Handover Mechanisms in Next Generation Heterogeneous Wireless Networks

A Thesis
submitted to
University of Technology, Sydney
by

Mo Li

In accordance with
the requirements for the Degree of

Doctor of Philosophy

Faculty of Engineering and Information Technology
University of Technology, Sydney
New South Wales, Australia
November 2008

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged with the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

_____

ACKNOWLEDGMENT

First of all, I would like to express my sincere gratitude to Dr. Kumbesan Sandrasegaran, who has directed and mentored my research work greatly. Dr. Sandrasegaran helped me in many other ways that supported, encouraged and motivated me during the pursuit of my Ph.D. degree. His endless efforts put into the weekly research meetings have made sure that both the flexibility of choosing research topics and the correct direction could be provided.

I would like to give my special thanks to my former thesis co-supervisor Dr. Tracy Tung. Dr. Tung guided and encouraged me on developing research ideas with her intensive knowledge in the mobility management field, although she left for U.S. at the midway of my Ph.D. candidature in 2007.

I would like to thank the University of Technology, Sydney and the Faculty of Engineering for their support through the APA scholarship and sponsoring my several conference trips.

I would like to extend my appreciation to Rachod Patachaianand for his comments to this thesis, and my fellow officemates: Xiaoan Huang, John Yong, Leijia Wu and Prashant Gami, who have made office a fun place.

Finally, I wish to thank my wife Liumei Lin, whose sacrifice to the family and support during the most difficult time in Australia gave me great encouragement. I am deeply indebted to my parents for their efforts to provide me with the best possible education.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

New access technologies such as IEEE 802.11 Wireless LAN are emerging as a new means of public wireless access. Working on public unlicensed bands, they are capable of providing high speed data services, but small radio coverage. The third generation cellular networks such as Universal Mobile Telecommunications System (UMTS) provide wide radio coverage, but have limited data rates. An integration of these heterogeneous wireless networks is expected to be an effective means of providing high speed data access in wide radio coverage in the Next Generation (NG) wireless networks. When a mobile user moves across these networks, it has to perform handover to maintain its services. During a handover, it is pivotal to guarantee both service continuity and service quality, which ensure that handover can be made seamlessly. To provide ubiquitous services, an extensive collaboration between network operators is anticipated to be an economic solution. Providing seamless handover and ubiquitous services in heterogeneous wireless networks presents many new research challenges.

The objective of this thesis is to develop new handover management techniques for supporting seamless handover and facilitating ubiquitous services in heterogeneous wireless networks. More specifically, new techniques for dealing with the extensive collaboration of NG network operators, and new techniques that enable the interworking of heterogeneous wireless technologies.

Regarding the extensive collaboration of network operators, a neighbour network trust information retrieval scheme is proposed for global roaming. With this scheme, an access network can obtain network trust information of its nearby access networks without relying on direct links with them. The retrieved trust information can be provided to an attached mobile user later to assist it with global roaming. Next, a handover decision algorithm that uses network trust information is presented. The proposed algorithm guarantees much more reliable handover in a multiple-operator

environment. It is demonstrated how quality of service is maintained and overall network load is balanced using the proposed handover algorithm. The thesis moves further to a proxy based authentication localisation scheme that focuses on the handover across two networks without a trust relation. The proposed scheme provides a secure and effective method of localising authentication at a third-party entity during a handover. This avoids resorting to a mobile's home network for identity verification in a handover, and thus, greatly reduces handover latency.

In terms of the interworking of heterogeneous wireless technologies, the thesis presents a multi-interface mobile terminal model for media independent handover. The presented model addresses the challenge of working with heterogeneous wireless technologies from the perspective of a mobile terminal. Under the proposed multi-interface architecture, a mobile terminal can work with multiple network interfaces, and still uses common upper layer protocols such as Mobile IPv4. Being compatible with IEEE 802.21 framework, it uses a cross-layer design approach.

Chapter 1

INTRODUCTION

New access technologies such as IEEE 802.11 Wireless Local Area Network (WLAN) are emerging as an effective means of public wireless access. The IEEE 802.11 standard, also known by its commercial trademark Wi-Fi, can provide high speed data services of up to 54Mbps with a radio range of less than 1km. Small radio coverage of such access technologies is due to several reasons, e.g. limitations of radio transmit power on using public unlicensed bands. In contrast, cellular networks can cover much wider areas. Current public wireless infrastructure is mainly built based on cellular network technologies such as the Third Generation (3G) Universal Mobile Telecommunications System (UMTS). Relying on Wideband Code Division Multiple Access (W-CDMA) techniques for radio access, UMTS supports wide radio coverage but at a relative low data rate of up to 384Kbps. Although more advanced standards such as High-Speed Downlink Packet Access (HSDPA) may achieve a data rate of up to 14.4Mbps in the downlink connection (as specified in the 3rd Generation Partnership Project (3GPP) release 5), it may be insufficient for many multimedia applications.

The growing demand for high speed data access at anytime, anywhere and on any device necessitates a new direction in the design of the Next Generation Wireless Networks. From a mobile user's viewpoint, some key features of the Next Generation Wireless Networks include high bandwidth, low latency, and ubiquitous coverage. However, none of the current wireless technologies can simultaneously satisfy these needs at low cost. Intuitively, the "high bandwidth" and "ubiquitous coverage" needs of a mobile user are best satisfied if it can freely hand over to any discovered networks to maintain its services at all times. The interworking of the current wireless technologies including 3G UMTS and WLAN, and other future technologies thus becomes an

economic and practical solution. Seamless integrating heterogeneous wireless networks would enable a ubiquitous high speed access infrastructure for mobile users.

For a mobile user in heterogeneous wireless networks, ubiquitous access can be further identified as having capabilities in two aspects: *service continuity* and s*ervice quality*. For service continuity, a mobile user is often engaged in a so-called handover operation to ensure that its user session is maintained continuously without being aware of the underlaying operation. In a Global System for Mobile communications (GSM) network, for example, a mobile user is connected to a Base Station (BS) via a radio link. If the mobile user moves to the coverage area of another BS, the radio link to the old BS would be eventually disconnected, and a radio link to a new BS should be established to continue the telephony conversation. Handover is referred to as the process of switching user connections in order to keep ongoing connections uninterrupted. On an integrated network infrastructure, a mobile user switches between heterogeneous wireless networks to obtain the best available connection to the network. A typical case of a handover for service continuity in heterogeneous wireless networks is the handover of a mobile user's radio link from a high speed data link e.g. IEEE 802.11 to a low rate data link e.g. UMTS when it moves out of a Wi-Fi hotspot.

