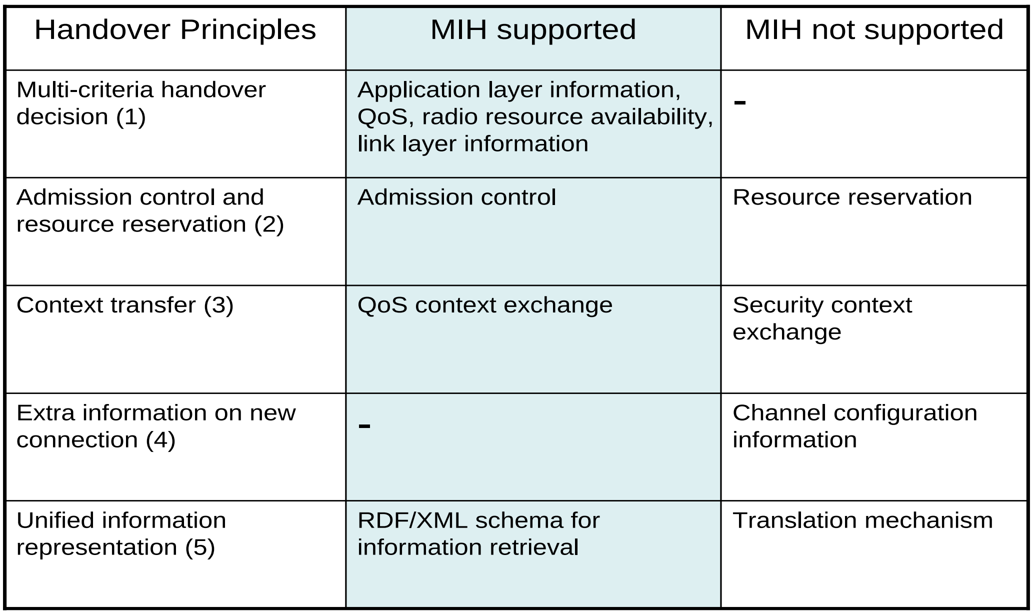
Media Independent Information Servicess (MIIS) - include a set of information elements (IEs) as well as the information structure, representation, and query-response-based mechanism for information transfer. MIIS provide a framework for MIH entities to discover information that can assist handover decision. For example, information of networks within a geographical area can be obtained from MIIS to enable effective handover decision-making and handover execution. The MIIS uses both resource description framework (RDF) [Rec04] and type-length-value (TLV) format to specify a media independent representation of information across different technologies. By using RDF which is an extensible framework, MIIS can support the creation of new IEs conforming to RDF schema. Vendor specific extensions can be added through the extended schema namespace defined in IEEE 802.21. One additional advantage of using RDF in MIIS is because RDF can support efficient responding to complex queries. For instance, MIIS can thus be used efficiently to identify neighboring networks that meet a complex set of criteria and requirements. This feature is beneficial if comparing with the normal scheme with multiple message exchanges that would require more bandwidth and lead to greater handover latency. For information that is not available locally, MIH protocol can be used to access remote information sources. The network element to store MIH related information is referred to as information server (IS).

With the provision of MIH services to support inter-technology handovers, IEEE 802.21 is regarded as a promising solution to achieve the goal of seamless mobility.

In order to demonstrate the capability of IEEE 802.21 MIH regarding to seamless handover support, analysis against the five seamless handover principles is made and summarized in Figure 16 [LSP08].

First, principle 1 is fully supported by IEEE 802.21 concerning the multi-criteria handover decision. MIH offers the capability to collect information from mobile nodes and network side covering all protocol stacks from physical to application layer. Additional information can also be retrieved from the MIIS server. Second, principle 2 is partially supported as admission control is achieved by queries to candidate networks, but the primitives to define resource reservation is missing. Third, IEEE 802.21 does not consider the exchange of security context whereas QoS information exchange is provided regarding to principle 3. Fourth, the standard does not indicate supports of connection information such as new channel configuration, as stated in principle 4. Finally, principle 5 is handled by the uniform RDF and extensible markup language (XML) schema adopted in MIIS. However, the translation of information between different systems is undefined.

Comparing with tight coupling solutions that are expensive and un-scalable to deploy, the IEEE 802.21 Media Independent Handover is a proposal for loose coupling



Handover Principles	MIH supported	MIH not supported
Multi-criteria handover decision (1)	Application layer information, QoS, radio resource availability, link layer information	-
Admission control and resource reservation (2)	Admission control	Resource reservation
Context transfer (3)	QoS context exchange	Security context exchange
Extra information on new connection (4)	-	Channel configuration information
Unified information representation (5)	RDF/XML schema for information retrieval	Translation mechanism

Figure 16: Support of seamless handover principles in IEEE 802.21 [LSP08].

that doesn't require the source and target systems to exhibit strong interdependency. The success of this standard, as pointed out, depends not only upon the activities within the IEEE 802.21 WG, but also upon the acceptance of this technology by other standards and industy forums [T$^+$09]. As standardizations must strike a balance between encouraging innovation and ensuring interoperability, some features are better left for industrial parties to address, hence creating opportunities for companies to distinguish themselves in the markets. At the same time, a mandatory set of common hooks and interfaces should be in place to ensure interoperability and allow large scale industrial deployment. The strong tie and close cooperation among IEEE, IETF, 3GPP, and international mobile telephony (IMT) advanced communities, as well as the extensible feature of IEEE 802.21, constitute a good base for the future success of IEEE 802.21 MIH standard.

4 Handover Security in Wireless Networks

As a key part of mobility management, handover management entails the task of transferring essential user context between two access networks, which in principle, must be conducted in a secure manner. However, due to the fact that existing security solutions add complexity and overhead to handover management, it is demanding to achieve both efficiency and security at the same time [C$^+$08]. To understand the security impact on handover, we provide a general discussion on essential security technique involved in handover management, covering the authentication, authorization, and accounting (AAA) framework, and the key management. Our focus is to analyze existing security schemes that can be utilized to secure handover within a well defined security enhanced wireless network.

4.1 Security in wireless and mobile environment

4.1.1 Authentication, authorization, and accounting

Given the untrustworthy nature of wireless and mobile environment, the security guarantee of network access is indispensable. With the increasing popularity of mobile services, mobile users often need to be attached to new domains convenient to their current locations [Per00]. To enable secure access for such mobile users, one key technique that is widely deployed nowadays is the authentication, authorization, and accounting (AAA) framework.

For current wireless and mobile networks, AAA provides a primary architecture for access control that ensures the mobility not to happen at the places without permission. When a mobile user needs to access an administrative domain that is not his or her home domain, the service providers in such foreign domain need a proper authorization to guarantee valid service for this visiting user. This directly leads to the authentication of user's identity as well as the accounting for billing and auditing purposes. By integrating logically three security services into a coherent framework, AAA supports security in a consistent manner.

To enable the dynamic configuration of access control, AAA adopts a modular way to perform security services through three major functions: authentication, authorization, and accounting. The authentication function in AAA provides methods to identify users who intend to obtain the access to system or network. By definition, authentication is the binding of an identity to a subject [Bis02]. The network-based authentication requires the subject e.g. mobile user to provide necessary information that can enable the system to confirm its identity. In principle, authentication consists of the user and the verifier. When a user attempts to confirm his identity to the verifier, he needs to provide one (or more) of the following information: 1) what the user knows, such as passwords or shared secret information; 2) what the user has, such as ID card or digital signature; 3) what the user is, such as biometric measures; and 4) where the user is, such as the location in the lab. According to the authentication policy, the verifier will validate the user's claim with auxiliary equipment. The authentication policy defines whether an authentication should be done and which mechanisms and algorithms should be applied. Concerning the AAA system, the security server functions as the verifier for the service users.

The authorization function in AAA provides methods to verify and grant rights and restrictions entitled to each user. The authorization policy is a key term of AAA authorization to define the condition and manner for corresponding actions performed by the user. The policy decision can be either positive or negative, permitting or prohibiting the action, respectively. The network-based authorization adopts the attribute-value (AV) pairs in its security data base to associate the specific right with the appropriate user under certain conditions. The attribute types in AAA authorization include the network access, IP address assignment, and QoS support, etc. The form of such authorization AV pairs can be described as 'subject S is allowed to perform action A on object O under the condition C' [R$^+$01]. Based on the authentication result, AAA authorization works by assembling the set of attributes which describe what a user is authorized to perform, and comparing them to the information in security data base. The result is returned to AAA to determine the actual capabilities and restrictions for each user according to the predefined authorization policies, which take into account technical (e.g., available bandwidth) and financial (e.g., remaining credit) aspects.

The accounting function in AAA provides methods to collect and aggregate information from each user regarding to their service utilization. The information recorded covers user identity, service type, service duration, executed commands, and size of traffic, etc. The accounting function enables the system to track the service and network resources consumed by each authorized user. When AAA accounting is activated, the responsible entity e.g., a network access server (NAS) which functions as the point of attachment to the domain will report accounting records to the AAA server. Following the format defined in corresponding AAA protocols, accounting records can be used to assist network management, service billing, auditing and trend analysis.

To coordinate three AAA functions, different AAA protocols are proposed by IETF for various application scenarios. The major AAA services are defined by AAA protocol in terms of message format, extensions, error handling, and generic system architecture. In current mobile networks, RADIUS (Remote Authentication Dial In User Service) [R+97] and Diameter [C+03] are two AAA protocols enjoying wide deployment.

As a centralized access management protocol, RADIUS protocol is designed originally for the dial-up connections of telecommunication systems. It defines protocol operations and message formats for transferring authentication, authorization, and configuration data between network access server (NAS) and RADIUS server. A NAS in RADIUS functions as a client to communicate with other RADIUS servers, which hold information for authentication and authorization. A RADIUS server can act as a client to other RADIUS servers in order to exchange AAA information. The accounting support is defined in an extension to deliver accounting records to RADIUS accounting server [Rig97].

As the successor of RADIUS, Diameter protocol is proposed to remove some deficiencies of RADIUS such as weak user-password protection [M+01a]. By providing a base protocol for header formats, extensions, mandatory commands, and attribute value pairs (AVPs), Diameter is regarded as a flexible and extensible AAA protocol [C+03]. The AAA functionality is implemented through Diameter extensions to allow the adaptation of different access technologies. The Diameter extension

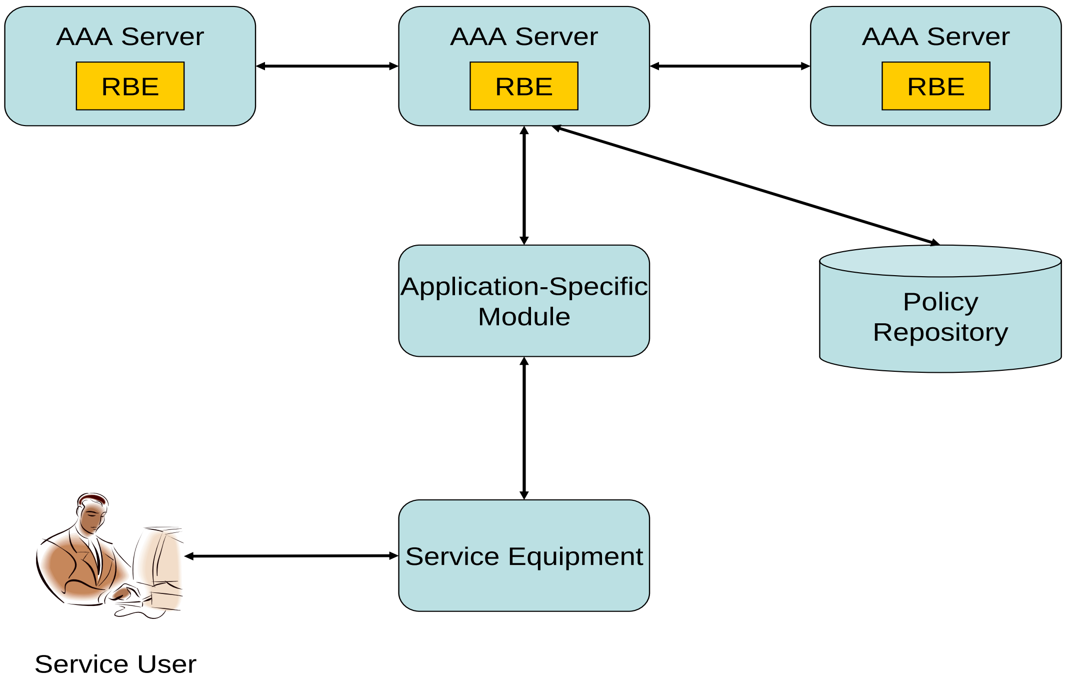


Figure 17: AAA architecture [L+00].

which is referred to as Diameter application defines special command codes and AVPs to implement specific functions. To support mobility, Diameter includes an extension for Mobile IP which supports the mobility of mobile users across different domains [C+05a]. Different from RADIUS, Diameter uses reliable transport protocol such as TCP or SCTP instead of UDP. Although Diameter is not fully compatible with RADIUS, its NAS application extension [C+05b] supports the RADIUS authentication as well as the authorization needed by NAS services. The extensible feature as well as mobility and security enhancement make Diameter a suitable protocol of AAA services for the next generation IP based networks.

A key assumption for the current AAA architecture is the multi-domain network topology where each administrative domain resides at least one AAA server. Distributed servers cooperate with each other to offer AAA services to users inside the network [R+01]. As illustrated in Figure 17, the existing architecture of AAA consists of functional components, AAA policies and three AAA services (authentication, authorization, and accounting). Different components are combined together by various protocols. The policy based model is adopted to assist inter-organizational operations, in which an AAA policy is defined as an aggregation of rules made up of conditions and corresponding actions [M+01b].

The essential components in the AAA architecture include Rule-Based Engine (RBE), Policy Repository (PR), Application-Specific Module (ASM), and Service Equipment (SE). Rule-Based Engine (RBE) is a central component residing in the AAA server, which evaluates conditions, makes a decision and executes actions according to AAA policies. Policy Repository (PR) stores the AAA policies to be used by AAA services. Application-Specific Module (ASM) is an interface between AAA server and Service Equipment (SE) which receives and delivers messages from both sides. SE is the equipment providing services to the user.

When AAA server receives a request from SE via ASM, or directly from other AAA servers, the request is inspected by RBE, which consults the PR for AAA policies. To evaluate policy conditions, other AAA servers may be consulted through direct communication between two servers using an AAA protocol such as Diameter. Take the mobile networks for example, ASM is realized as an access point or NAS in home or foreign domain, while SE is the mobile node. The ASM forwards the access request to the AAA server to authenticate the user. Based on the response from AAA server, ASM accepts or rejects user's request. The PR holds essential user information which is used by AAA server. The information includes user session state, accounting data, and log actions, etc. Concerning the mobility aspect, the security and trust relationship between different AAA entities are crucial. To authenticate and authorize user in a foreign domain, the trust relationship should be maintained along the chain of servers that belong to both home and foreign domains [V+00].

4.1.2 Key management

The management of cryptographic keys plays a critical role in security. Because information security protected by cryptography not only relies on the strength of cryptographic algorithms, but also on the protection mechanisms and protocols associated with the keys, a poor design of key management may even compromise strong algorithms, leading to the failure of whole cryptographic system. This crucial fact implies that cryptographic keys should be protected against modification and unauthorized disclosure during their lifetime. A proper key management provides a secure foundation for the generation, storage, distribution, and destruction of cryptographic keys.

In current cryptographic system, key management refers to the distribution of cryptographic keys; the mechanisms used to bind an identity to a key; and the generation, maintenance, and revoking of such keys [Bis02]. One assumption on which key management relies is that the identity can correctly define the principal, implying that a key bound to a user A is really A's key rather than that of user B. This assumption is based on the authentication of principal's identity and the proper identity representation. The discussion of identity representation can be found in the related document [Bis02], and its content is beyond the scope of this thesis.

Concerning the central component of key management, a cryptographic key is a parameter used in conjunction with a cryptographic algorithm that determines its operation in such a way that an entity with knowledge of the key can reproduce or reverse the operation, while an entity without knowledge of the key cannot [B$^+$07]. The secrecy provided by cryptographic system resides in the generation and selection of keys. In order to prevent attackers from guessing the keys, the key generation process is implemented by a random number generator to produce keys that are theoretically unpredictable and irreproducible. Normally a cryptographic key is attached to an expiration date, which determines the lifetime of the key to be used in the cryptographic system.

For network security, the differences between the communication and the principals e.g. users distinct two types of keys: interchange keys and session keys. By definition, an interchange key is a cryptographic key associated with a principal to a communication, while a session key is a cryptographic key associated with the communication itself [Bis02]. Compared with the session key, the interchange key does not change over communication sessions, which makes it ideal to be used in authentication. The session key is generated and used for each communication session and discarded when the session ends. Session keys are used mostly for the encryption of data transfer and normally do not authenticate any principal.