Apart from handover for service continuity, handover can be initiated for optimising service quality. Unlike a mobile user of a GSM network who makes solely voice calls, a mobile user of UMTS or WLAN may have multiple data services such as voice, video, and messaging being carried on top of the Internet Protocol (IP) at the same time. In this case, different priorities need to be applied to these services so as to guarantee a certain level of performance to a specific data flow, especially when the network capacity is limited.

Therefore, service continuity and service quality would be essential to mobile users in the Next Generation Heterogeneous Wireless Networks. When multiple wireless networks provide overlapped radio coverage to a mobile user, handover can be an effective method of optimising service quality. This is achieved by avoiding those wireless networks with less satisfactory conditions. For example, a mobile user being served by a UMTS network may choose to switch to an available Wi-Fi hotspot to obtain higher bandwidth for its bandwidth-hunger applications once the Wi-Fi service

becomes available. Handover between networks of the same type occurs for balancing network load or other reasons. Maintaining a mobile user's service continuity and service quality in a handover makes sure that the handover is performed seamlessly. *Seamless handover* makes the transfer of a mobile user's network connections transparent (without perceptible interruption of services) to upper layer applications. With seamless handover, a mobile user can obtain service portability and application persistence across heterogeneous wireless networks. Generally, seamless handover will be a key enabling technique for the Next Generation Heterogeneous Wireless Networks.

1.1 Motivation

In the Next Generation Wireless Networks, various types of wireless networks including UMTS and WLAN are expected to be interconnected to support ubiquitous high speed data services. These wireless systems were designed independently and targeting different service types, data rates, and users, and thus require an intelligent interworking approach to be effective. In heterogeneous wireless networks, both the mobile user and the interconnected wireless networks play an important role in determining how service continuity and service quality can be served in a handover. Providing service continuity and service quality in a heterogeneous network environment would create several research challenges in the Next Generation Heterogeneous Wireless Networks and these challenges are itemised below:

- Access technologies: Heterogeneous wireless networks that employ a number of radio technologies may have overlapped radio coverage. A mobile user needs to switch between access networks to maintain service continuity and optimise service quality. How does a mobile user deal with heterogeneous access technologies?

- Network architectures: Heterogeneous wireless networks rely on different network architectures and protocols for transport, routing, mobility management and so forth. How will they be interconnected in an integral manner to facilitate the cooperation between themselves?

- Network conditions: Network conditions such as bandwidth, delay, jitter and so forth may vary across wireless networks, and result in different service quality to

be provided. How does a mobile user deal with the variation in network conditions, and maintain service quality when crossing heterogeneous wireless networks?

The above intrinsic *heterogeneities* call for the design of a well-organised common infrastructure to integrate heterogeneous wireless networks.

On the other hand, the increased affordability of the WLAN technology is encouraging both wireless Internet Service Providers such as T-Mobile and retailers such as Starbucks to deploy Wi-Fi services in airports, train stations, hotels and so forth. According to Analysys Consulting Ltd. (Cambridge), the current cost of transferring 1 Mb over an IEEE 802.11 network is between 0.2 and 0.4 eurocent, compared with between 3 and 38 eurocents for General Packet Radio Service (GPRS) networks. By the year 2007, 66,921 Wi-Fi hot spots have been deployed in the U.S., which was up 56% from the previous year. This is underpinned by the recent advances in technologies, e.g. Wi-Fi mesh architecture, which enables the delivery of Wi-Fi services in citywide areas. All these factors have contributed to the independent roll-out of global Wi-Fi services by a large number of small WLAN operators. The Return On Investment (ROI) of reusing the high speed Wi-Fi hotspots that have already been deployed by these operators would outweigh building a new high speed network from scratch in the Next Generation Wireless Networks. However, several research challenges would arise from dealing with a large number of independent network operators and these challenges are itemised below.

- Interoperability: Services will be jointly provided by autonomous networks of multiple network operators. This is referred to as the *multiplicities* of network operators in this thesis. How will multiple network operators collaborate with each other in an effective manner to make best use of their network infrastructures?

- Large number of operators: A large number of network operators are expected to co-exist and collaborate in the Next Generation Wireless Networks. In such circumstances, mobile users who are responsible for handover decision will require increased levels of control over how services can be secured in handover. This will be complicated by versatile trust relationships between network operators.

The above *multiplicities* call for the design of a well-engineered network trust framework to accommodate multiple network operators securely and efficiently.

In summary, the research challenges arising from the Next Generation Heterogeneous Wireless Networks can be classified into two areas: 1) the *heterogeneity* issues from the interworking of heterogeneous wireless technologies; 2) the *multiplicity* issues from the co-existence and extensive collaboration of a large number of network operators. Seamless handover across heterogeneous wireless networks is achieved on the promise of the aforementioned network heterogeneities being dealt with properly. This requires the appropriate solutions at various system levels: mobile terminals, network architectures, protocols and so forth. Apart from the challenges at the system level, seamless handover across multiple network operators will bring about the challenges at the operation level. This requires the corresponding solutions for the both parties: mobile user and network operator to address the discussed multiplicities.

1.2 Thesis Overview

In this thesis, four research contributions have been made for seamless handover in the Next Generation Heterogeneous Wireless Networks. Three of the contributions address the multiplicity issues, while the fourth contribution focuses on the heterogeneities caused by a mobile user's interfacing with heterogeneous access technologies. The research work of this thesis is summarised in Figure 1.1.

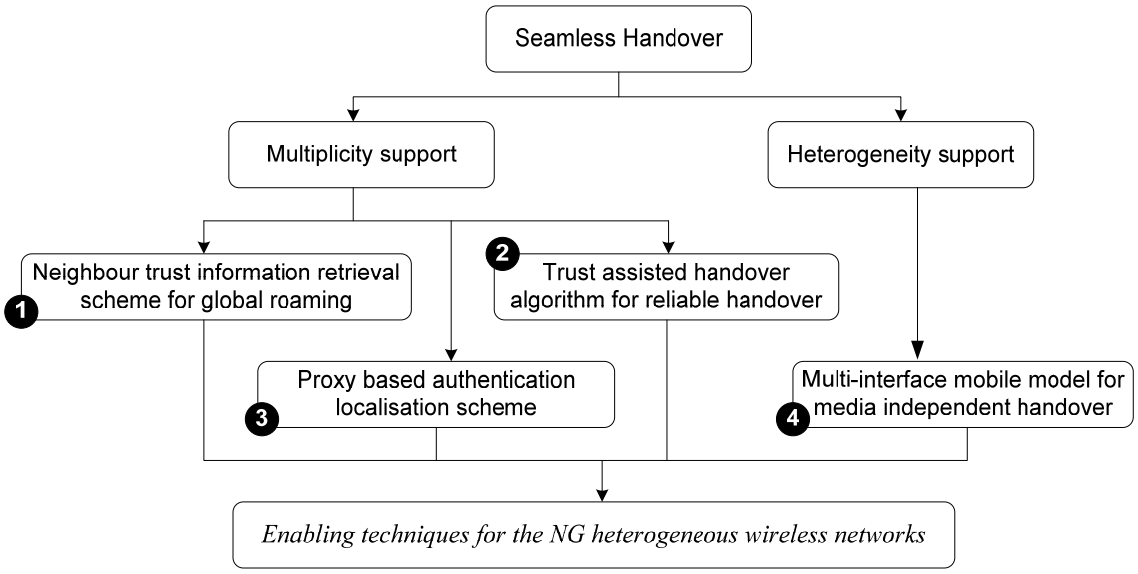


Figure 1.1 Summary of thesis contributions on handover management in the Next Generation Heterogeneous Wireless Networks