Identified by the National Institute of Standards and Technology (NIST) [NIS09], the life circle of key management can be divided into 4 phases as depicted in Figure 18, including pre-operational phase, operational phase, post-operational phase, and destroyed phase [B$^+$07]. In pre-operational phase, keying materials are not available for cryptographic usage and are under the generation process. In operational phase, keying materials are available for cryptographic operations such as key exchange, data encryption, and data decryption. In post-operational phase, the keys are no longer used while the access to keying materials may still be possible. Keys of the post-operational phase are archived. In the final destroyed phase, keys are no longer

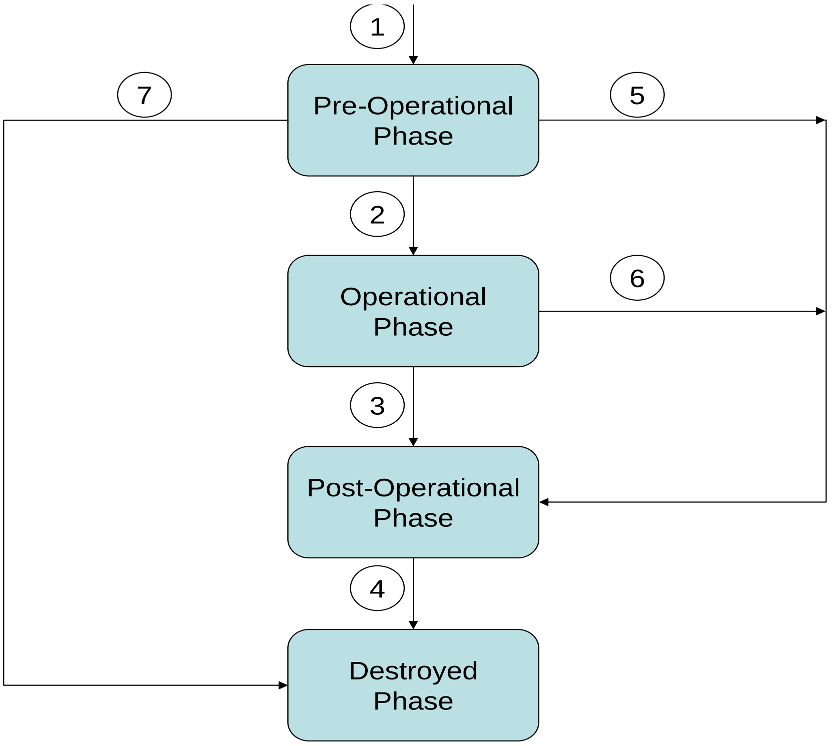


Figure 18: Key management phases [B+07].

available and all related records are deleted.

According to the flow diagram in Figure 18, seven transitions are identified, in which keys shall not be able to transfer back to any previous phase. For transition 1, the cryptographic key is generated, but it is not authorized to be used; For transition 2, once the required key attributes are established, key is ready to be used and distributed; For transition 3, when keys are no longer needed for normal use such as encryption, but access to the keys needs to be maintained, the management procedure will go to the post-operational phase; For transition 4, when the key in post-operational phase is no longer needed, it will go to the destroyed phase; Transition 5 and 6 deal with the case that if a key is compromised, the management procedure will go to the post-operational phase; For transition 7, keys that are produced but never used will be destroyed by directly going from the pre-operational phase to the destroyed phase.

For key management, one fundamental concept is the cryptographic key infrastructure which provides necessary support for key operations in cryptographic systems. The cryptographic key infrastructure is responsible for regulating the key-identity binding, selecting security algorithms, and implementing key exchange protocols.

In current cryptographic key infrastructure, key exchange protocols are essential to guarantee the safety of key distribution. The goal of key exchange protocol is to enable secure communication between two parties basing on the long term trust. To achieve this goal, solutions to key exchange protocol should meet three criteria: 1) the key in use can not be transmitted in clear text; 2) a trusted third party in key infrastructure may be used; 3) the cryptographic algorithms and protocols are publicly known, while the only secret data is the cryptographic keys involved [Bis02].

Currently, the Internet Key Exchange version 2 (IKEv2) [Kau05] is a standard key exchange protocol used in Internet Protocol Security (IPsec) architecture [KS05] to enable secure key exchange. The major task of IKEv2 for IPsec is to set up security association (SA), which is a set of cryptographic methods and associated key materials to be used by IPsec for Encapsulating Security Payload (ESP) [Ken05b] and/or Authentication Header (AH) [Ken05a] used for data encryption and authentication, respectively.

As illustrated in Figure 19, the operation of IKEv2 consists of two phases that are referred to as IKE phase 1 and IKE phase 2. In IKE phase 1, the main goal is to establish a secure and authenticated channel between two communicating peers. IKEv2

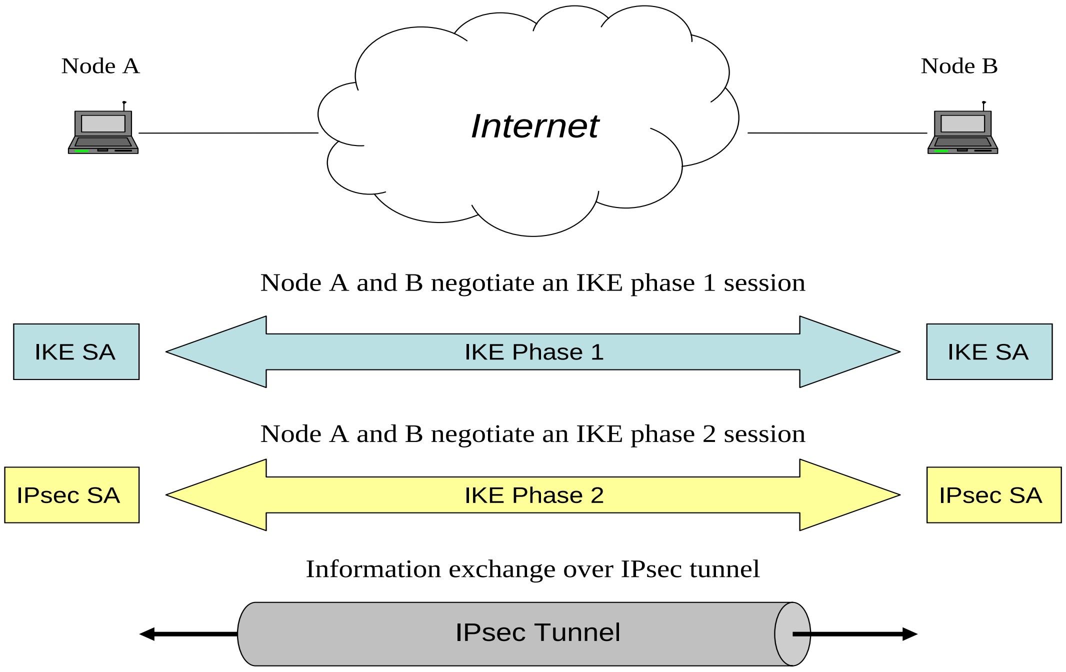

Node A
Node B
Internet
Node A and B negotiate an IKE phase 1 session
IKE SA
IKE Phase 1
IKE SA
Node A and B negotiate an IKE phase 2 session
IPsec SA
IKE Phase 2
IPsec SA
Information exchange over IPsec tunnel
IPsec Tunnel

Figure 19: IKEv2 operation.

performs mutual authentication by using the long term shared trust held by each peer to validate peer identity. The Diffie-Hellman key exchange algorithm [Res99] is used in phase 1 to generate a shared secret key to encrypt further IKEv2 communication. The negotiation in phase 1 results in a shared bi-directional IKE SA on both peers that is used to set up secure communication channel. In IKE phase 2, by using the existing IKE SA established in phase 1, communicating peers negotiate IPsec SA parameters through the secure channel. The phase 2 negotiation yields a minimum of two uni-directional IPsec SAs, one inbound and one outbound, respectively, for IPsec operation. Once the IKEv2 negotiation is complete, further communication between peers is protected by IPsec by going through the secure tunnel.

Since the characteristics of different cryptographic systems vary from each other, key exchange protocols should follow specific requirements to reach the corresponding goals. For AAA key management, one major goal is to distribute keying materials to entities in the situations where key derivation can not be used by such entities e.g. a wireless access point, because such entities may lack the resources necessary to implement various authentication mechanisms required, or it is undesirable for each entity to engage in a separate key management conversation [HA07]. For classical cryptographic systems, where principals share a common trusted key, key agreement is often used. In such environment, key agreement enables the communication principals to establish keying materials based on the information contributed by the principals without actually sending the keying materials. The shared symmetric key in classical cryptographic system is used to calculate shared secrets and then derive other keying materials for further use. Only approved key agreement schemes shall be used in such cases [B$^+$07].

4.1.3 Security requirement for wireless mobility

To enable secure data communication, security requirements are needed to regulate design and implementation. Based on previous research efforts for decades, major security requirements for data communication can be summarized as follows: 1) confidentiality is necessary for secure communication which provides resistance to interception and eavesdropping; 2) authentication guarantees the integrity of data and the identity of principal, corresponding to data modification and impersonation; 3) anti-replay helps detect message that is a replay of previous session; 4) non-repudiation prevents denial and fabrication; 5) access control prevents unauthorized access; 6) availability ensures the resources are always accessible [XMH06, Sta06, Bis02].

Because wireless communication is based on transmitting signals via electromagnetic waves over unbounded media, this nature enables the pervasiveness of data transmission regardless of physical boundaries. Nevertheless, as wireless communication is conducted over such open channel, it suffers from a number of vulnerabilities such as eavesdropping, identity cheating, and data modification [BH07]. These challenges make security issue a key concern in wireless networks.

At the same time, the fast growing popularity of mobile services demands a careful design of security solutions for existing wireless environment. As mobile users move frequently across different wireless domains with their devices running applications, this dynamic change of location and environmental context bring challenges to the existing security solutions. Besides the fundamental security requirements, security solutions in a wireless and mobile environment should also take the mobility aspects into account.

In general, security approaches supporting mobility should obey basic principles that are believed to be crucial to the suitability and acceptance of mobility protocols in wireless networks. The principles contain the following elements as efficiency, scalability, transparency, and manageability: 1) for efficiency, it demands that the design should induce little overhead and computing requirement; 2) for scalability, it demands that the scheme should be able to work in networks with many mobile users and frequent handovers; 3) for transparency, it demands that the implementation should bring as little changes as possible to existing systems, especially to mobile terminals; 4) for manageability, it requires that the mechanisms and protocols should be managed efficiently without complicated operations [M$^+$00].

As the general solution to wireless mobility, mobility protocols such as Mobile IP exert huge impact on the security infrastructure and key management. Concerning the widely deployed AAA infrastructure, new components are introduced to the architecture to integrate with mobility protocols, including the AAA Broker, local AAA authority, and home AAA authority, as shown in Figure 20. Such changes generate new security requirements to both AAA entities as well as AAA key management.

Take Mobile IP for example, in order to provide better mobility support, IETF has identified essential requirements that integrate security implementations with Mobile IP [G$^+$00]. The key elements are summarized as follows: 1) Mobile node (MN) and AAA home entity need to authenticate each other before access is permitted. A security association (SA) should be shared between MN and AAA home entity; 2)

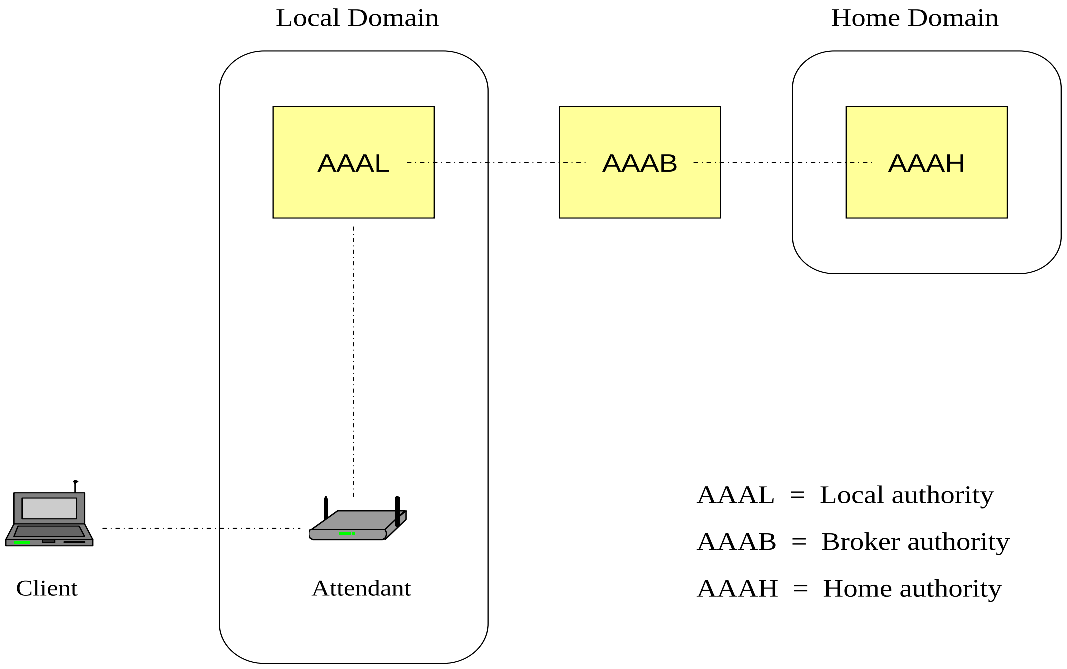


Figure 20: AAA architecture using a broker [G^{+}00].

Each local attendant such as an access point should have a security relationship with the local AAA server; 3) The local authority has to share security relationship with external authorities that are able to validate customer credentials; 4) The attendant has to protect against reply attacks; 5) To keep the mobile node's credentials safe, intervening nodes such as attendants or local AAA server must not learn any confidential information which might compromise the credentials; 6) For scenarios where one attendant need to manage requests from many customers at the same time, the attendant should be easy and inexpensive to implement, so that it can be replicated in the foreign domain to handle multiple requests; 7) The local AAA server has to share or dynamically establish security associations with home AAA server. To provide a scalable solution, AAA broker entity may be used, which has SAs with both local AAA and home AAA server. AAA broker can act as a proxy between local AAA and home AAA and relay shared secret key to them in order to set up SA; 8) After successful authentication, MN is allowed to access the network and use services such as mobility protocol functionality; 9) AAA protocol should enable transport of mobility related messages, e.g. Mobile IP registration, as part of the initial registration sequence to be handled by AAA server. Any mobility related message transported via AAA entities should be considered opaque to AAA system; 10) For handover within one administrative domain, local AAA server should be able to provide necessary authentication without involving the home AAA server; 11) For handover between different administrative domains, local AAA server in new domain may contact AAA server in previous domain to verify authenticity of customers and/or obtain session keys for security operations.

Concerning the key management within AAA framework, security requirements are needed for current wireless networks with diverse protocols employed. To improve interoperability, the following requirements constitute a good base for security designers covering essential aspects of AAA based key management: 1) Cryptographic algorithm independent, which requires that AAA key management protocol must be cryptographic algorithm independent, so that possible cryptographic algorithms can be negotiated and selected to provide resilience against compromise of a particular cryptographic algorithm; 2) Strong and fresh session keys, which requires that session keys must be strong and fresh, with each session deserving an independent session key so that the disclosure of one session key does not aid the attacker in discovering any other session keys; 3) Limit key scope, which requires that entities must not have access to keying material that is not needed to perform their role; 4) Replay detection mechanism, which requires that key exchange must be replay protected to enable the recipient to discard any message that was recorded during a previous legitimate dialogue and presented as though it belonged to the current dialogue; 5) Authenticate all parties, which requires that each entity in the key management protocol must be authenticated to others with whom it communicates; 6) Peer and authenticator authorization, which requires that peer and authenticator authorization must be performed and entities must demonstrate possession of the appropriate keying material, without disclosing it; 7) Keying material confidentiality and integrity, which requires that confidentiality and integrity of all keying material must be maintained; 8) Confirm ciphersuite selection requires that the selection of the "best" ciphersuite should be securely confirmed; 9) Uniquely named keys, which requires that all keys must be uniquely named, and the key name must not directly or indirectly disclose the keying material; 10) Prevent the Domino effect, which requires that compromise of a single entity must not compromise keying material held by any other entities within the system; 11) Bind key to its context, which requires that keying material must be bound to the appropriate context. The discussions of key management security can be found in related IETF document [HA07], while its details are beyond the scope of this thesis.