Chapter 2 provides an overview of handover management in the Next Generation (NG) Heterogeneous Wireless Networks. It identifies the major research issues related to handover management and classifies the issues into two categories: *Mobility Engineering* and *Handover Methodology*. The fundamental issues in each category are explained. And their corresponding requirements in the NG heterogeneous wireless networks are highlighted and the current solutions are briefed.

Chapter 3 introduces handover related security mechanisms in various wireless networks. It presents the concept of Authentication, Authorisation, and Accounting (AAA) and explains its applications in cellular networks and IEEE 802.11 WLAN. A generic AAA architecture that is designed for the interworking of heterogeneous wireless networks is introduced. The required AAA protocols and their implementations are discussed. This chapter is intended to provide background on the handover solutions for the multiplicity challenges.

In Chapter 4, a neighbour network trust information retrieval scheme is proposed for global roaming in the NG heterogeneous wireless networks. The proposed scheme provides an effective means of retrieving and distributing network trust information between two neighbouring networks without a trust relation. A network builds up the

network trust pattern of its nearby networks using mobile users' handover records without relying on direct links with its neighbours. In this scheme, a mobile user can obtain necessary network trust information dynamically from its serving access network, and thus has sufficient flexibility to work with a large number of network operators.

In Chapter 5, a network trust information assisted handover algorithm is presented for handover in a multi-operator environment. On a cost basis, the proposed algorithm uses network trust relationship in additional to other handover metrics such as signal strength, delay, network load and so forth. This provides a mobile user with an intelligent approach of dealing with complex trust relationships between network operators. According to the proposed algorithm, network trust information that is needed in a handover can be considered as an integral part of handover decision algorithm rather than as a separated add-on. This effectively avoids unnecessary handover attempts, and ensures reliable handover in a multi-operator environment. The algorithm can work without compromising Quality of Service (QoS).

Chapter 6 introduces a proxy-based authentication localisation scheme for handover between networks without a trust relation. The proposed scheme presents a method of relaying a mobile's home network's authentication authority to a third party proxy based on a proposed trust association model. To provide secure and controllable authentication localisation, a set of fast authentication related algorithms are presented. The proposed trust association model along with the fast authentication algorithms enable the authentication required in a handover to be locally processed. This avoids resorting to the home AAA server for identity verification in a handover. The proposed scheme is particularly effective for the handover taking place between two networks without a trust relation, and can result in great performance improvement.

In Chapter 7, a generic multi-interface mobile terminal model is described. The proposed model provides a multi-interface architecture that supports multiple heterogeneous network interfaces being alternatively used and media independent handover. Its design focus is to enable the multi-interface function and support ordinary upper layer protocols such as TCP/IP and Mobile IPv4 at the mobile end without modifying wireless network infrastructure already in use. This makes sure that the proposed model that is compatible with the latest IEEE 802.21 standard can be

practically implemented for media independent handover. In this model, the cross-layer design is introduced for improving handover performance.

Chapter 8 summarises the contributions of this thesis in the four areas. It identifies the areas for future research in handover management for the NG heterogeneous wireless networks.

1.3 Related Publications

The following publications have been produced based on the contributions included in this thesis.

JOURNALS

I. Mo Li and K. Sandrasegaran, "A Proxy Based Authentication Localisation Scheme for Handover between Non Trust-Associated Domains," submitted to *ACM Mobile Computing and Communications Review (MC2R)*, 2009.

II. Mo Li and K. Sandrasegaran, "A Dynamic Trust Information Retrieval Scheme for Global Seamless Roaming," submitted to *Wiley Security and Communication Networks, Special Issue on Security in Mobile Wireless Networks*, 2009.

III. Mo Li, K. Sandrasegaran, and T. Tung, "Trust-Assisted Handover Approach in Hybrid Wireless Networks," *Springer Wireless Personal Communications, Special Issue on Resource and Mobility Management and Cross-Layer Design for the Support of Multimedia Services in Heterogeneous Emerging Wireless Networks*, to be published.

BOOK CHAPTERS

IV. K. Sandrasegaran and Mo Li, "Identity Management," in Handbook of Research on Wireless Security, Y. Zhang, J. Zheng, and M. Ma, Eds. New York: Information Science Reference, February 2008.

CONFERENCES

V. Mo Li and K. Sandrasegaran, "A Proxy Based Authentication Localisation Scheme for Handover between Non Trust-Associated Domains," submitted to the IEEE International Conference on Communications (ICC Communication and Information Systems Security Symposium), Dresden, Germany, June 2009.

VI. Mo Li, K. Sandrasegaran, and T. Tung, "Performance Evaluation of A Multi-Interface Model for Media Independent Handover," appeared in the proceeding of The

7th International Symposium on Communications and Information Technologies (ISCIT), Sydney, Australia, October 2007.

VII. Mo Li, K. Sandrasegaran, and T. Tung, " A Multi-Interface Proposal for IEEE 802.21 Media Independent Handover," appeared in the proceeding of the Sixth IEEE International Conference on Mobile Business (ICMB), Toronto, Canada, July, 2007.

VIII. Mo Li, K. Sandrasegaran, and T. Tung, "Trust-Assisted Handover Decision Algorithm in Hybrid Wireless Networks," appeared in the proceeding of IEEE Wireless Communications & Networking Conference (WCNC), Hong Kong, March, 2007.

IX. Mo Li, K. Sandrasegaran, and T. Tung, "An Analysis of Prioritized Hybrid Interworking Requirements in Next-Generation Wireless Data Networks," appeared in the proceeding of IEEE International Conference on Computer & Communication Engineering (ICCCE), Kuala Lumpur, Malaysia, May 2006.

X. Mo Li and K. Sandrasegaran, "Federated Authentication in Next-Generation Wireless Networks," appeared in the proceeding of 8th International Symposium on DSP and Communication Systems (DSPCS & WITSP), Noosa Heads, Australia, December 2005.

XI. Mo Li and K. Sandrasegaran, "Network Management Challenges for Next Generation Networks," appeared in the proceeding of the 30th IEEE Conference on Local Computer Networks (LCN), Sydney, Australia, November 2005.

XII. Mo Li, K. Sandrasegaran, and X. Huang, "Identity Management in Vertical Handovers for UMTS-WLAN Networks," appeared in the proceeding of the Fourth IEEE International Conference on Mobile Business (ICMB), Sydney, Australia, July 2005.

TECHNICAL REPORTS

XIII. K. Sandrasegaran, X. Huang, and Mo Li, "Digital Identity in Next Generation Networks," University of Technology, Sydney (UTS), Sydney, Alcatel Research Partner Program (ARPP) August 2005.

XIV. K. Sandrasegaran, Mo Li, and X. Huang, "Identity Management in Next Generation Networks," University of Technology, Sydney (UTS), Sydney, Alcatel Research Partner Program (ARPP) May 2005.

Chapter 2

HANDOVER MANAGEMENT

This chapter provides an overview of handover management and discusses various issues related to handover management. These related issues are classified into two categories: mobility engineering and handover methodology. First, mobility engineering is described, and its current solutions are presented. Thereafter, handover methodology is explained.

2.1 Introduction to Handover

Handover is a process of transferring an active mobile user session from one Base Station (BS) or Access Point (AP) to another in order to keep the user's connection uninterrupted. In the traditional circuit-switched wireless networks such as GSM, handover is employed mainly for maintaining a mobile user's telephony voice. The handover in such a circumstance is motivated by the fact that the coverage area of a single BS transceiver can not cover the whole service area. The coverage area of one or more BS transceivers at a single physical site is referred to as a cell. In Frequency Division Multiple Access (FDMA) based systems, a cluster is a group of cells in which frequencies are not reused. Clusters can be repeated with careful planning to minimise interference among cells using the same frequency so as to enlarge radio coverage as shown in Figure 2.1. Such a flat compound architecture can be supplemented by more intelligent radio resource management techniques such as the macrocell/microcell overlay [1], which consists of large-size macrocells and small-size microcells for balancing network capacity and network control load associated with handover.

When a mobile user connection to an AP or BS degrades below an acceptable threshold, it has to switch the session to a neighbouring cell. If the neighbouring cell employs the same type of access technology, handover across these cells is often seen being done

horizontally. *Horizontal handover* refers to handover between base stations using the same type of network interface. This is common in homogeneous circuit-switched cellular systems such as GSM and Code Division Multiple Access (CDMA) networks. Apart from being implemented in circuit-switched cellular systems, horizontal handover can be utilised for maintaining the continuity of wireless data services. Such applications are seen in packet-switched cellular systems such as General Packet Radio Service (GPRS) and Universal Mobile Telecommunications System (UMTS) networks.

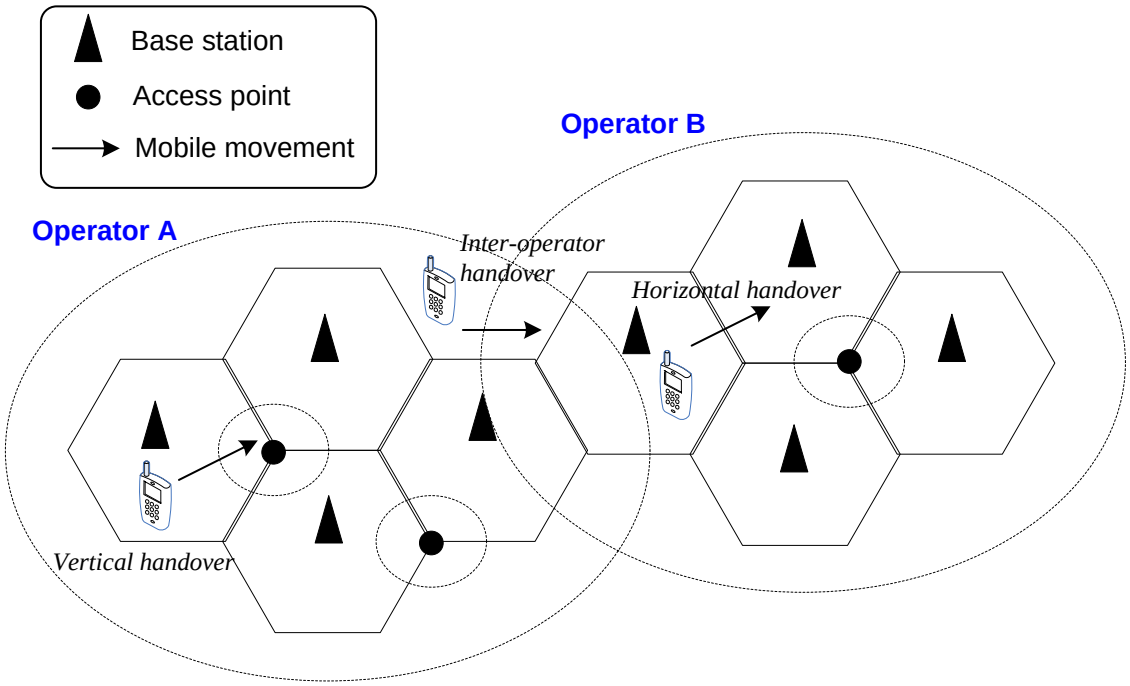


Figure 2.1 An overview of handover scenarios

The growing demands for high speed wireless data services drive the development of new access technologies. Wireless Local Area Networks (WLAN) such as IEEE 802.11 are able to provide high speed access but with small radio coverage. For example, IEEE standard 802.11g [2] supports a data rate up to 54Mbps but an outdoor coverage of 140 meters. In contrast, the Third Generation (3G) cellular systems, e.g. UMTS, can offer much wider coverage through more complex network architecture. However, the data rates they can offer are not unfavourable for many real-time applications, which need high bandwidth. Thus, an integration strategy of the two technologies (e.g. 3G UMTS and WLAN) has been proposed and widely accepted in the literature [3-8] as an economical and feasible solution for providing ubiquitous access [9]. This integration is

expected to result in a heavy demand for handover between heterogeneous wireless networks, which is recognised as an important feature of the NG wireless networks [6]. Handover between two networks based on different access technologies is known as *vertical handover*. Vertical handover is further divided into two categories: upward (move-out) vertical handover and downward (move-in) vertical handover [10]. An upward vertical handover occurs from a BS/AP with smaller radio coverage to a BS/AP with wider coverage. A downward vertical handover occurs in the reverse direction to an upward vertical handover. Apart from handover for radio link quality reasons, vertical handover can be initiated for optimising service quality for wireless data services. In this context, vertical handover has to deal with the heterogeneities existing in the interconnected wireless networks.