4.2 Handover security

Handover security guarantees a safe transfer of critical context for mobile users moving across different wireless access networks. To offer rich and safe mobile services in the upcoming wireless IP networks with heterogeneous accesses, organizations including IETF and IEEE propose solutions for handover security, covering access authentication, key management, and secure IP mobility. Meanwhile, 3GPP works on the architecture framework to integrate various protocols required by handover security. By studying major protocols and architecture, we analyze the security impact on handover in terms of system performance.

4.2.1 Handover security schemes

Due to the resource constraints on mobile devices, security mechanisms for handover should be designed carefully to avoid negative impact on performance and energy consumption while preserving the security at a satisfactory level. In the past decade, a number of schemes have been introduced by the Standards Developing Organizations (SDO) such as IEEE, IETF, and 3GPP to enhance handover security. As a base for existing handover security mechanisms, their contributions can be summerized into three categories: access authentication, key management, and secure IP mobility.

As the first step of handover security, access authentication enables the access control of network resources by validating the identity of entities involved in a handover. A typical access authentication in wireless and mobile environment takes place between a mobile node and its corresponding authentication server supported by authentication protocols. Currently, the Extensible Authentication Protocol (EAP) [A$^+$04] proposed by IETF provides a flexible and extensible framework to utilize various authentication implementations and becomes a standard protocol for access authentication in existing wireless networks.

By defining a base protocol to transport authentication messages, EAP can be used for both wired and wireless links. It operates in a peer-to-peer manner where independent and simultaneous authentication can take place on both ends of the link at the same time. As EAP adopts the lock-step mechanism, which supports only a single packet in flight, it is not suitable for bulk data transport. For network access authentication where IP connectivity may not be available, EAP supports message transport without requiring IP [80204]. At the same time, EAP provides its own mechanisms for retransmission and duplicate elimination while the message ordering is expected from the lower layer. For protocol function, the following terms are introduced by EAP: EAP method, authenticator, peer, backend authentication server, and EAP server [A$^+$04].

First, as illustrated in Figure 21, an EAP method is the mechanism for a specific authentication implementation. Because EAP does not provide specific authentication mechanisms, the EAP method is the implementation of security schemes and algorithms. Currently, multiple methods are developed by SDOs, such as the Extensible Authentication Protocol Transport Layer Security (EAP-TLS) [SAH08] for WLAN and the Extensible Authentication Protocol Authentication and Key Agreement (EAP-AKA) [AH06, ALE09] for 3G networks. Second, an authenticator in EAP is the end of the link that initiates the EAP authentication, which usually is an access point or a gateway router. Third, a peer is the end of the link that responds to the authenticator. Fourth, a backend authentication server is the entity that provides authentication services to the authenticator by executing EAP methods. Fifth, EAP server is the entity that terminates the EAP authentication method with the peer. As an abstract concept, there are two cases for the EAP server: for the case where no backend authentication server is used, EAP server is part of the authenticator, and for the case where authenticator operates as a relay agent for the backend authentication server, EAP server is located on the backend authentication server.

Nowadays, EAP has been implemented on hosts and routers that connect to Internet through switch circuits or dial-up lines with PPP [Sim94, A$^+$04]. IEEE adopts EAP in many of its standards such as the IEEE 802.1X [80204] for IEEE 802 wired media, and IEEE 802.11i [IEE07] for IEEE 802.11 wireless media. Regarding the protocol stack and operations, Figure 21 presents an example of EAP stack working with IEEE 802 wireless systems and EAP packet format.

As shown in the figure, the authenticator works in the pass-through mode, which is required by the AAA protocols such as Diameter EAP application [EHZ05]. EAP messages are encapsulated using specific protocol such as the Extensible Authentication Protocol over LAN (EAPoL) [80204] and are transmitted between peer and authenticator via the lower layer data link protocol such as IEEE 802.11 [IEE07]. Between the authenticator and the EAP server, AAA protocols such as Diameter protocol [C$^+$03] are used to transport authentication messages via TCP/IP.

To illustrate protocol operation, EAP-AKA full authentication message flow is pre-

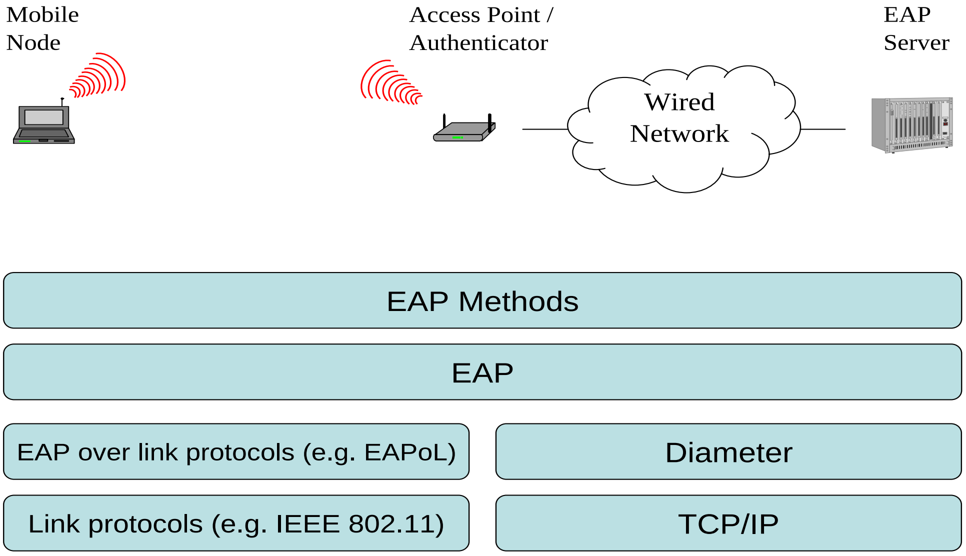
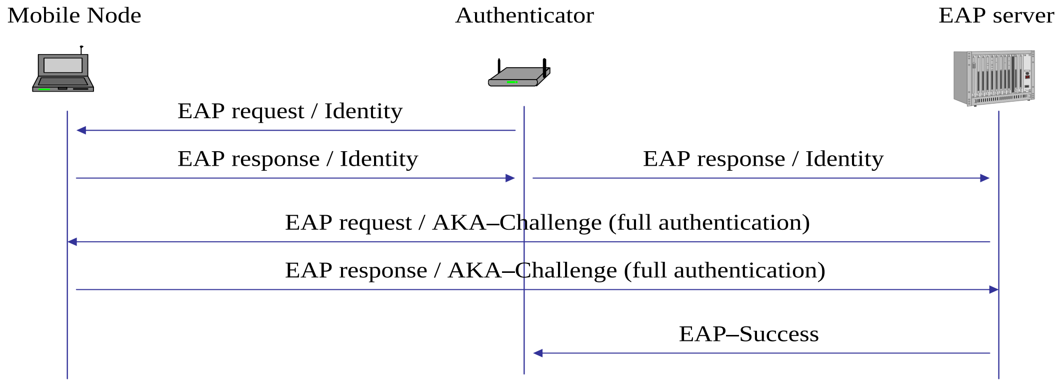


Figure 21: EAP protocol stack with IEEE 802 networks [A^{+}04].

sented in Figure 22 (a) as an example. In general, EAP authentication proceeds by the following four general steps: 1) authenticator sends a request to the peer; 2) peer sends a response to authenticator in reply to a valid request. If backend authentication server is used, the response will be forwarded further to the backend authentication server; 3) authenticator or backend authentication server sends additional request messages and peer replies with response. This exchange continues as long as needed, and after a pre-defined number of rounds, authenticator or backend authentication server will end the conversation; 4) based on the previous exchange of authentication messages, authenticator or backend authentication server will send the final message as EAP Success or EAP Failure to the peer.

In EAP-AKA full authentication, AKA method [3GP09] is adopted and encapsulated in EAP packet to provide mutual authentication. As shown in Figure 22 (b), every EAP packet begins with the 5-byte ’Code’, ’Identifier’, ’Length’, and ’Type’ fields. EAP method specific Type-Data fields immediately follow the basic EAP header fields. For EAP-AKA, the Type-Data begins with the EAP-AKA header which consists of 1-byte ’Subtype’ field and 2-byte ’Reserved’ field. Following the EAP-AKA header, the rest of Type-Data consists of various attributes that take the



(a) EAP-AKA full authentication

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Code      |  Identifier   |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |    Subtype    |           Reserved            |

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Attribute Type |    Length     |           Value...

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

(b) EAP-AKA packet format

Figure 22: (a) EAP-AKA full authentication; (b) packet format [AH06].

generic TLV form with ’Type’, ’Length’, and ’Value’ fields.

As depicted in Figure 22, an EAP-AKA full authentication starts with an EAP-request sent from authenticator to MN, in which the ’Code’ field is set to 1 and ’Subtype’ field is set to 5 indicating an EAP-AKA Identity request. Once receiving the EAP-request/Identity message, MN replies with an EAP-response containing its network identity recognizable in the system. The EAP-response/Identity message from the MN will be forwarded to the EAP server by the authenticator, where authenticator is functioned in the pass-through mode. After obtaining MN’s identity, EAP server sends an EAP-request/AKA-Challenge to the MN to start the AKA mutual authentication process. When the EAP-request/AKA-Challenge from EAP server is verified as valid, MN replies with an EAP-response/AKA-Challenge indicating that MN has successfully authenticated the EAP server and this EAP exchange will be accepted by MN’s local policy. If EAP server verifies the attributes of MN’s reply as valid, EAP server will send a final EAP-Success message to the authenticator, confirming the successful authentication and authorizing the access right of MN.

The advantages of using EAP in access authentication lies in its flexible framework, in which the authenticator devices such as access point do not need to understand each authentication method and can simply relay the messages to the corresponding backend authentication server. The separation of the authenticator from the backend authentication server allows one server to serve many authenticators, which eases the management of security policy. This separation also centralizes the decision making and keeps the complexity and cost of each authenticator low. By acting in the pass-through mode, access points do not need frequent upgrades for all the upcoming authentication methods. At the same time, it should be noted that this separation complicates the security analysis and brings management overhead such as key management.

Besides access authentication, another critical issue in handover security is the key management. Because all authentication and integrity checks involve the key operation, how to derive and distribute keys safely in a distributed environment is challenging. Currently, 3GPP is working on its System Architecture Evolution (SAE) / Long Term Evolution (LTE) [3GP08a, 3GP08b] that aims at integrating multiple wireless access technologies to deliver rich and secure services. Two main protocols are proposed for SAE/LTE to handle authentication and key distribution during the handover: Universal Mobile Telecommunications System Authentication and Key Agreement (UMTS-AKA) [3GP09] and Extensible Authentication Protocol Authentication and Key Agreement (EAP-AKA) [AH06] with its latest enhancement EAP-AKA' [ALE09]. Both UMTS-AKA and EAP-AKA enable mutual authentication and key distribution for handover security.

For key management of access authentication, the trust relationship between different networking entities can be summarized as follows [Nak07]: Peer-Server, Authenticator-Server, and Peer-Authenticator. The Peer-Server relationship has two aspects, permanent and transient. Permanent relationship is based on the credentials that are configured for peers at the time of subscription. The transient one with limited life time is based on the master session key (MSK) and extended master session key (EMSK) that are created as a result of successful authentication between the peer and server. The Authenticator-Server relationship is a permanent trust, which is based on the existing AAA infrastructure, where the AAA security is provided by lower layer IPsec [KS05] or TLS [DR08]. The Peer-Authenticator relationship is a dynamic one, which needs to be established during the handover. For scalability, peers share long term trust only with the server, but when needed, transient shared keys are established dynamically between the peer and authenticator for secure communication.

To illustrate the operations of access authentication and key management in wireless and mobile environment, Figure 23 presents a general EAP based procedure adopted by 802.11i, 802.1X, and 802.16e for handover security. As shown in the Figure 23, the first phase of security capability discovery is to allow the peer e.g. a mobile node to obtain necessary information from authenticator for initiating access authentication. In this phase, although peer and authenticator are exchanging messages, none of the entities are authenticated nor do they obtain the keys. The second phase is for authentication and key generation, where mutual authentication is fulfilled by authentication schemes such as EAP-AKA and UMTS-AKA. The key generation is done at this phase for the transient keys that will be used for the communication session. The third phase is for key delivery and derivation, where after the second phase, peer and server already obtain enough information to derive the session keys. The EAP server will distribute the derived session key to the authenticator. The fourth phase is for the negotiation between peer and authenticator on additional keys that can be used for communication security, with an example of the temporal key in IEEE802.11 system that is used for link level encryption of data over the wireless link [IEE07].

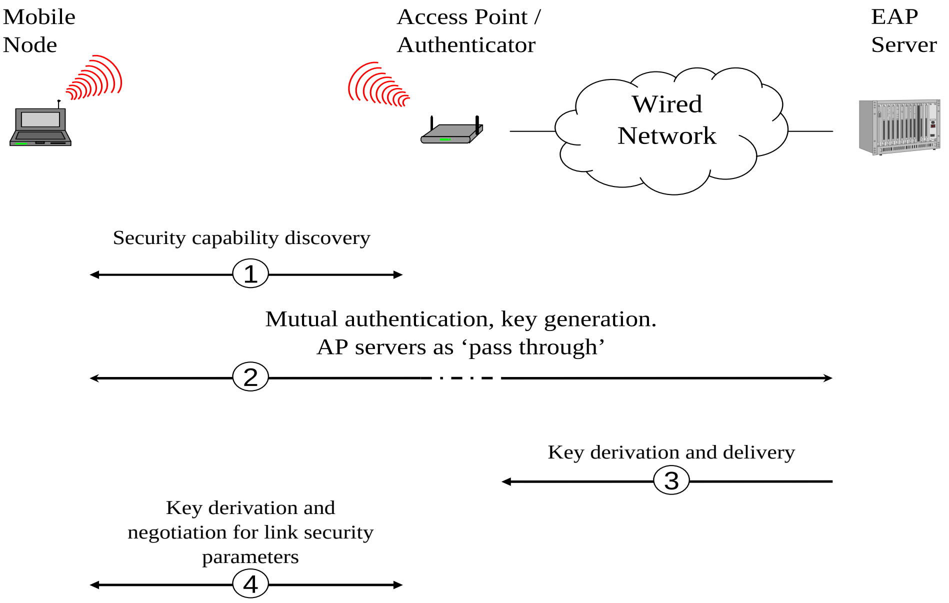

Mobile Node
Access Point / Authenticator
EAP Server
Wired Network
Security capability discovery
1
Mutual authentication, key generation.
AP servers as ‘pass through’
2
Key derivation and delivery
3
Key derivation and negotiation for link security parameters
4

Figure 23: Key management in EAP based authentication [IEE07].

Finally for secure IP mobility, existing IP mobility schemes all demand authentication of both parties in a handover as well as secure transport of signaling messages. By enabling authentication and encryption of IP packet, IPsec [KS05] provides sufficient security support required by the current IP based mobile environment. Concerning the handover phase where the mobile node moves into a new network, the registration and binding update should be transported to corresponding entities either in the local network or the home network for mobility management. In order to achieve secure handover for such IP mobility, IKEv2 [Kau05] is used by IPsec to mutually authenticate and dynamically establish IKE security association (SA) on the communicating parties. Based on the information exchange protected by IKE SA, network entities can derive IPsec SAs for Encapsulating Security Payload (ESP) [Ken05b] and/or Authentication Header (AH) [Ken05a] and a set of cryptographic algorithms to be used to protect the following IP mobility management traffic. In Proxy Mobile IP [G$^+$08, LYC08], IPsec and IKEv2 are mandatory components to secure the PMIP management signaling. The details of IKEv2 operation in IPsec for mobility management can be found from related RFC documents [Kau05, KT06, Ero06, DE07, DE08], and the details are beyond the scope of this thesis.

4.2.2 2-Phase model of handover security

Taking into account the growing demand for secure and high quality services on current mobile systems with limited resources and capabilities, security implementations on such systems need to consider to what extent the performance will be affected [ONY03]. For handover management, the main challenge lies in how to achieve a balance between security and performance for a handover. Because the answer to this challenge can shed light on the design and development of future handover security schemes, an in-depth analysis of the security impact is hence of great value as a guideline to promote the overall understanding of this challenging topic.

To analyze the security impact, we identify two general phases of handover security in existing wireless and mobile networks: the phase of secure connectivity and the phase of secure reachability. In our 2-Phase model, the first secure connectivity phase represents the access authentication with key distribution when a mobile node attaches to a network. By passing the authentication, mobile node gains the access right to the network with link connectivity. In the second phase of secure

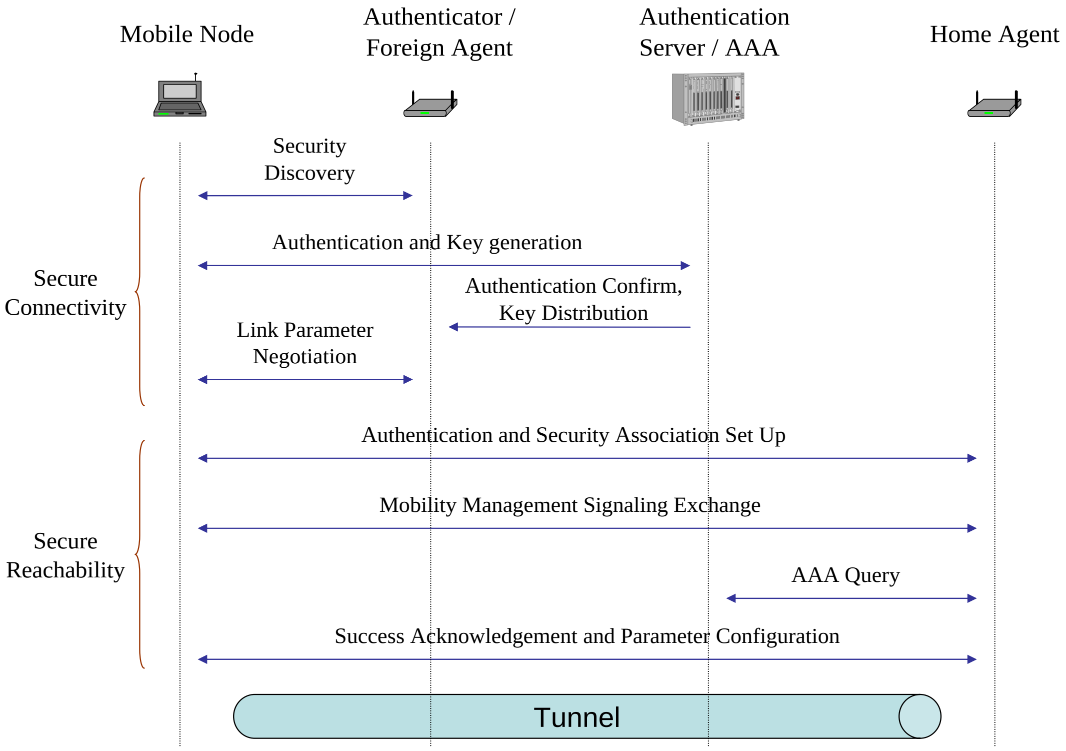

Mobile Node
Authenticator / Foreign Agent
Authentication Server / AAA
Home Agent
Security Discovery
Authentication and Key generation
Authentication Confirm, Key Distribution
Link Parameter Negotiation
Secure Connectivity
Authentication and Security Association Set Up
Mobility Management Signaling Exchange
AAA Query
Success Acknowledgement and Parameter Configuration
Secure Reachability
Tunnel

Figure 24: Two phases of handover security.

reachability, key exchange protocol is adopted to set up security associations (SA) to protect the mobility management signaling, e.g. registration and binding update in Mobile IP. Mobility management messages and AAA queries are included in the second phase as integrated parts of handover. Once the mobile node successfully passes two phases, its handover process is complete. A communication tunnel is established between mobile node and its home network, hence making the mobile node fully reachable by other entities in the network.

Regarding the protocol operation in our 2-Phase model, Figure 24 depicts the major message flows in a typcial wireless access network with involved networking entities including the mobile node, authenticator, authentication server, AAA server, foreign agent, and home agent.

In the first phase of secure connectivity, the main goal is to authenticate corresponding entities in the network including mobile node, authenticator, and authentication server. The access authentication starts with the 'security discovery' between MN and corresponding authenticator, in which authentication related parameters such as supported algorithms and lifetime of keys are negotiated and agreed. The key derivation and distribution are carried out by key management protocols together with the access authentication process during the 'authentication and key generation' exchange. Mutual authentication and key management are supported by technology specific protocols, such as EAP-AKA and UMTS-AKA [AH06, 3GP09] adopted in 3G networks, and EAP-TLS [SAH08] used in WiFi networks. Between the authenticator and authentication server, the transport of authentication messages is conducted over AAA protocols such as RADIUS or DIAMETER [R^{+}97, C^{+}03]. Once the authentication server successfully authenticates the mobile node, confirmation will be sent to the authenticator along with necessary information to derive session keys. Based on the answer from the authentication server, authenticator will initiate link specific negotiation with mobile node to configure link parameters used for data encryption and/or privacy protection over the wireless access, such as the Privacy and Key Management Protocol version 2 (PKMv2) [Man02, XMH06] used in 802.16e networks.

In the second phase of secure reachability, the main goal is to set up necessary security association (SA) on peer entities to protect their mobility management signaling. For IP mobility, peer authentication is supported by IPsec and key exchange protocols such as IKEv2 [Kau05]. By using IKEv2, communicating peers are authenticated in the first IKE phase in order to set up IKE SAs on each peer. Protected by the previously established IKE SAs, the IPsec SA is established in the second IKE phase. Once the IPsec SAs are established on both peers, mobility management signaling such as registration and binding update in Mobile IP can be transported securely though IPsec tunnel. To obtain user credential and validate mobility management messages, home agent (HA) consults the AAA server in the home network through AAA protocols. Based on the response from the AAA server, HA will make the decision to conduct IP mobility operations. By passing the two phases of handover security, the mobile node obtains the access right to the network with an established IP communication tunnel enabling its reachability in the IP network.

It should be noted that the entities and signaling depicted in Figure 24 represent only their logical meaning for the sake of simplicity to convey a general view, while the implementations in real practice may vary in different cases. A typical example in Mobile IP is that the authenticator and the foreign agent (FA) may locate on different devices instead of on the same one as depicted in this figure. The major entities and relationships described in Figure 24 form a general framework for our analysis of handover security in IP mobility.

4.2.3 Security impact

As a key part in handover security, the trust relationship between different network entities affect all aspects of security implementations including authentication, key management, AAA operation, and protocol design. Based on the 2-Phase model, we highlight the major relationships between communicating peers in Figure 25 to illustrate their security impact.

In secure connectivity phase, a long term trust exists between a MN and its authentication server. The authentication of MN is based on this long term trust. An example of this relation can be found between a mobile subscriber and his 3G network, where a contract is made between this subscriber and his operator with shared credentials for access authentication. Between the authenticator and the authentication server, the trust relationship is provided by the infrastructure where network providers guarantee the safety of operations. Between MN and authentica-

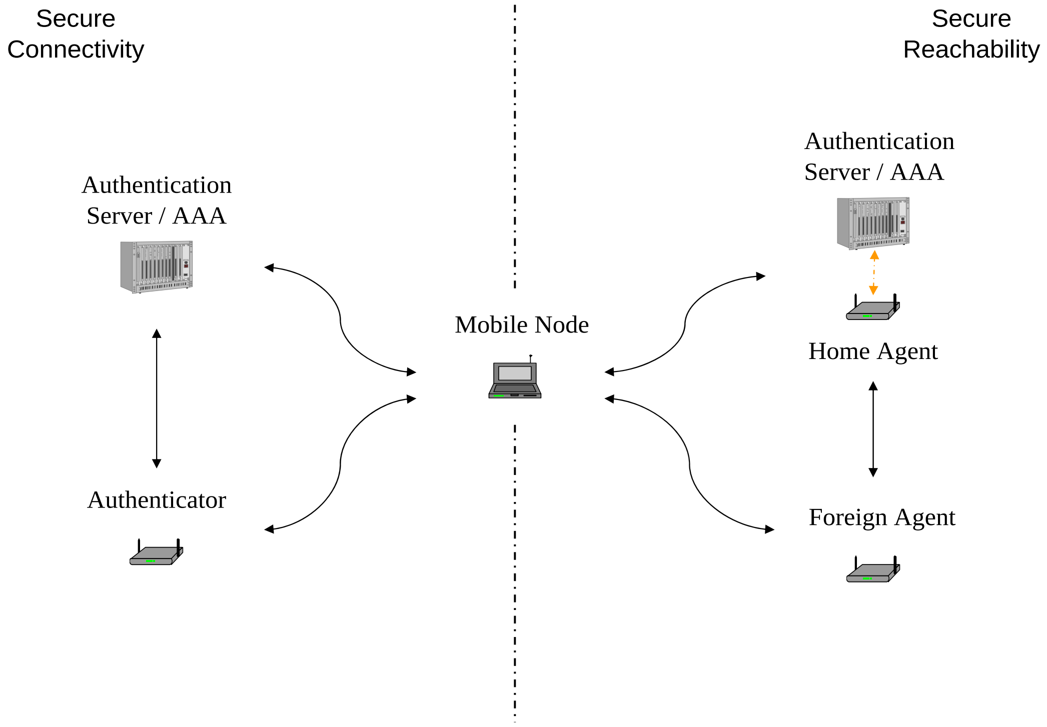


Figure 25: Trust relationship of handover security.

tor, the trust relationship is a temporary one which needs to be built on demand. The key management of this phase is based on the long term trust between MN and authentication server. Various session keys are derived from the master key of long term trust, and further distributed to corresponding entities to set up temporary trust relationship for the purposes of integrity and confidentiality protection.

In secure reachability phase, MN has a long term trust with its home network consist of home agent (HA) and AAA server. When MN starts exchanging mobility management messages with its home network, this long term trust is used to authenticate and authorize the MN. Within the home network, the trust relationship between HA and AAA server is guaranteed by the home network service provider to ensure the security and interoperability of different mobility schemes. The trust relationship between the foreign agent (FA) and HA is a dynamic one that requires authentication to be set up. In the case of IKEv2 protocol for IP mobility, IKEv2 enables the mutual authentication between FA and HA to set up the IKE SAs. The IKE SAs are used further to set up IPsec SAs to protect the transport of mobility management signaling. Between MN and FA, the trust relation requires authentication to be set up, and its establishment in real practice depends highly on the link technologies in use.

In recent heterogeneous networking environment with multiple access technologies and network vendors, security policies and agreements among different networks bring complexity to handover security management. Since different network vendors can choose different cryptographic algorithms and AAA protocols according to their needs, potential interoperability problems may arise when inter-technology handovers across different vendor networks occur.

As previous research has pointed out that handover security schemes can degrade performance due to the overhead introduced by authentication, key management, and cryptographic operations [ONY03, PK04, LW05, AW07], in order to understand the correlations between handover security and performance, we identify the following performance aspects that are affected by existing handover security schemes: handover completion time, throughput, and quality of service (QoS).

Firstly, the handover completion time is the period between the start and the end of a handover, which determines how fast a mobile node can obtain its connectivity and reachability in the network. The overall handover completion time is affected by the latency introduced by security implementations in terms of access authentication, SA establishment, AAA operation, encryption/decryption, and key management.

Among various factors leading to handover latency, access authentication is regarded as the most critical one contributing towards the performance degradation in a wireless network with high mobility user running real-time applications [AW07]. To guarantee strong protection, mutual authentication is adopted by a number of existing authentication protocols such as EAP-AKA and EAP-TLS. As described in the 2-Phase model, since more authentication messages are needed for mutual authentication, the extra processing time on transmission and identity validation will prolong the completion time of secure connectivity phase. This directly leads to the delay of handover completion time.

In the secure reachability phase, SA establishment is needed to distribute shared secrets to corresponding entities for cryptographic purposes. As to IP mobility, in order to negotiate IPsec SA, multiple rounds of IKEv2 exchange can lead to potential delay to complete mobility signaling. The AAA operation involved in both two phases is to enable authentication and authorization of each entity, while the AAA related database processing and request/response messages add latency to the handover completion time. For data integrity and confidentiality, encryption/decryption supports the safe transport of authentication and mobility management messages, but such operations can add up latency of handover. Meanwhile, the choice of algorithm and key length exerts huge impact on the security processing time. For instance, compared with symmetric key mechanisms, the public key methods consume more energy and require more time for computation [PK04]. Because cryptographic and authentication schemes rely on the secret keys, key management operations are essential to the handover security. Nevertheless, the time needed to generate, derive, and further distribute secret keys to corresponding entities prolongs the overall handover completion time.

Secondly, the negative impact on throughput comes mainly from the latency induced by handover security schemes. Because many existing transport protocols such as TCP will slow down the transmission rate while encountering excessive latency and potential packet losses that exceed the threshold, the reduced sending rate hence brings down the overall throughput for data transmission.

Thirdly, frequent handovers can generate a huge amount of security related traffic that take up the bandwidth reserved for data transmission, thus affecting both the throughput and the quality of service. In addition, as handover security schemes involve security policy and agreement among different networks, this complexity results in difficulty to guarantee a stable level of QoS for mobile services.

Concerning the latency introduced by handover security schemes, one major cause comes from the security related message exchange between the mobile node and its home network. In practice, the distance between the home network and the visiting network where MN currently locates could be far enough to generate huge propagation delay that can affect performance and user experience, especially for delay-sensitive applications such as VoIP. In addition, the possibility of packet loss could increse due to the long distance transmission across various domains. This can prolong the overall handover completion time due to the need of re-negotiation and retransmission of lost packets. Besides the propagation delay, mutual authentication and SA set up both require multiple rounds of message exchange that need to reach the home network, possibly crossing the Internet. The time required by processing and transmitting such messages contributes further to the latency and puts off the handover completion time.

Besides the performance aspects discussed, it should be noted that the impact of handover security can vary in different scenarios. When a handover occurs with an on-going communication session, e.g. VoIP call, the latency requirement is more critical, which requires the security schemes such as authentication and key management to minimize their overhead and negative effects on performance. Meanwhile, if there is no on-going session during a handover, the performance requirement in terms of latency is less critical, with more flexibility for security operations.

5 Securing Handover in Local Administrative Domain

In mobile and wireless networks, handover is a complex process that involves multiple layers of protocol and security executions. Nowadays, a great challenge faced by handover comes from the overhead of security implementations that affects performance. By studying the impacts of handover security, we propose our approach to seek a balance between handover security and performance in the security enhanced Local Administrative Domain (LAD).

5.1 Local administrative domain

According to our analysis of security impact, the lengthy delay introduced by security schemes constitutes a major part of performance degradation in handover. Besides the latency factor, authentication and SA negotiation bring communication overhead to the Internet in terms of signaling traffic. In addition, security related computations entail extra energy consumption and hence poses a huge challenge to existing mobile terminals with limited power supply and computing capability.

Given the existing challenges, careful investigation and design are needed in order to make a handover both secure and seamless. Taking into account the heterogeneity of access technologies and security policies in current networks, we propose the Local Administrative Domain (LAD) with localized optimization as our approach to seek a balance between handover security and performance in a well defined security enhanced domain.

As depicted in Figure 26, an LAD in our proposal is defined as an access domain consisting of a collection of sub-nets, network entities, and AAA databases under a common administration. The network entities operating under such administration can be assumed to share administratively created trust. Optimization can be implemented in the domain to enhance the performance and handover security.

A simplified handover scenario is also presented in Figure 26. When a mobile node

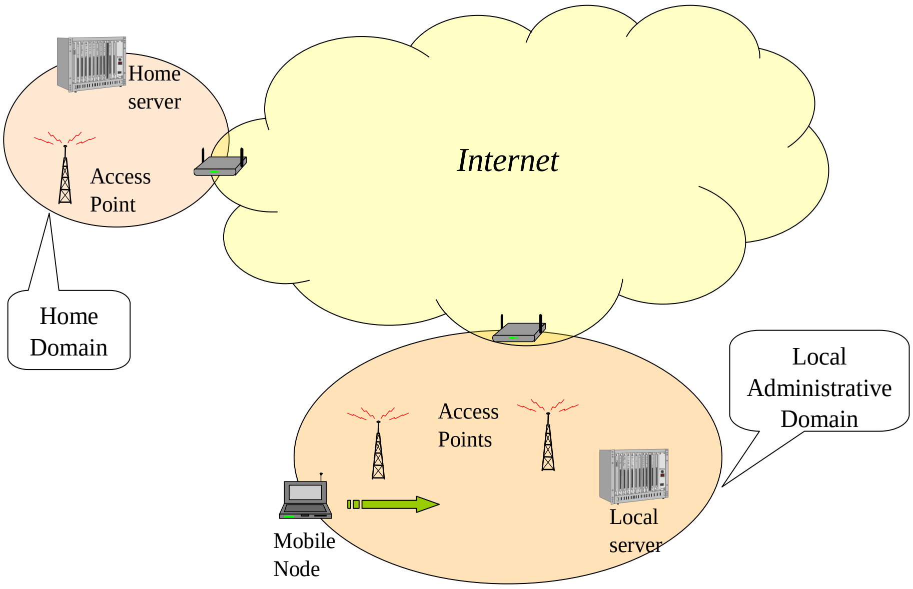


Figure 26: Handover Scenario in Local Administrative Domain.

leaves its home network and enters a LAD, handover occurs and is handled by the responsible entities in this local domain. When the mobile node changes its access point within the LAD, localized handover schemes will be used to guarantee a good level of security and performance.