With the proliferation of wireless local area networks such as IEEE 802.11, it is envisaged that multiple wireless networks may be overlapped and may serve the same service area to form a wireless overlay network [10]. The wireless overlay networks could be heterogeneous in nature and belong to different network operators. The multiple ownership of the wireless overlay networks would make the task of interworking heterogeneous networks more complex and challenging. Network operators may apply diverse security policies in their administrative domains which may prevent others from using their networks. A mobile user has a reasonable expectation that security of each individual domain will not be an obstacle [11], when they are switched across heterogeneous network domains.

The perspective on “Cooperative Network” [12] that enables seamless communications on mobile devices operating in networks composed of heterogeneous technologies has encouraged an initiative to build a secure and trusted environment. A mobile user in a cooperative network is allowed to handover between networks of different technologies and allowed to select a network based on its preference of network operators. The term *inter-operator handover* denotes a handover between two networks belonging to different operators. This handover could be either a vertical inter-operator handover or a horizontal inter-operator handover (as shown in Figure 2.1). An inter-operator handover implies a transfer of trust association established for handover attachment to another operator.

The objective of supporting various forms of handover on a mobile user is to provide ubiquitous access across heterogeneous wireless networks and a number of network operators without manual user intervention [12]. This is supplemented by the demands for comprehensive and personalised services, stable system performance and service quality [13]. Seamless handover in the NG wireless networks needs additional capabilities in network architectures, protocols and control mechanisms, all of which combine to facilitate the smooth interworking of heterogeneous wireless systems. Accordingly, the interworking raises a number of research issues, which can generally classified into two categories: Mobility Engineering and Handover Methodology.

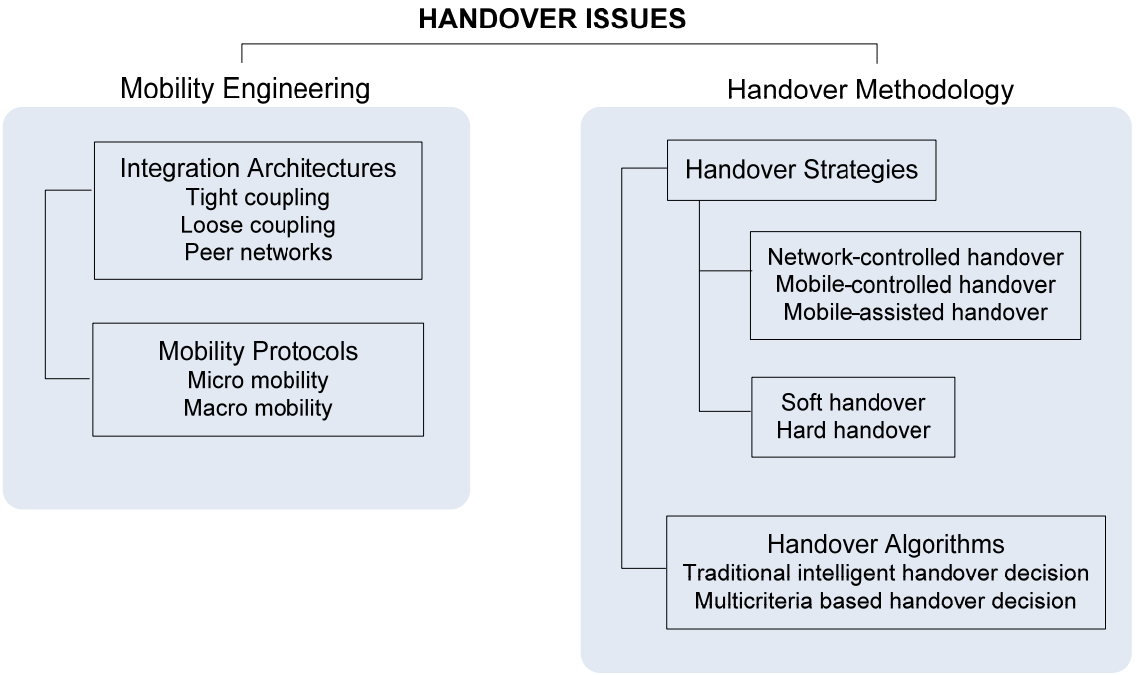


Figure 2.2 Handover management issues in the NG heterogeneous wireless networks

Figure 2.2 summarises the major issues in each category. Mobility engineering provides basic building blocks, which underpin handover functionality. Mobility engineering provides a common platform for all the mobile users, and comes with network infrastructure. Integration architectures and mobility protocols lie in this category. Handover methodology, on the other hand, specifies the way, in which handover should be performed. Unlike network protocols being “hardcoded” for all the mobile users, handover methodology can be made different for each individual mobile user for

optimised service quality. Handover methodology is comprised of handover strategies and handover algorithms.

2.2 Mobility Engineering

For the NG wireless networks, mobility engineering provides basic interworking architecture, on top of which seamless mobility can be enabled. On the mobile terminal level, it involves the consolidation of multiple functions into small, portable devices with integrated cellular and WLAN interfaces [14]. On the network level, it addresses the convergence of current heterogeneous wireless systems through appropriate interworking mechanisms. Generally, mobility engineering covers integration engineering and mobility protocols as shown in Figure 2.2. Due to the heterogeneities of the interconnected wireless systems in their access technologies, protocols, and architectures, mobility engineering has to cope with a series of challenges in [4] in the NG wireless networks.

2.2.1 Integration Architecture

Currently, there exist disparate wireless networks such as WLAN for local areas, UMTS for wide areas, and satellite networks for global communications. These networks are designed for specific service needs and vary widely in regard to their air interface technologies, network architectures, protocols and signalling process. The integration of these heterogeneous networks is recognised as an effective solution for service provisioning for the NG mobile users [8]. Consequently, the NG wireless network infrastructure is expected to converge into a heterogeneous, distributed, all-IP network architecture [6, 15, 16]. All-IP based infrastructure should seamlessly integrate the current and emerging wireless architectures such as UMTS, IEEE 802.11 and so forth.