At the moment, various standard development organizations (SDO) are working cooperatively to improve the handover security performance from different perspectives. Concerning IP mobility, Proxy Mobile IP (PMIP) improves the global mobility scheme by limiting the mobility management within the PMIP domain. By utilizing network-controlled handover method, PMIP can alleviate the overhead of mobility management signaling conducted over the wireless air link [G+08, LYC08, LF08]. For AAA operation, the hierarchical AAA approach is proposed to reduce the cost coming from remote authentication and hence improves the overall performance of handover [W+08].

Based on EAP framework, the IETF HOKEY group [HOK09] focuses on the localized re-authentication approach to eliminate the latency resulting from multiple rounds of authentication exchange between a mobile node and its home network with key distribution supports. The approach for root key derivation from Extended Master Session Key (EMSK) is proposed to enable the usage of domain-specific root key within specific key management domains [S+08]. In order to avoid repetition of full EAP exchange for frequent mobility with less computational overhead, EAP-AKA provides a solution to handle re-authentication when a mobile node changes its access authenticator [AH06]. The lately proposed EAP Re-authentication Protocol (ERP) defines an EAP based method-independent re-authentication scheme with keying hierarchy to support efficient local authentication [ND08].

To compare three EAP based authentication mechanisms, a typical access authentication scenario is depicted in Figure 27, in which four major entities are presented: mobile node, authenticator, EAP/ERP local server, and EAP/ERP home server. In this scenario, full authentication takes place between the mobile node and its home server, which requires the local server to communicate with the home server to finish EAP-AKA authentication. The fast re-authentication occurs when the mobile node changes its authenticator in the LAD. Both EAP-AKA and ERP are able to complete re-authentication without involving the home server. As shown in Figure 27, compared with EAP-AKA re-authentication, ERP requires only 1 round-trip time between the mobile node and local server to finish the fast re-authentication, which yields better performance.

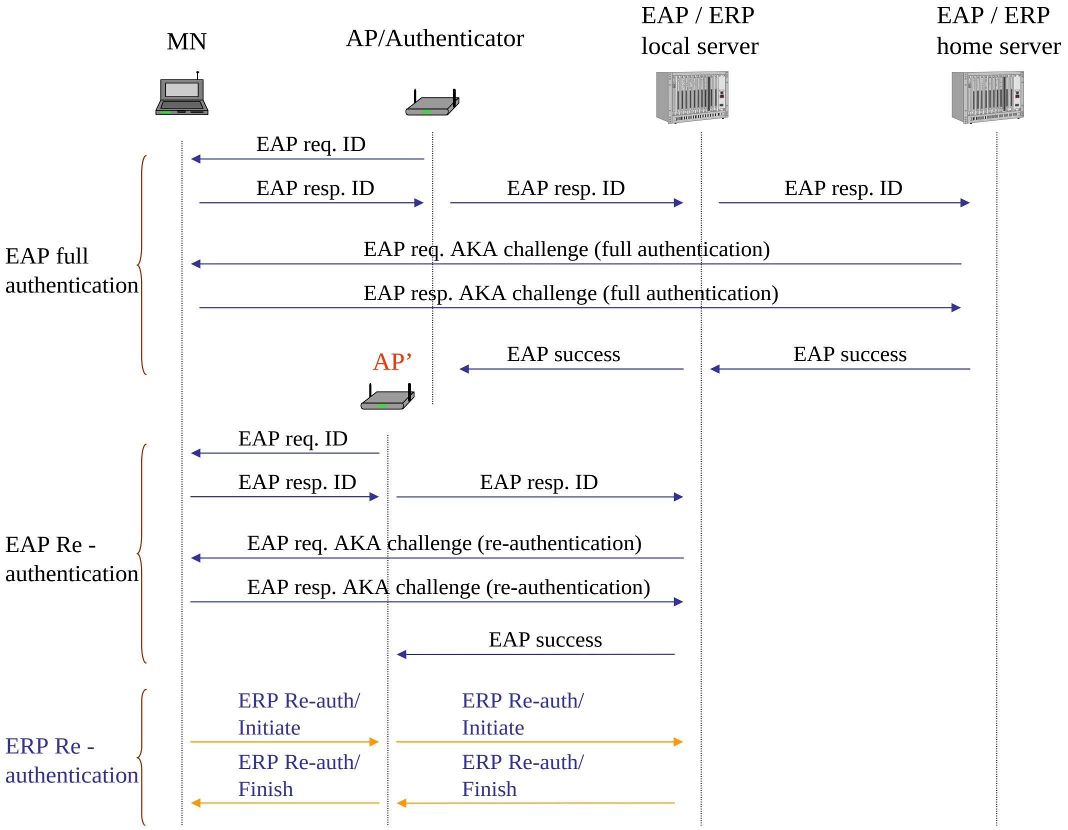


Figure 27: EAP-AKA and ERP re-authentication signalling [AH06, ND08].

Previous work to improve the performance of handover security forms a solid base for future design and development. However, as each solution targets at a specific part of handover security, there is a lack of comprehensive view that is supported by use case to utilize the existing mechanisms in a coherent manner.

Based on the two-phase handover security model, we integrate the following solutions to our proposed local administrative domain: First, for secure connectivity, EAP-AKA is used as the base to support mutual authentication and key management for access control, owing to its advantage of reusing 3G infrastructure. For scalability and performance, ERP protocol with its method-independent feature is adopted as the fast re-authentication mechanism for access authenticaiton to incorporate multiple authentication methods. Local EAP/AAA servers are used in the LAD to assist localized AAA operations. Second, for secure reachability, Proxy Mobile IP is adopted as the main solution to support wireless IP mobility. IKEv2 and IPsec are used for SA set up and cryptographic operations to protect mobility management signaling. For AAA operation, both RADIUS and DIAMETER are candidates to enable AAA functionality for interoperability and scalability concerns.

By utilizing solutions provided by SDOs, our proposal renders the following advantages: 1) by conducting access authentication and AAA operation within the LAD, the frequent security signaling across the Internet can be avoided, which can greatly reduce the propagation delay hence shortening the authentication completion time; 2) by using EAP-AKA and ERP re-authentication mechanism, less security signaling is needed, thus promoting the bandwidth usage of both wired and wireless links; 3) as EAP-AKA adopts symmetric key, less computation time is needed for authentication and key derivation, hence reducing the authentication and saving energy on the mobile terminal; 4) ERP provides a method-independent generic framework to support efficient re-authentication and distribution of EAP keying materials; 5) Proxy Mobile IP enables the mobility management without the involvement of mobile terminals, which greatly reduces the energy consumed by transmitting security and mobility related signaling over the wireless link; 6) the local AAA operation with key management enables the fast authentication and authorization procedure, which helps shorten the overall handover completion time.

It should be noticed that the cross-vendor security agreement is not concerned in our proposal at current phase, since such agreement involves commercial and social complexity and requires careful investigation and analysis. Our proposal of Local Administrative Domain targets at the existing handover security solutions and seeks a balance between security and performance in such a security enhanced multi-access domain. By enabling handover with satisfactory performance and security guarantee, our proposal offers a flexible and extensible structure open to future improvement and upgrade.

5.2 Performance evaluation

To analyze handover performance using the security schemes proposed in Local Administrative Domain (LAD), we implement two EAP based access authentication protocols, EAP-AKA and ERP [AH06, ALE09, ND08], in a simulation environment - the Network Simulator version 2 (ns-2) [NS 08]. Based on the security impact analysis, our experimentation reflects the performance improvement in secure connectivity phase, which yields a positive impact on the overall performance of handover.

5.2.1 Objective and methods

The main goal of our test is to evaluate the performance improvement gaining from the handover security schemes adopted in our proposal, the Local Administrative Domain (LAD). Our focus is on the access authentication latency given its critical impact on overall handover performance [AW07]. As our first step to achieve a balance between handover performance and security, we aim at promoting the understanding of how to make a handover in wireless mobile environment both secure and seamless.

As defined in our 2-Phase model, access authentication occurs in the secure connectivity phase. Given a typical roaming scenario when a mobile node is moving to a foreign domain, the full authentication messages often need to be transmitted to the home network over the Internet. Due to the long distance between foreign domain and home domain, the lengthy propagation delay and potential packet loss can cause performance degradation. As we adopt EAP-AKA and ERP in our proposed LAD, both protocols provide efficient localized re-authentication mechanisms to avoid excessive latency from frequent full authentications across the Internet. The localized optimization in LAD shortens the handover completion time, and hence leads to better handover performance.

To test the efficiency of fast re-authentication, we implement both EAP-AKA and ERP protocols in ns-2 and conduct performance tests covering two handover scenarios, as depicted in Figure 28. In scenario 1, handover occurs in the home domain with a base network topology including mobile node, access points, and EAP/ERP server. Scenario 2 covers the handover in a foreign LAD where full authentication messages are handled by home EAP/ERP server while localized re-authentication is handled by the local EAP/ERP server residing in the LAD.

In order to obtain results containing statistical significance, we conduct simulation test with 100 repetitions for each scenario under different parameters such as link bandwidth and Internet latency. By comparing the re-authentication of both EAP-AKA and ERP with the full authentication which is performed at connection set up time, we provide detailed analysis and means to estimate the authentication latency for each scheme.

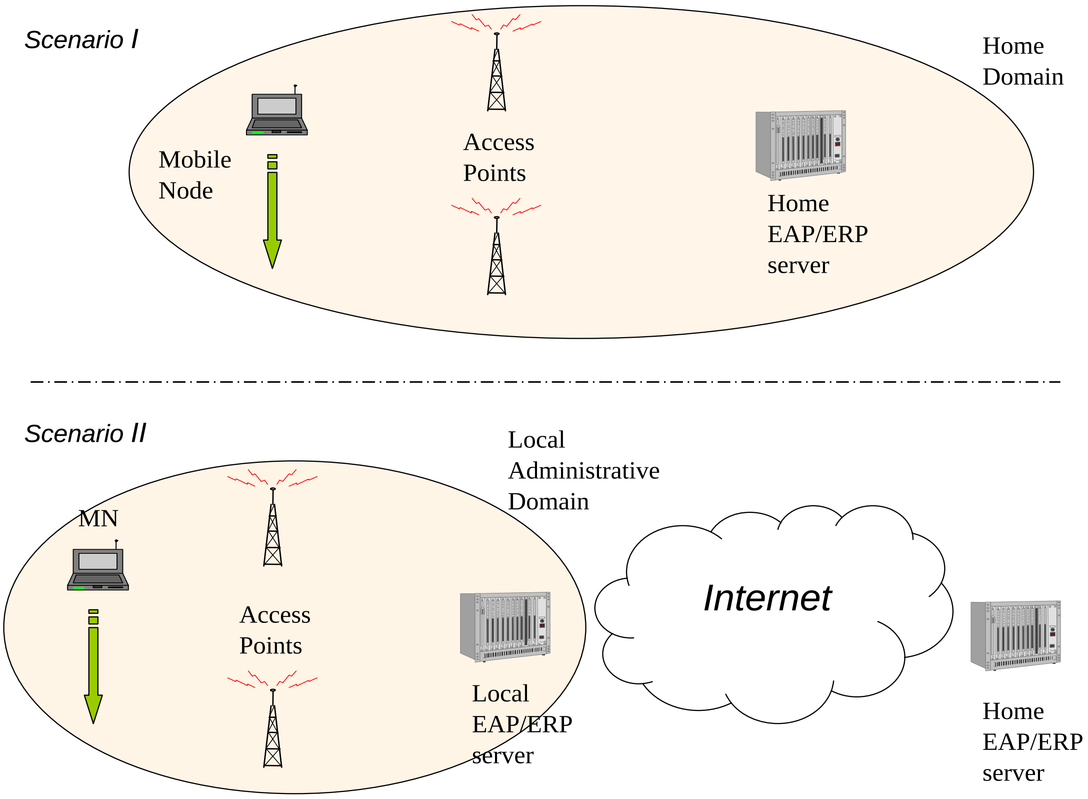


Figure 28: Test scenarios.

5.2.2 Simulation platform and protocol implementation

The simulation platform we use for testing and protocol implementation is the Network Simulator version 2 (ns-2). With the latest released version 2.34, ns-2 is an object-oriented and event driven network simulator developed in C++ and Object-oriented Tcl (OTcl) [OTc09]. Owing to its open and extensible architecture, ns-2 is currently used by numerous research projects as simulation platform for wired and wireless IP networks covering a wide range of protocols [NS 08].

In ns-2, different networking protocols are implemented as ns-2 network component objects such as ns agents and ns applications. An agent in ns-2 is a network object responsible for data transfer e.g. a TCP agent and UDP agent. On the other hand, an application in ns-2 is responsible for generating data traffic, e.g. a Telnet application and FTP application. Most of the existing protocols in ns-2 are implemented in C++ to achieve satisfactory execution performance. At the same time, in order to offer a convenient interface for ns-2 users to set up network topology and run simulation tests, ns-2 adopts OTcl script interpreter and network set up modules.

As an event driven simulator, ns-2 deploys the event scheduler to manipulate events. For ns-2, an event contains a scheduling time, a unique packet identification, and an pointer to a event handler. Each ns-2 packet identified by its event ID is thus associated with a scheduling time with one pointer to an object that handles the event. The ns-2 event scheduler is responsible for keeping track of simulation time and scheduling events in the event queue. When the scheduled time for an event arrives, the event scheduler will invoke appropriate network components pointed by this event to fulfil corresponding operations.

In order to control network components coded in C++, ns-2 uses the OTcl linkage to build a connection between C++ objects and corresponding OTcl objects. The linkage concept in ns-2 is illustrated in Figure 29 together with a general view of ns-2 simulation. Because OTcl linkage matches each C++ object with its OTcl object, OTcl interpreter can handle and control network objects compiled in C++ via the linkage. This linkage feature in ns-2 offers not only a clear interface to its users, but also an extensible framework for protocol developers to add new protocols to ns-2.

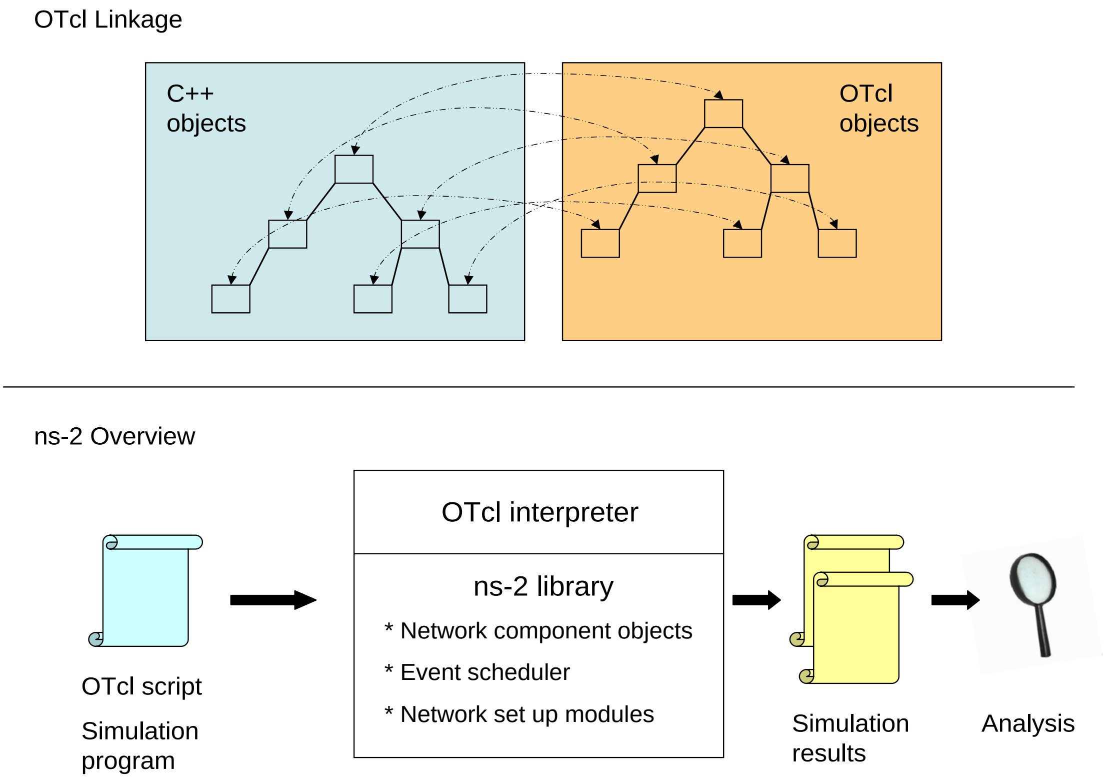


Figure 29: OTcl linkage and ns-2 overview. [CC09]

As shown in Figure 29 for a general view of ns-2 simulation, the OTcl script language and OTcl interpreter together provide a fast prototyping feature for ns-2 users to set up simulation test. By linking various ns objects and libraries to form a coherent framework, ns-2 can generate simulation results according to users' will and requirement. In addition, ns-2 regulates and forms the simulation results in a proper format ready for data analysis.