A number of wireless systems may employ disparate mechanisms for the same function, e.g. radio access, authentication, and so forth. For example, IEEE 802.11g [2], a WLAN standard operating on the 2.4 GHz unlicensed band, provides theoretical data rate up to 54Mbps. Its security architecture recommends the authentication through IEEE 802.1X framework [17], which employs the Extensible Authentication Protocol (EAP) [18] such as EAP-Transport Layer Security (EAP-TLS). While, 3G wireless systems, e.g. UMTS uses Wideband Code Division Multiple Access (W-CDMA) for its underlying

air interface, and builds its Packet-Switched (PS) core reusing General Packet Radio Service (GPRS). The UMTS Authentication and Key Agreement (AKA) [19] is conducted in a challenge-response manner. When these different wireless networks with disparate mechanisms are interconnected, specific integration architecture is needed. The integration architecture is introduced to mitigate network heterogeneities arising from interconnection, and provide smooth services for mobile users. The integration of heterogeneous wireless networks for mobility management can be achieved in three types of architectures: tight coupling, loose coupling and peer networks as illustrated in Figure 2.3.

In the *tight coupling*, a WLAN is connected to a UMTS's core network via the *Gn* interface. The rationale behind the tightly-coupled approach is to make the WLAN appear to the UMTS core network as another Serving GPRS Support Node (SGSN) area. Both the signalling and data traffic of the WLAN are routed to the UMTS's core network. The WLAN may be deployed by the UMTS operator or by an independent operator. The WLAN is assigned a Routing Area Identity (RAI), and shares the same address pool as the UMTS Radio Network Controller (RNC) under the same Gateway GPRS Support Node (GGSN). All these functions can be achieved by a SGSN emulator as demonstrated in Figure 2.3. The SGSN emulator hides the details of the WLAN to the UMTS core, and implements all the protocols required for interworking. Therefore, the user's mobility across the WLAN-UMTS boundary is treated as an inter-SGSN update procedure by the UMTS's mobility management. Within the WLAN, the mobility on the same Extended Service Set (ESS) follows the WLAN's mobility management procedure. But, an intra-SGSN routing area update is employed for the mobility across ESSs. Reference [4] describes an interworking architecture that enables the integration of WLAN with 3G Public Land Mobile Network (PLMN) in a tightly coupled manner. In the tight coupling architecture, user data traffic is routed to a Packet Data Gateway (PDG), which locates at a mobile user's home PLMN. The mobile user requests access to a coupled WLAN to its 3G authentication server. The 3G Partnership Project (3GPP) specifies a reference model for the WLAN-3GPP interworking in its specification TS 23.234 [7]. Its specified reference model allows 3GPP PS services to be provided via a 3GPP visited network or the home network using the *Wn* reference point, which connects a WLAN to a WLAN Access Gateway (WAG) on 3GPP network.

In terms of access control signalling and account information, WLAN interfaces to 3GPP network via the *Wa* reference point that corresponds to a 3GPP Authentication, Authorisation, and Accounting (AAA) proxy.

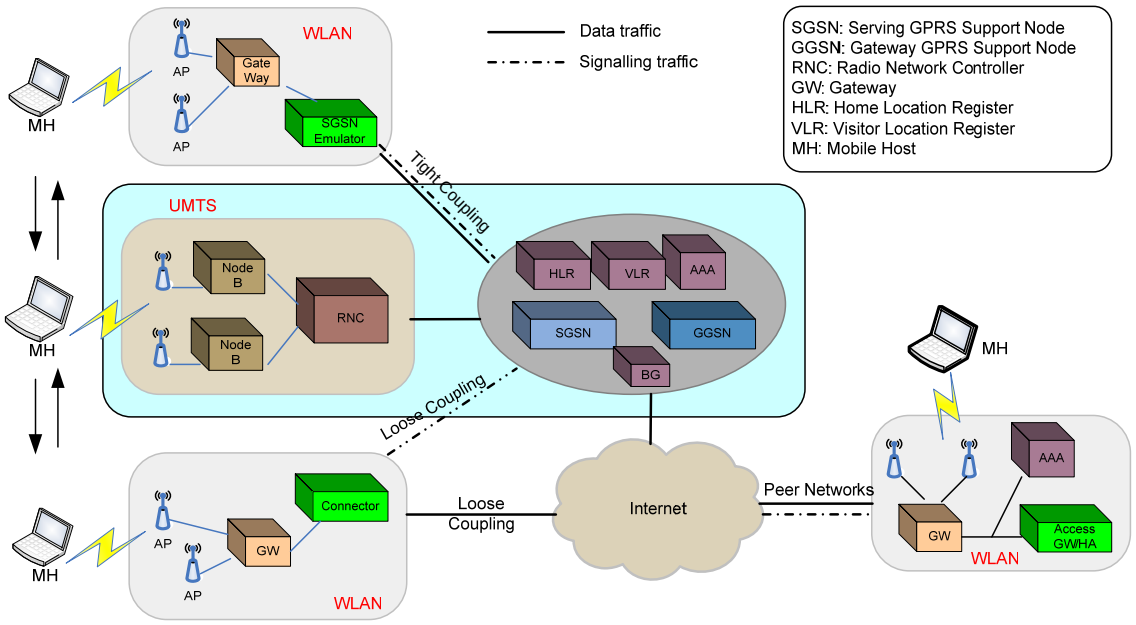


Figure 2.3 Three alternatives for the integration of UMTS and WLAN networks

The *loose coupling* makes use of IP connectivity provided by local WLAN operator. The attached WLAN appears as another router area to the UMTS. However, the WLAN interfaces directly to the UMTS's core network for control signalling. As a result, the complexity of tunnelling data traffic to the UMTS is avoided. The high speed data traffic of the WLAN is never injected into the UMTS's core network. In this approach, different protocols can be utilised to handle AAA and mobility management in the WLAN and UMTS portions of network. However, to make seamless handover possible, they have to interoperate. Figure 2.3 shows a big picture of how a WLAN can be integrated with a UMTS network via a connector. Unlike the SGSN emulation technique that is applied in the tight coupling, the connector in between WLAN and UMTS relays control signalling to UMTS's core network. Several WLAN ESSs could be connected to the connector for system interworking with UMTS. The mobility of a mobile user within the same network is handled by local mobility management. For handover across two heterogeneous networks, a UMTS handover procedure or a customised procedure should be executed. Reference [4] presents a 3G/WLAN interworking architecture for the 3G-based access control and charging scenario. The

proposed architecture incorporates a WLAN network by routing the AAA signalling to/from a 3G AAA proxy residing in the associated visited 3G PLMN. A WLAN AAA proxy with the WLAN acts as the “connector” between WLAN and 3G PLMN. 3GPP endorses the loose coupling idea in its specification TS 23.234 for interworking [7]. *Chen et al.* presented a practical loose coupling architecture for the integration of GPRS and WLAN systems, which has been put into commercial operation [20]. By placing a logical gateway on the conjunctional point of GPRS and WLAN, the proposed architecture can keep mobility management functions in GPRS and WLAN as they are. The research on the integration architecture [4, 5, 20] argues that the loose coupling will be deployed earlier than the tight coupling due to the architecture complexity in the tight coupling.