In order to support protocol implementation, ns-2 defines its own packet format which consists of a stack of headers and optional data space. The header stack in a packet contains all registered headers such as IP header, TCP header, trace header, and headers added by developers. The optional data space is used by ns-2 application for data traffic. Each header contains protocol specific information fields and can be accessed by network objects such as ns-2 agent using the corresponding offset value.

Our implementation for access authentication protocols in ns-2 consists of two major parts: a new packet header containing protocol related information, and a new agent class defining protocol specific operations. The format of new header and protocol behavior defined by our agent are illustrated in Figure 30.

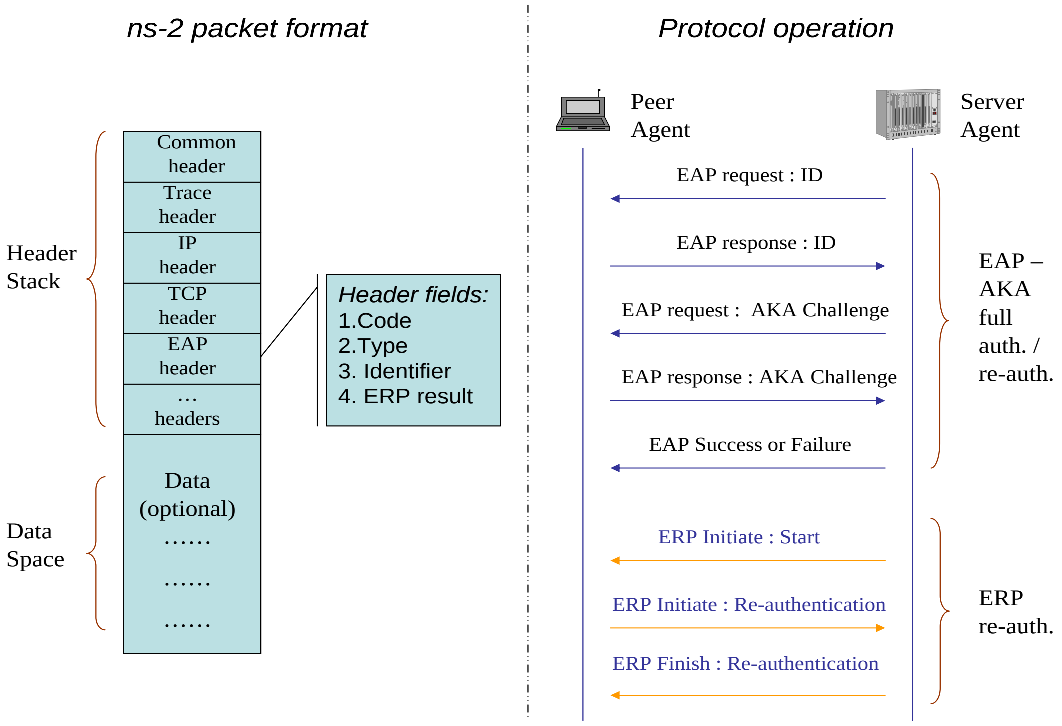


Figure 30: Packet format and protocol operation.

First, we create a new header containing EAP related information. The new header is registered in ns-2 with an pre-defined offset, so that ns-2 agent can access this header to manipulate protocol fields. As shown in Figure 30, we define four EAP fields in our new header: Code, Type, Identifier, and ERP Result.

In our new header, the Code field is used to identify the type of each EAP packet with six types of values: value 1 represents EAP Request; value 2 represents EAP Response; value 3 represents EAP Success; value 4 represents EAP Failure; value 5 represents ERP Init; and value 6 represents ERP Finish. The Type field is used to indicate the EAP method information with four types of values: value 1 represents the EAP-AKA Identity; value 2 represents the EAP-AKA Challenge; value 3 represents the ERP Start; and value 4 represents the ERP Re-authentication. The Identifier field is used to synchronize the behavior of sender and receiver by requiring that the Identifier of EAP Request must match the current outstanding EAP Response. The ERP Result field is used to indicate the re-authentication result of an ERP exchange.

Second, based on the header format we implement a new ns-2 agent to transfer authentication packets following the protocol specifications [A^+04, AH06, ALE09, ND08]. As illustrated in Figure 30, both EAP-AKA and ERP protocols are implemented in our agent.

For EAP-AKA protocol, the message flow of re-authentication is the same as the full authentication. The difference between EAP-AKA re-authentication and full authentication is illustrated by the handover scenario in LAD, since the re-authentication is handled by local server in LAD, while the full authentication needs to be handled by the home server in the home domain. Upon receiving authentication packet, the agent will check the values contained in each EAP header field, make necessary modifications, and send out packet according to the evaluation.

For ERP protocol, we implement the complete ERP re-authentication procedure in which the server side starts the ERP Init with Type field set as ERP start. Following the ERP protocol specification, three authentication messages are needed to fulfil re-authentication, including the ERP Init - Start message to initiate ERP exchange, and two following messages ERP Init and ERP Finish for re-authentication.

Because our new EAP agent is implemented in C++ as a network object, we use OTcl linkage to bind this object to its corresponding OTcl object, so that ns-2 users can call the agent using OTcl script via the OTcl interpreter. For testing, we also link the network parameters such as propagation delay and packet size to the OTcl space to enable dynamic parameter change on demand.

Concerning the cryptographic operations involvded in key management and AAA operation that are carried out by our agent, we adopt an indirect way for implementation. The computation time for such cryptographic operations are emulated by adding a delay to the agent on the processing of corresponding messages. By doing this, we reveal the computation latency involved in those operations and logically combine them into our protocol operations.

It should be noticed that our implementations of EAP-AKA and ERP protocols are slightly modified to conform to the design requirement of ns-2 extension. Therefore we do not cover all the features and scenarios specified in the protocol specifications [AH06, ALE09, ND08]. For instance, retransmission and fall-back behavior are not included in the current agent implementation. The reason for our simplified protocol implementation is that we aim at obtaining direct and priliminary results as the first step. Due to the time limit and implementation consideration, protocol features that are not included in current version will be added to the future implementation. Nevertheless, our implementation contains the essential parts required by each access authentication scheme. By covering such key features, our ns-2 implementations are capable of reflecting correct protocol behavior in the simulation environment.

5.3 Test arrangement

To measure the authentication latency of three target EAP based schemes, EAP-AKA full authentication, EAP-AKA re-authentication, and ERP re-authentication, we set up a test environment in ns-2 version 2.34 [NS 08] with our EAP/ERP agent and packet header deployed on corresponding entities. The network topology for our simulation is illustrated in Figure 31. As shown in the figure, our test covers two major scenarios. As discussed in Section 5.2.1, scenario 1 targets at performance evaluation of three access authentication schemes in the home domain by comparing the latency values obtained from each scheme. Scenario 2 is to test the performance improvement for handover security in LAD by comparing the EAP-AKA full authentication with other two localized authentication schemes.

In scenario 1, three types of entities form the test network: one mobile node (MN), two authenticators functioning as access points for the MN, and one EAP/ERP home server handling authentication operations. We deploy EAP/ERP agents on the

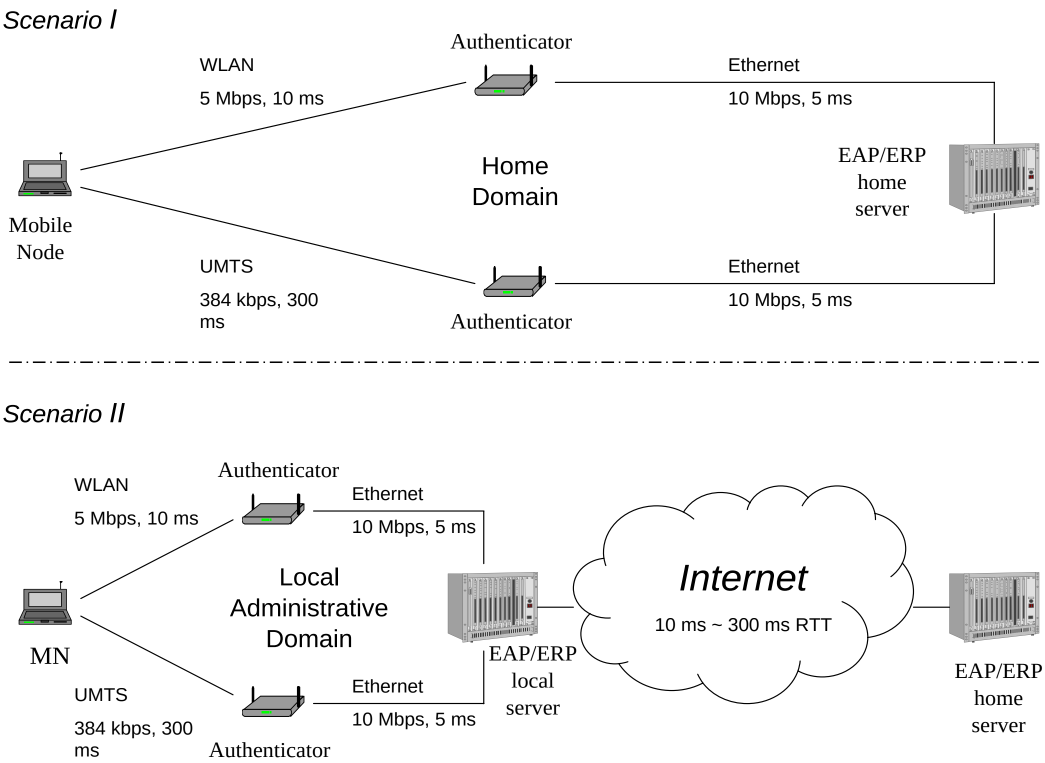


Figure 31: Network topology for performance test.

MN and server to fullfil authentication process following the protocol specifications. From mobile node to authenticators, two types of access links are simulated for wireless connectivity. The transmission rate for WLAN link is set to 5 Mbps with 10 ms as propagation delay. For UMTS link the rate is set to 384 kbps with 300 ms as propagation delay. Between authenticators and EAP/ERP server, wired connections are simulated as ethernet links with 10 Mbps transmission rate and 5 ms propagation delay.

Scenario 2 consists of two domains: the home domain with one EAP/ERP server and the LAD containing one MN, two authenticators, and one local EAP/ERP server. The characteristics of links inside LAD follow the same setting as used in scenario 1. As shown in Figure 31, the LAD domain and home domain are separated by the IP based Internet connection. For EAP-AKA full authentication, each message needs to be transferred between the local server and home server across the Internet. To simulate the impact of Internet connection, we assign the link latency with a series of selected values ranging from 10 ms to 300 ms.

Because the Internet latency parameter plays a critical role in the test, the values selected for our simulation need to reflect the condition in real networking environment. As an example, assuming the home domain for a mobile node is in Helsinki while the mobile node is handovered to a LAD in France at the moment, this mobile node needs to conduct full authentication acrossing the Internet between France and Helsinki in order to obtain access right to the visiting network. In this example, re-authentication schemes require only authentication exchange within such LAD in France. To make our simulation realistic, we obtain 6 sets of values from the real network by running the PING program provided by Linux operating system (kernel version 2.6.27) to probe 6 different locations with physical distance variance. The values collected are summarized in Table 1 containing four fields: location, round-trip time (RTT), mean deviation, and selected probing sites.

For each latency value presented in the table, we run the probing test with 100 repetitions. Each time the PING program sends 50 ICMP Echo Requests to the corresponding destination with message size equal to the EAP-AKA full authentication message. The probing test is done at Department of Computer Science, University of Helsinki. As shown in the table, the target destinations cover 2 European cities (Uppsala, Sophia Antipolis), 2 American cities (New York, Stanford), and 2 Asian cities (Beijing, Tokyo). The choice of locations exhibits diversity of distance, which yields a good coverage for latency values. The set of 'Mean RTT' values in Table 1 are used in our test as the link latency parameter to simulate the Internet latency.

Concerning another key parameter in our test, the AAA related cryptographic computation time required by key management and AAA operations, we assign a delay value to our EAP/ERP agent during its processing of authentication messages to

Table 1: Internet latency probing results

Location	Mean RTT(ms)	Mean Deviation(ms)	Probing Site
Uppsala, Sweden	9.99	0.29	www.uu.se
Sophia Antipolis, France	58.20	0.30	www-sop.inria.fr
New York, U.S.A.	114.53	0.35	www1.cs.columbia.edu
Stanford, U.S.A.	206.66	0.66	www-cs.stanford.edu
Beijing, China	197.97	0.73	www.tsinghua.edu.cn
Tokyo, Japan	306.51	0.81	www.u-tokyo.ac.jp

simulate this computation effect. Considering the load variation on AAA server in real environment, we select the ns-2 random number generator to obtain a series of independent delay values for our test. Based on previous research [K^{+}06, Kor08, I^{+}08], the value for EAP-AKA full authentication is generated under exponential distribution with 500 ms as the mean value, ranging from 200 ms to 9000 ms. The value for re-authenticaiton schemes is generated under exponential distribution with 300 ms as the mean value, ranging from 45 ms to 9000 ms. Following such setting, we generate 100 independent random numbers for EAP-AKA full authentication and use them as the input of Internet latency parameter in each test repetition. In order to compare the performance between EAP-AKA re-authentication and ERP re-authentication, we choose to generate 100 independent random numbers and use the same number set for both schemes in our measurement.

For the size of authentication message in our simulation, given that the cryptographic keys used by authentication schemes vary in different scenarios according to security level and system requirement, we select 178 bytes as the mean value for EAP-AKA messages and 132 bytes for ERP messages [PK04, ND08].

The performance metric we measure is the authentication latency, which is calculated as the time between the start of first authentication message and the receival of the final authentication message. Since our test does not cover retransmission behavior, we simulate all the connections as error free links in order to obtain direct values of authentication latency.

Following the test design, we measure the authentication latency in our simulation environment for three access authentication schemes based on the protocol behavior illustrated in Figure 30. For scenario 1, we allocate the EAP/ERP agents on mobile node and home server to exchange authentication messages. The full authentication and re-authentication in scenario 1 involve mobile node, authenticator, and home server. Our test covers three authentication schemes conducted over two types of wireless access technologies - UMTS and WLAN, thus yielding two test categories in scenario 1.

For scenario 2, the full EAP-AKA authentication involves the mobile node, authenticator, and the home EAP/ERP server across the simulated Internet connection, while the re-authentication is handled by the local EAP/ERP server within the LAD. We therefore allocate two pairs of agents on the corresponding entities, with one pair connecting the mobile node and the home EAP/ERP server and another pair connecting the mobile node and the local EAP/ERP server. The test in scenario 2 covers two types of wireless access technologies and six different values for Internet latency. For Internet latency, we adopt the 'Mean RTT' values from Table 1 as the link latency in test. Three authentication schemes are tested against twelve cases, with 6 cases for UMTS access and 6 for WLAN access, respectively.

All the test cases in both scenarios are repeated over 100 times by taking the aforementioned random number as the input for AAA related computation latency on each repetition. In each test case, one type of authentication scheme is simulated which involves message exchange between the mobile node and the EAP/ERP server to complete access authentication. The ns-2 simulator keeps track of the essential networking events and records them into the trace file for further performance analysis.

5.4 Test results

Based on the analysis of trace files generated by ns-2, we obtain a series of measurement results on authentication latency according to our test arrangement for two scenarios. Since the test setting of scenario 1 is different from that of scenario 2, we discuss and analyze two scenarios separately.

For scenario 1, we focus on the performance comparison of three authentication schemes occurring in the home domain. The simulation results are summarized in Table 2, covering two categories for UMTS access and WLAN access, representing the slow link with 300 ms delay and the fast link with 10 ms delay, respectively. We provide four fields in the table to compare different aspects for each scheme.

The 'Total Latency' field shows the mean time required to complete the authentication associated with a standard deviation obtained from 100 test repetitions. As the value in 'Total Latency' includes two parts, the transmission time for authentication messages and the time spent on AAA related cryptographic computations, we provide two fields to indicate the percentage of each part in the total latency. The 'RTT % of Total' shows the percentage of transmission time required for authentication messages against the total latency. The value of 'RTT % of Total' is calculated by dividing the transmission time to the mean value of corresponding total latency in each case. At the same time, the 'AAA % of Total' shows the percentage of time spent on AAA related computations in the total latency, which is obtained by subtracting the value of 'RTT % of Total' by 1. The 'No. of Msg' shows the total number of messages required to complete each authentication scheme.

From the values in ’Total Latency’ for scenario 1, we can tell that the performance of ERP protocol is better than EAP-AKA re-authentication and full authentication. This performance improvement comes mainly from two parts: 1) ERP requires only three messages to complete authentication, compared to the five messages required by other two schemes; 2) the re-authentication requires less time to finish the AAA related computation. For both UMTS and WLAN, the authentication latency of ERP is at least 40% less than that of EAP-AKA full authentication. Compared with EAP-AKA re-authentication, ERP requires 30% less authentication time in UMTS access, and for WLAN access it requires 10% less time.

Since in scenario 1 all the authentication messages are handled inside the home domain, the propagation delay from wireless access exerts a huge impact on authentication. As shown in Table 2, the total latency values in the UMTS category are much higher than the ones in WLAN access for all three types of authentication schemes. This impact is also reflected in the RTT percentage field. For the UMTS access which has longer propagation delay than the WLAN access, the RTT required to exchange authentication messages takes a large portion of total latency, with the highest value up to 81.7% under the EAP-AKA re-authentication. On the other hand, the effect of AAA related computation is obvious in the WLAN case, with the AAA percentage value above 80% for all three schemes. This implication is due to that in WLAN case less time is required for message exchange.

As aforementioned, in order to compare the performance of EAP-AKA re-authentication and ERP re-authentication, we use the same set of random numbers as AAA latency

Table 2: Test results of scenario 1

UMTS access				
Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	2215.72 ± 448.88	69.7	30.3	5
EAP-AKA re-auth.	1889.31 ± 281.21	81.7	18.3	5
ERP re-auth.	1266.60 ± 281.48	72.9	27.1	3

WLAN access				
Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	748.59 ± 448.24	10.3	89.7	5
EAP-AKA re-auth.	422.40 ± 281.47	18.3	81.7	5
ERP re-auth.	388.81 ± 280.98	11.8	88.2	3

parameter for both schemes. This is reflected in our results as the standard deviation of total latency for both schemes are very close and fall in the range between 280 ms to 282 ms.

In scenario 2, we target at comparing the performance of localized authentication schemes adopted in LAD against the EAP-AKA full authentication. Given our test design, we divide the measurement results into two main categories in terms of wireless access. For each category, six test cases are covered using six different Internet latency values. We adopt both figures and tables to assist our analysis, with the figures conveying the authentication latency of three schemes under different Internet latency values, and the tables summarizing the measurement results using the same fields as in Table 2.

Concerning the category of UMTS access, we present the results in Figure 32 and Table 3 covering six cases. For performance comparison, the diagram in Figure 32 depicts the mean values of total authentication time required for each authentication mechanism under six different Internet latency parameters. As shown in the figure, ERP re-authentication yields better performance than both EAP-AKA full authentication and EAP-AKA re-authentication owing to its localized optimization in LAD.

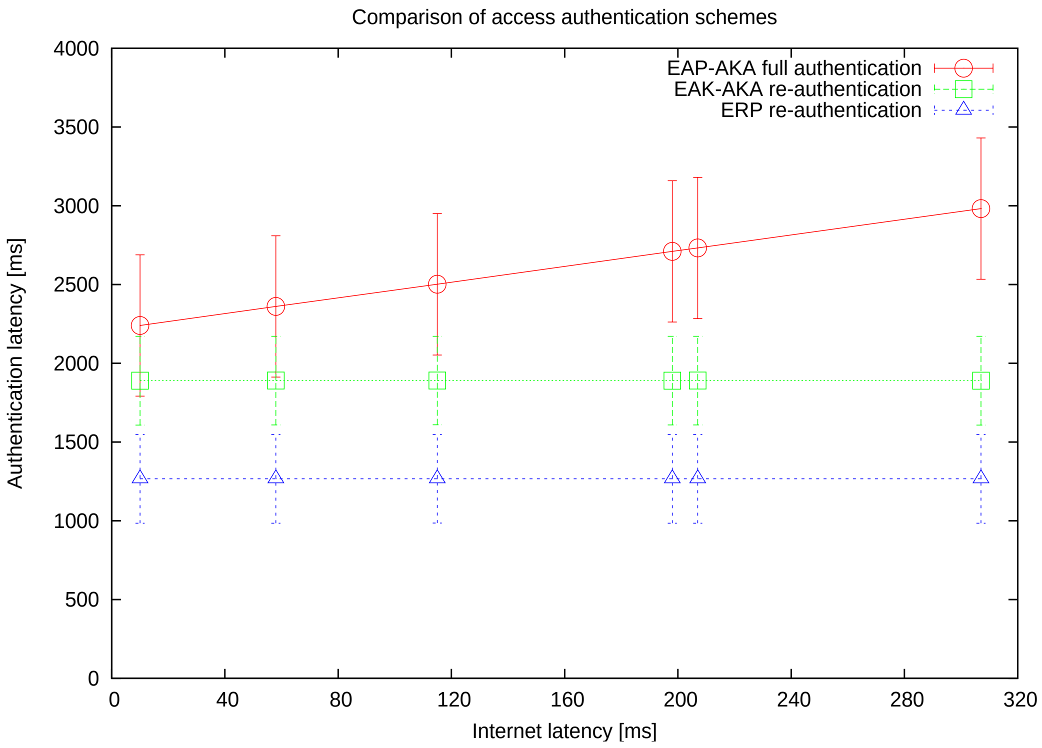


Figure 32: Test results of scenario 2, UMTS access.

Table 3: Test results of scenario 2, UMTS access

Case 1: 10 ms latency

Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	2240.23 ± 448.61	70.0	30.0	5
EAP-AKA re-auth.	1889.90 ± 281.92	81.7	18.3	5
ERP re-auth.	1266.59 ± 281.40	72.9	27.1	3

Case 2: 58 ms latency

Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	2360.66 ± 448.45	71.6	28.4	5
EAP-AKA re-auth.	1890.09 ± 281.30	81.7	18.3	5
ERP re-auth.	1266.47 ± 281.43	72.9	27.1	3

Case 3: 115 ms latency

Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	2501.54 ± 448.75	73.2	26.8	5
EAP-AKA re-auth.	1890.09 ± 280.92	81.7	18.3	5
ERP re-auth.	1266.54 ± 281.15	72.9	27.1	3

Case 4: 198 ms latency

Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	2710.15 ± 448.57	75.2	24.8	5
EAP-AKA re-auth.	1889.82 ± 281.38	81.7	18.3	5
ERP re-auth.	1266.71 ± 281.48	72.9	27.1	3

Case 5: 207 ms latency

Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	2731.97 ± 448.42	75.4	24.6	5
EAP-AKA re-auth.	1890.10 ± 281.19	81.7	18.3	5
ERP re-auth.	1266.64 ± 281.41	72.9	27.1	3

Case 6: 307 ms latency

Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	2981.86 ± 448.33	77.5	22.5	5
EAP-AKA re-auth.	1889.53 ± 281.49	81.7	18.3	5
ERP re-auth.	1266.71 ± 281.48	72.9	27.1	3

Because full authentication requires communication with the home domain across the Internet, the total latency for EAP-AKA full authentication grows along with the augment of Internet latency. As re-authentication schemes are handled locally, their latency measurement results are not affected by the Internet delay parameter compared with full authentication. Comparing two re-authentication schemes from EAP-AKA and ERP, since ERP requires less messages to complete its authentication, under the slow access link such as UMTS, the latency value for ERP is 30% less than that of EAP-AKA re-authentication.

From Table 3, we observe a steady increase in the 'RTT % of Total' for EAP-AKA full authentication from case 1 to case 6 with the highest value up to 77.5 % in the total letency. Due to the long propagation delay over wireless link in this category, the 'RTT % of Total' values are high for all three schemes with the highest value achieved by EAP-AKA re-authentication as 81.7 %. Because the time required to exchange authentication message takes long under slow wireless link, the 'AAA % of Total' constitutes a smaller portion of total latency in this category.

Concerning the WLAN access, we present our measurements in Figure 33 and Table 4. From the diagram depicted in Figure 33, we observe that the performance of

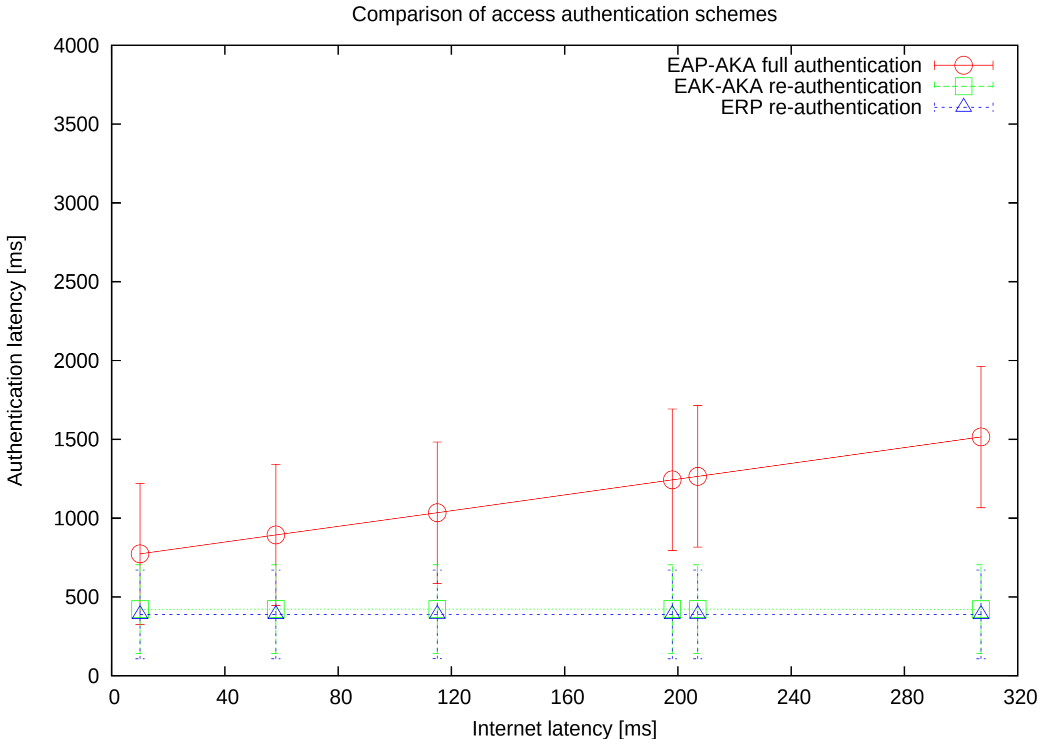


Figure 33: Test results of scenario 2, WLAN access.

Table 4: Test results of scenario 2, WLAN access

Case 1: 10 ms latency				
Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	772.98 ± 448.35	13.2	86.8	5
EAP-AKA re-auth.	422.34 ± 281.15	18.3	81.7	5
ERP re-auth.	388.95 ± 281.35	11.8	88.2	3
Case 2: 58 ms latency				
Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	893.71 ± 448.16	24.9	75.1	5
EAP-AKA re-auth.	422.45 ± 281.33	18.3	81.7	5
ERP re-auth.	388.81 ± 281.35	11.8	88.2	3
Case 3: 115 ms latency				
Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	1034.05 ± 448.72	35.1	64.9	5
EAP-AKA re-auth.	422.67 ± 280.82	18.3	81.7	5
ERP re-auth.	389.53 ± 281.41	11.8	88.2	3
Case 4: 198 ms latency				
Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	1243.20 ± 448.76	46.0	54.0	5
EAP-AKA re-auth.	423.22 ± 280.91	18.3	81.7	5
ERP re-auth.	388.90 ± 281.57	11.8	88.2	3
Case 5: 207 ms latency				
Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	1264.89 ± 448.67	46.9	53.1	5
EAP-AKA re-auth.	422.72 ± 281.01	18.3	81.7	5
ERP re-auth.	389.19 ± 281.45	11.8	88.2	3
Case 6: 307 ms latency				
Auth. Type	Total Latency(ms)	RTT % of Total	AAA % of Total	No. of Msg
EAP-AKA full	1514.83 ± 448.80	55.7	44.3	5
EAP-AKA re-auth.	422.36 ± 281.27	18.3	81.7	5
ERP re-auth.	388.90 ± 281.42	11.8	88.2	3

ERP is better than EAP-AKA full authentication and re-authentication. Another interesting observation is that with fast access link, the performce of EAP-AKA re-authentication is very close to that of ERP. This mainly comes from the fact that both two schemes adopt the localized approach for authentication, and the tranmission delay in this case is comparably insignificant to the total latency. This can be observed also by comparing the diagrams in Figure 3 and 4 that the latency gap between ERP and EAP-AKA re-authentication in UMTS case is wider than that in WLAN case. Since EAP-AKA re-authentication requires two more messages to complete the authentication than ERP, under slow link access such as UMTS, the effect of long propagation delay is obvious to prolong the total latency.

As shown in Table 4, the Internet latency has huge impact on the total authentication time for full authentication scheme, with the 'Total Latency' value growing from 772.98 ms to 1514.83 ms regarding to case 1 and case 6, respectively. This effect is also indicated in the 'RTT % Total' field, as the value increases from 13.2% to 55.7% for full authentication. Another observation from the table is that under fast wireless access, the time required for AAA related computation occupies a large part of total latency, with the highest value up to 88.2% for ERP scheme.

It should be noticed that in order to compare three schemes across different scenarios, we apply the same set of random numbers generated by ns-2 to each authentication scheme for all 12 test cases. This ensures that the values for AAA related computation latency follow the same distribution in different scenarios for each scheme and enables us to compare the performance of each scheme under different test parameters such as the Internet latency.

In summary, the localized mechanisms yield better performance comparing to the counterpart without localized optimization for access authentication as expected. The performance of access authentication depends on three major factors including the characteristics of access links, the Internet latency, and the AAA related computation time.

6 Conclusions and Future Work

In wireless and mobile networks, handover security enables the protection of integrity, confidentiality, and availability of user credentials and network resources during a handover. The security impact from authentication, key management, and cryptographic operations affects the handover performance.

To understand the interaction between performance and handover security, we divide the process of handover security into two phases supported by our in-depth analysis on trust relationships and performance aspects that are affected by existing security schemes. As an attempt to seek a balance between handover security and performance, we propose the local administrative domain with localized security optimization to promote handover performance.

In summary, the contributions of this thesis include the 2-Phase model for handover security, the design of local administrative domain, and our protocol implementations in ns-2 for performance measurement. Because handover security is a challenging topic, our analysis on the impact of handover security is hence valuable as a guideline to promote our understanding and shed lights on the design and development of future handover security schemes.

As our first step to understand the security impact on handover, we measure and compare the performance of different access authentication schemes occurring in the secure connectivity phase. In the future we plan to extend our work to the secure mobility phase and implement mobility protocols such as Proxy Mobile IP in ns-2.

References

3GP06 3GPP2, Fast handoff for HRPD. 3GPP2 X.P0043-0 version 1.0, 3GPP2, November 2006.

3GP07a 3GPP, 3GPP system architecture evolution: Architecture enhancements for non-3GPP accesses. Release 8. 3GPP TS 23.402 1.3.0, Technical specification group services and system aspects, September 2007.

3GP07b 3GPP, GPRS tunneling protocol (GTP) across the GN and GP interface (release 7). Stage 3. 3GPP TS 23.060 7.5.0, General packet radio service (GPRS), March 2007.

3GP08a 3GPP, 3GPP forum, 2008. `http://www.3gpp.org/`. [06.11.2009]

3GP08b 3GPP, Requirements for further advancements for E-UTRA (LTE-Advanced), 2008. `http://www.3gpp.org/FTP/Specs/html-info/36913.htm`. [22.10.2009]

3GP09 3GPP, 3G Security; Security Architecture. Release 8. 3GPP TS 33.102

8.3.0, Technical specification group services and system aspects, June 2009.

80204 802.1X, I., Local and Metropolitan Area Networks: Port-Based Network Access Control. IEEE Standard 802.1X-2004, Institute of Electrical and Electronics Engineers, December 2004.

A$^+$04 Aboba, B. et al., Extensible Authentication Protocol (EAP). RFC 3748, Internet Society, June 2004.

AH06 Arkko, J. and Haverinen, H., Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA). RFC 4187, Internet Society, January 2006.

ALE09 Arkko, J., Lehtovirta, V. and Eronen, P., Improved Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA'). RFC 5448, Internet Society, May 2009.

AW07 Agarwal, A. K. and Wang, W., On the Impact of Quality of Protection in Wireless Local Area Networks with IP Mobility. *Mobile Networks and Applications*, 12,1(2007), pages 93–110.