The *peer network* architecture allows control signalling for interworking to pass through the public Internet instead of using a dedicated link via a PDG at UMTS network. The WLAN can be operated by a different operator and function independently by incorporating its own AAA services. The WLAN ESSs terminate on a gateway at the WLAN as in the aforementioned architectures. The gateway handles a mobile user’s roaming to other peer networks. For example, a UMTS subscriber switches to a peered WLAN network as illustrated in Figure 2.3. The mobility of a mobile user within individual network infrastructure is handled by its local mobility protocols. In the peer network architecture, mobility protocols like Mobile IP [21] plays a key role in supporting a mobile user’s mobility across heterogeneous wireless networks. This is because an access-technology-independent mobility protocol would decouple the integrated heterogeneous networks in their functionalities, and thus allow each network to function on their own. An AAA server residing in the WLAN manages its own subscribers and communicates with other AAA servers in peer networks for roaming related operations. Rather than having a WLAN appear as an attached routing area to the UMTS, the peer network architecture maximises the autonomy of each type of network. *Shi et al.* described an IEEE 802.11 and cellular network integrated architecture from the perspective of roaming and authentication in [22]. The AAA structure introduced in [22] was proposed to make IEEE 802.11 architecture and signalling processes work with cellular networks. Reference [8] described an IP-based interconnection solution that works on a global IP infrastructure like the Internet.

According to [8], heterogeneous wireless networks can be integrated using a third party, Network Interoperating Agent, for establishing Service Level Agreements (SLA) among network operators.

2.2.2 Mobility Protocols

IP has been recognised as the foundation for next generation integrated wireless networks [15, 23], since it takes advantage of the widely installed base of IP devices. With various integration architectures discussed in Sec. 2.2.1, two heterogeneous wireless networks can be interconnected and exchange controlling signalling and user traffic with each other. One of the challenges is the design of a mobility protocol suite that could achieve the desired mobility across various types of access networks. The protocol should provide optimal performance across varied network environments. This section discusses three levels of mobility support in the context of all-IP-based [15] wireless networks: macro mobility, micro mobility, and access mobility, which are explained below.

- Macro mobility: The movement of a mobile user between two network domains is referred to as macro mobility. For example, the movement of the mobile user MH from Domain I to Domain II shown in Figure 2.4. A domain is an autonomous wireless network, which may include a number of subnets in various geographical regions.

- Micro mobility: The movement of a mobile user between two subnets within a domain is referred to as micro-mobility. For example, the movement of the mobile user MH from Subnet A to Subnet B shown in Figure 2.4. A subnet is an identifiably separated part of a network, which may consist of a group of network devices at the same geographic location.

- Access mobility: The movement of a mobile user between the access points or the base stations within the scope of the same access router or single radio network controller RNC. For example, the movement of the mobile user MH from AP 1 to AP 2 shown in Figure 2.4. The two access points AP 1 and AP 2 are under the management of the same subnet “Subnet A”.

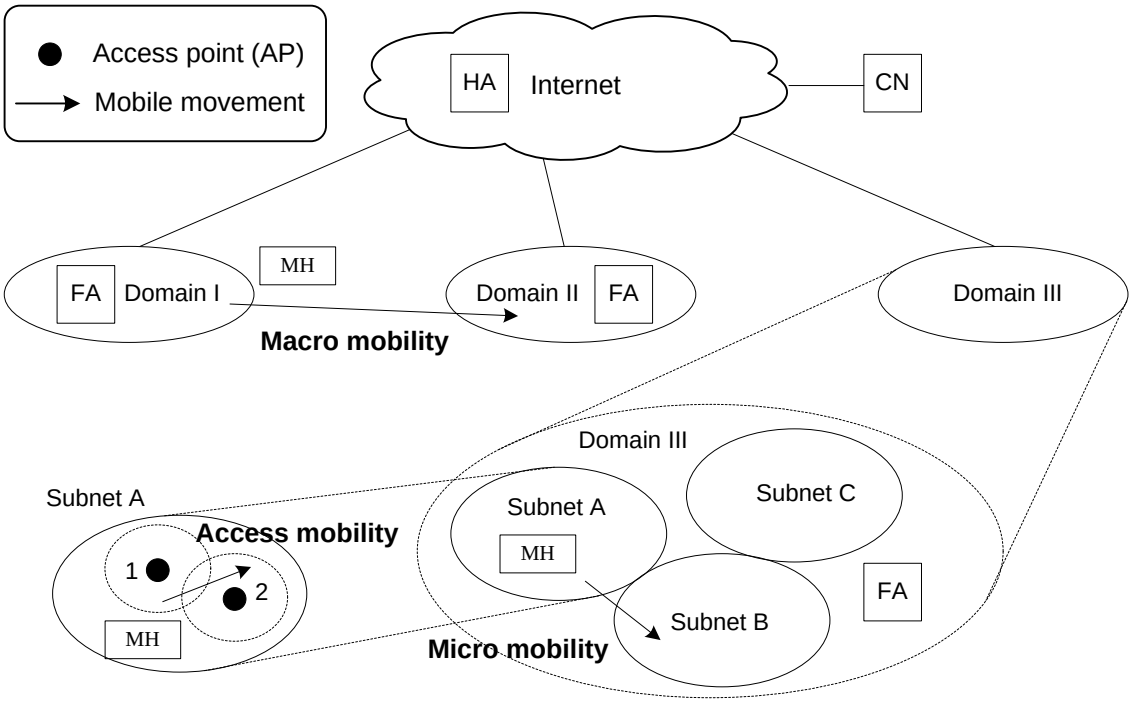


Figure 2.4 Mobility support in the all IP-based wireless networks

On an IP network, a host is identified by an IP address that uniquely identifies its point of attachment to the network. This address is usually associated with a physical subnet. Therefore, the host must reside on its subnet in order to keep a connection to the Internet. The need of a topologically correct address [24] prohibits the host from moving and retaining its connection to the Internet using the base IP protocol. Mobile IP was proposed with the objective of supporting mobile hosts with continuous IP connectivity on a macro mobility level.