AXM04 Akyildiz, I., Xie, J. and Mohanty, S., A Survey of Mobility Management in Next-Generation All-IP-Based Wireless Systems. *Wireless Communications*, 11,4(2004), pages 16–28.

B$^+$07 Barker, P. et al., Recommendation for Key Management - Part 1: General (Revised). NIST Special Publication 800-57, NIST, March 2007.

BH07 Buttyan, L. and Hubaux, J.-P., *Security and Cooperation in Wireless Networks*. Cambridge University Press, 2007.

Bis02 Bishop, M., *Computer Security: Art and Science*. Addison-Wesley Professional, 2002.

C$^+$02 Campbell, A. et al., Comparison of IP micromobility protocols. *Wireless Communications*, 9,1(2002), pages 72–82.

C$^+$03 Calhoun, P. et al., Diameter Base Protocol. RFC 3588, Internet Society, September 2003.

C$^+$05a Calhoun, P. et al., Diameter Mobile IPv4 Application. RFC 4004, Internet Society, August 2005.

C$^+$05b Calhoun, P. et al., Diameter Network Access Server Application. RFC 4005, Internet Society, August 2005.

C$^+$06 Caro, G. et al., A Cross-Layering and Autonomic Approach to Optimized Seamless Handover. *In Proceedings of the 3rd Annual Conference on Wireless On Demand Network Systems and Services*, January 2006.

C$^+$08 Clancy, T. et al., Handover Key Management and Re-Authentication Problem Statement. RFC 5169, Internet Society, March 2008.

CC09 Chung, J. and Claypool, M., NS by Example, 2009. `http://nile.wpi.edu/NS/`. [21.10.2009]

CD05 Choi, J. and Daley, G., Goals of Detecting Network Attachment in IPv6. RFC 4135, Internet Society, August 2005.

CDM08 CDMA2000, CDMA Development Group, 2008. `http://www.cdg.org/`. [30.10.2009]

CKK02 Chiussi, F., Khotimsky, D. and Krishnan, S., Mobility Management in Third-Generation All-IP Networks. *Communications Magazine*, 40,9(2002), pages 124–135.

DE07 Devarapalli, V. and Eronen, P., Mobile IPv6 Operation with IKEv2 and the Revised IPsec Architecture. RFC 4877, Internet Society, April 2007.

DE08 Devarapalli, V. and Eronen, P., Secure Connectivity and Mobility Using Mobile IPv4 and IKEv2 Mobility and Multihoming (MOBIKE). RFC 5266, Internet Society, June 2008.

DH98 Deering, S. and Hinden, R., Internet Protocol, Version 6 (IPv6) Specification. RFC 2460, Internet Society, December 1998.

DR08 Dierks, T. and Rescorla, E., The Transport Layer Security (TLS) Protocol Version 1.2. RFC 5246, Internet Society, August 2008.

E$^+$07 Emmelmann, M. et al., Moving Toward Seamless Mobility: State of the Art and Emerging Aspects in Standardization Bodies. *Wireless Personal Communications*, 43,3(2007), pages 803–816.

EHZ05 Eronen, P., Hiller, T. and Zorn, G., Diameter Extensible Authentication Protocol (EAP) Application. RFC 4072, Internet Society, August 2005.

EN02 Endler, M. and Nagamuta, V., General Approaches for Implementing Seamless Handover. *In Proceedings of the second ACM International Workshop on Principles of Mobile Computing*, August 2002, pages 17–24.

Ero06 Eronen, P., IKEv2 Mobility and Multihoming Protocol (MOBIKE). RFC 4555, Internet Society, June 2006.

G$^+$00 Glass, S. et al., Mobile IP Authentication, Authorization, and Accounting Requirements. RFC 2977, Internet Society, October 2000.

G$^+$08 Gundavelli, S. et al., Proxy Mobile IPv6. RFC 5213, Internet Society, August 2008.

GSM08 GSM, GSM World, 2008. `http://www.gsmworld.com/`. [30.10.2009]

HA07 Housley, R. and Aboba, B., Guidance for Authentication, Authorization, and Accounting (AAA) Key Management. RFC 4962, Internet Society, July 2007.

HO09 Hoeper, K. and Ohba, Y., Distribution of EAP based keys for handover and re-authentication, draft-ietf-hokey-key-mgm-06. Internet Draft, Work in Progress, April 2009.

HOK09 HOKEY, Handover Keying (hokey) WG, 2009. `http://www.ietf.org/html.charters/hokey-charter.html`. [15.7.2009]

HY03 Hui, S. and Yeung, K., Challenges in the migration to 4G mobile systems. *IEEE Communication Magazine*, 41,12(2003), pages 54–59.

I$^+$08 Izquierdo, A. et al., Using the EAP Framework for Fast Media Independent Handover Authentication. *In Proceedings of the 4th Annual International Conference on Wireless Internet. WICON 08*, November 2008, pages 1–8.

IEE06 IEEE 802.16 Task Group, IEEE 802.16 Task Group m (TGm), 2006. `http://ieee802.org/16/tgm/`. [22.10.2009]

IEE07 IEEE 802.11 Working Group, IEEE 802.11 - 2007, 2007. http://standards.ieee.org/getieee802/download/802.11-2007.pdf. [02.11.2009]

IEE08a IEEE 802.11 Working Group, IEEE 802.11k - Amendment 1: Radio Resource Measurement of Wireless LANs, 2008. http://ieeexplore.ieee.org/servlet/opac?punumber=4544752. [15.01.2009]

IEE08b IEEE 802.11 Working Group, IEEE 802.11r-2008: Amendment 2: Fast Basic Service Set (BSS) Transition, 2008. http://ieeexplore.ieee.org/servlet/opac?punumber=4573290. [15.01.2009]

Ins81 Institute, I. S., Internet Protocol. RFC 791, Internet Society, September 1981.

INT07 INTERNATIONAL TELECOMMUNICATION UNION, Key Global Telecom Indicators for the World Telecommunication Service Sector, 2007. http://www.itu.int/ITU-D/ict/statistics/at_glance/KeyTelecom99.html. [22.10.2009]

INT08 INTERNATIONAL TELECOMMUNICATION UNION, Telecommunication Standardization Sector (ITU-T), 2008. http://www.itu.int/ITU-T/. [06.11.2009]

J$^+$05 Jung, H. et al., Fast Handover for Hierarchical MIPv6 (F-HMIPv6). Internet Draft, Work in Progress, April 2005.

JPA04 Johnson, D., Perkins, C. and Arkko, J., IP Mobility Support in IPv6. RFC 3775, Internet Society, June 2004.

K$^+$06 Kwon, H. et al., USIM based Authentication Test-bed For UMTS-WLAN Handover. IEEE INFOCOM 2006 Poster and Demo, Electronics and Telecommunications Research Institute, April 2006.

Kau05 Kaufman, C., Internet Key Exchange (IKEv2) Protocol. RFC 4306, Internet Society, December 2005.

Kem05 Kempf, J., Instructions for Seamoby and Experimental Mobility Protocol IANA Allocations. RFC 4065, Internet Society, July 2005.

Ken05a Kent, S., IP Authentication Header. RFC 4302, Internet Society, December 2005.

Ken05b Kent, S., IP Encapsulating Security Payload (ESP). RFC 4303, Internet Society, December 2005.

Koo08 Koodli, R., Mobile IPv6 Fast Handovers. RFC 5268, Internet Society, June 2008.

Kor08 Korhonen, J., *IP Mobility in Wireless Operator Networks*. Ph.D. thesis, University of Helsinki, 2008.

KS05 Kent, S. and Seo, K., Security Architecture for the Internet Protocol. RFC 4301, Internet Society, December 2005.

KT06 Kivinen, T. and Tschofenig, H., Design of the IKEv2 Mobility and Multihoming (MOBIKE) Protocol. RFC 4621, Internet Society, August 2006.

L^+00 Laat, C. et al., Generic AAA Architecture. RFC 2903, Internet Society, August 2000.

L^+05a Liebsch, M. et al., Candidate Access Router Discovery (CARD). RFC 4066, Internet Society, July 2005.

L^+05b Loughney, J. et al., Context Transfer Protocol (CXTP). RFC 4067, Internet Society, July 2005.

LF08 Lei, J. and Fu, X., Evaluating the Benefits of Introducing PMIPv6 for Localized Mobility Management. *In Proceedings of the International Wireless Communications and Mobile Computing Conference. IWCMC 08*, August 2008, pages 74–80.

LSP08 Lampropoulos, G., Salkintzis, A. and Passas, N., Media-Independent Handover for Seamless Service Provision in Heterogeneous Networks. *Communications Magazine*, 46,1(2008), pages 64–71.

LW05 Liang, W. and Wang, W., A Quantitative Study of Authentication and QoS in Wireless IP Networks. *In Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2005*, March 2005, pages 1478–1489.

LYC08 Leung, K., Yegani, P. and Chowdhury, K., WiMAX Forum/3GPP2 Proxy Mobile IPv4, draft-leung-mip4-proxy-mode-10.txt. Internet Draft, Work in Progress, November 2008.

M$^+$00 Mink, S. et al., Towards Secure Mobility Support for IP Networks. *In Proceedings of the International Conference on Communication Technology (WCC - ICCT 2000)*, August 2000, pages 555–562.

M$^+$01a Mitton, D. et al., Authentication, Authorization, and Accounting: Protocol Evaluation. RFC 3127, Internet Society, June 2001.

M$^+$01b Moore, B. et al., Policy Core Information Model – Version 1 Specification. RFC 3060, Internet Society, February 2001.

M$^+$08 Montavont, N. et al., Analysis of Multihoming in Mobile IPv6, draft-ietf-monami6-mipv6-analysis-05. Internet Draft, Work in Progress, September 2008.

Man02 Mandin, J., 802.16e Privacy and Key Management (PKM) version 2, 2002. `http://wirelessman.org/tge/contrib/C80216e-04_131r1.pdf`. [10.7.2009]

MIH09 MIH, IEEE Standard for Local and metropolitan area networks- Part 21: Media Independent Handover. IEEE Standard, IEEE 802.21, January 2009.

MK04 Manner, J. and Kojo, M., Mobility Related Terminology. RFC 3753, Internet Society, June 2004.

MN06 Moskowitz, R. and Nikander, P., Host Identity Protocol (HIP) Architecture. RFC 4423, Internet Society, May 2006.

Mon01 Montenegro, G., Reverse Tunneling for Mobile IP, revised. RFC 3024, Internet Society, January 2001.

Nak07 Nakhjiri, M., Use of EAP-AKA, IETF Hokey and AAA Mechanisms to Provide Access and Handover Security and 3G-802.16M Interworking. *In Proceedings of the IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007).*, September 2007, pages 1–5.

ND08 Narayanan, V. and Dondeti, L., EAP Extensions for EAP Re-authentication Protocol (ERP). RFC 5296, Internet Society, August 2008.

NIS09 NIST, National Institute of Standards and Technology, 2009. www.nist.gov/. [01.11.2009]

NS 08 NS community, ns-2 wiki, 2008. http://nsnam.isi.edu/nsnam/index.php/Main_Page. [22.10.2009]

ONY03 Ong, C., Nahrstedt, K. and Yuan, W., Quality of Protection for Mobile Multimedia Applications. *In Proceedings of the 2003 International Conference on Multimedia and Expo.*, July 2003, pages 137–140.

OSI94 OSI, Information technology - Open Systems Interconnection - Basic Reference Model: The basic model, ITU-T X.200 series of recommendations, 1994. http://www.itu.int/rec/T-REC-X.200/en. [15.1.2009]

OTc09 OTcl Project, OTcl Home, 2009. http://otcl-tclcl.sourceforge.net/otcl/. [19.10.2009]

OWZ09 Ohba, Y., Wu, Q. and Zorn, G., Extensible Authentication Protocol (EAP) Early Authentication Problem Statement, draft-ietf-hokey-preauth-ps-08. Internet Draft, Work in Progress, June 2009.

Per97 Perkins, C., *Mobile IP: Design Principles and Practices.* Addison-Wesley Longman Publishing Co., Inc., 1997.

Per00 Perkins, C., Mobile IP Joins Forces with AAA. *IEEE Personal Communications*, 7,4(2000), pages 59–61.

Per02 Perkins, C., IP Mobility Support for IPv4. RFC 3344, Internet Society, August 2002.

PK04 Prasithsangaree, P. and Krishnamurthy, P., A New Authentication Mechanism for Loosely Coupled 3G-WLAN Integrated Networks. *In Proceedings of the IEEE 59th Vehicular Technology Conference, VTC 2004*, September 2004, pages 2998–3003.

R$^+$97 Rigney, C. et al., Remote Authentication Dial In User Service (RADIUS). RFC 2138, Internet Society, April 1997.

R$^+$01 Rensing, C. et al., A Survey on AAA Mechanisms, Protocols, and Architectures and a Policy-based Approach beyond: Ax. Technical Report 111, Swiss Federal Institute of Technology Zurich, May 2001.

R+02 Rosenberg, J. et al., SIP: Session Initiation Protocol. RFC 3261, Internet Society, June 2002.

Rah93 Rahnema, M., Overview of the GSM system and protocol architecture. *Communications Magazine*, 31,4(1993), pages 92–100.

Rec04 Recommendation, W., Resource Description Framework (RDF): Concepts and Abstract Syntax, 2004. `http://www.w3.org/TR/rdf-concepts/`. [25.2.2009]

Res99 Rescorla, E., Diffie-Hellman Key Agreement Method. RFC 2631, Internet Society, June 1999.

Rig97 Rigney, C., RADIUS Accounting. RFC 2139, Internet Society, April 1997.

S+05 Soliman, H. et al., Hierarchical Mobile IPv6 Mobility Management (HMIPv6). RFC 4140, Internet Society, August 2005.

S+08 Salowey, J. et al., Specification for the Derivation of Root Keys from an Extended Master Session Key (EMSK). RFC 5295, Internet Society, August 2008.

SAH08 Simon, D., Aboba, B. and Hurst, R., The EAP-TLS Authentication Protocol. RFC 5216, Internet Society, March 2008.

Sch03 Schiller, J., *Mobile Communications.* Addison-Wesley, second edition, 2003.

Sim94 Simpson, W., The Point-to-Point Protocol (PPP). RFC 1661, Internet Society, July 1994.

SK98 Stemm, M. and Katz, R., Vertical handoffs in wireless overlay networks. *Mobile Networks and Applications*, 3,4(1998), pages 335–350.

Sta05 Stallings, W., *Wireless Communications and Networks.* Prentice Hall, second edition, 2005.

Sta06 Stallings, W., *Cryptography and Network Security: Principles and Practices.* Prentice Hall, fourth edition, 2006.

Ste07 Stewart, R., Stream Control Transmission Protocol. RFC 4960, Internet Society, September 2007.

T$^+$09 Taniuchi, K. et al., IEEE 802.21: Media independent handover: Features, applicability, and realization. *IEEE Communication Magazine*, 47,1(2009), pages 112–120.

UMT08 UMTS, UMTS Forum, 2008. http://www.umts-forum.org/. [30.10.2009]

V$^+$00 Vollbrecht, J. et al., AAA Authorization Framework. RFC 2904, Internet Society, August 2000.

Val99 Valkó, A., Cellular IP: A New Approach to Internet Host Mobility. *SIGCOMM Computer Communication Review*, 29,1(1999), pages 55–65.

VoI09 VoIP, I. E. C., Voice over Internet Protocol: Definition and Overview, 2009. http://www.iec.org/online/tutorials/int_tele/index.asp. [30.6.2009]

W$^+$08 Wang, L. et al., Research on the Hierarchy AAA Scheme for Interworking Authentication in Heterogeneous Networks. *In Proceedings of International Conference on MultiMedia and Information Technology. MMIT 08*, December 2008, pages 586–589.

Wi-08 Wi-Fi, Wi-Fi Alliance Home Page, 2008. http://www.wi-fi.org/. [30.10.2009]

WiM06 WiMAX, WiMAX end-to-end network systems architecture (stage 2: Architecture tenets, reference model and reference points). Work-in-progress draft, Mobile WiMAX Forum, August 2006.

WiM08 WiMAX, WiMAX Forum, 2008. http://www.wimaxforum.org/home/. [30.10.2009]

WLA08 WLAN, IEEE 802.11 Wireless Local Area Networks, 2008. http://www.ieee802.org/11/. [30.10.2009]

WPA08 WPAN, IEEE 802.15 Working Group for WPAN, 2008. http://www.ieee802.org/15/. [30.10.2009]

XMH06 Xu, S., Matthews, M. and Huang, C., Security Issues in Privacy and Key Management Protocols of IEEE 802.16. *In Proceedings of the the 44th annual Southeast regional conference*, March 2006, pages 113–118.