Macro Mobility Support

In Mobile IP [21], macro mobility support is provided by redirecting packets for a mobile host to its current location using IP tunnelling. Mobile IP defines two new functional entities: Home Agent (HA) and Foreign Agent (FA), both of which act as tunnel endpoints. Mobile Host (MH) and Correspondent Node (CN) are end systems that are involved in a connection. A MH attaches to a FA for local access when it roams outside of its home network. The MH retains a permanent IP address assigned by its home network, which is also known as its home IP address. The home IP address is complemented by a dynamically obtained Care of Address (CoA) that uniquely

identifies the current location of the MH. The HA maintains a record of the MH's CoA, and forwards the data packets sent to the MH to the corresponding FA through an IP tunnel. Mobile IP consists of the following procedures for mobility management:

- Agent discovery: A MH detects whether it has moved from a subnet to another by periodically checking the received agent advertisement messages broadcasted by the FA. Agent advertisement messages are used by FA and HA to advertise their presence. Alternatively, the MH can send agent solicitations to request agent advertisements in an attempt to discover a new agent.

- Registration: Having obtained a CoA from a discovered FA, the MH has to register with the HA. The registration provides the HA with the current location information of the MH. After receiving a registration request, the HA sets up a mobility binding containing the MH's home IP address and its current CoA.

- Tunnelling: All the packets sent to the MH are intercepted by the HA. The HA encapsulates and forwards the packets through a tunnel established between FA and the MH's CoA. The FA decapsulates and sends the received packets to their final destination, the MH.

With the above functions, Mobile IP offers handover control to a mobile user. For example, when the MH moves from Domain I to Domain II as shown in Figure 2.4, it first obtains a new CoA from the FA at Domain II. Then, the MH registers this new CoA with its HA. The HA establishes a new tunnel terminating at the FA of Domain II, and removes the tunnel pointed to the old FA at Domain I. Once the new tunnel is set up, the HA tunnels packets destined to the MH to the current location of the MH, the FA at Domain II. All the processes are kept transparent to the CN, which continuously sends the packets to the MH's home IP address.

Mobile IP works fairly well when the visited foreign network is near a mobile user's home network. However, when the distance between the visited foreign network and the home network is sufficiently large, a mobile host's handover operation would induce large signalling delay for registration and binding update operations. This high latency in handover may be noticeable to applications, and hence no longer transparent to end

users. Moreover, assuming a large number of mobile hosts changing networks frequently, heavy burden would be laid on the HA and the FAs. The inefficiencies of Mobile IP could cause severe quality of service degradation for mobile users. To reduce the effect of these inefficiencies, micro mobility protocols that employ a hierarchy of FAs are introduced. IP micro mobility protocols can complement Mobile IP by offering fast and almost seamless handover control in limited geographical areas [24].

Micro Mobility Support

Micro mobility protocols operate in a restricted administrative domain and provide the MH within that domain with connections to the core network, while keeping signalling cost, packet loss, and handover latency as low as possible [25]. The basic idea behind all micro mobility protocols is the same: to keep the frequent updates generated by local changes of subnets away from the home network and only inform the HA of major changes, e.g. changes of domain. Several micro mobility protocols have been proposed, such as Cellular IP [26], Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [27], Intra Domain Mobility Management Protocol (IDMP) [28] and so forth.

In this section, Cellular IP [26] is as an example to show the basic mechanism of micro mobility solutions. Cellular IP [26] is proposed to provide local mobility support for frequently moving mobile users. By installing a Cellular IP Gateway (CIPGW) for each domain, the scheme localises handover without performing renewed registration. The CIPGW acts to the outside world as a foreign agent. The packets destined for a MH reach the CIPGW first, and then are forwarded to the MH based on the collected routing information. By allowing simultaneous forwarding of packets destined for a mobile node along multiple paths, the scheme supports soft handover. For example, with Cellular IP, a mobile user MH moving between the adjacent Subnet A and Subnet B shown in Figure 2.4 could be made to receive packets via both attached subnets. In Cellular IP, the routing cache is applied to maintain the position of an active MH up to subnet level accuracy. The routing states in the routing cache are dynamically updated when the MH hands over to other subnets.

Micro mobility solutions such as Cellular IP utilise host-specific routes in the routers to forward packets. Apart from routing-based solutions, micro mobility can be achieved using hierarchical registration and tunnelling to reduce mobility-related signalling load. IDMP manages mobility of users in a hierarchy of two levels: The first hierarchy covers different network domains; The second hierarchy consists of subnets within one domain. A new entity, Mobility Agent (MA), is placed in each domain for mobility management within that domain. MA provides a Global CoA (GCoA) to the MH for global location resolution. The MH's GCoA is unchanged as long as it is within the associated domain. Another new entity, Subnet Agent (SA), handles mobility management on a subnet level. SA supplies the second CoA to the MH, named as Local CoA (LCoA). The MH updates its LCoA when it changes its subnet. All the packets destined to the MH are tunnelled to the MA first using its GCoA. After being decapsulated at the MA, the original packets are encapsulated for the second time, and sent to the current LCoA of the MH. The encapsulated packets will be received by the SA serving the MH, and then forwarded to the MH after the decapsulation. When the MH moves to another subnet within the same domain, it registers the newly obtained LCoA at the MA. The HA is informed whether the MH has changed its domain. The hierarchical approach localises the scope of registration update and thereby reduces the global signalling load.

Access Mobility Support

Access mobility is often termed as layer 2 or link layer mobility because it largely relies on specific access technologies. Thus, the support on access mobility varies from one network to another. In cellular networks such as GSM, link transfer can be made from one channel to another channel on the same Base Station (BS) or from one BS to another BS, which involves the same network controller. In a GPRS or UMTS network, a virtual link providing layer 2 service is formed composing two segments of the established GPRS Tunnelling Protocol (GTP) tunnel between GGSN and MH. The SGSN handles inter-radio network controller mobility, and the GGSN manages inter-SGSN mobility. When a MH switches to a different Radio Network Controller (RNC) within the scope of the same SGSN, the GTP tunnel is redirected from the serving RNC to the new RNC.

For IEEE 802.11 based WLAN, link layer mobility is supported by the Inter-Access Point Protocol (IAPP) recommended in IEEE 802.11F [29]. The IAPP specifies how the information could be exchanged between APs to achieve multi-vendor interoperability inside a Distribution System (DS). The DS is a collection of interconnected BSSs (The *Basic Service Set* is the basic building block of an IEEE 802.11 WLAN, and consists of a number of APs.). The IAPP is a communication protocol, and does not deal directly with the delivery of 802.11 data frames to a MH. In the IAPP-enabled communication system, a Remote Authentication Dial-in User Service (RADIUS) registry defines the AP members and maintains the mapping of their wireless medium addresses to the DS network layer addresses. When a MH attaches to a new AP, the new AP retrieves the IP address of the old AP that the MH was associated with from the RADIUS registry. Then, the new AP requests the MH's context transfer by sending an IAPP MOVE-Notify to the old AP. The context block will be carried in an IAPP MOVE-Response message returned by the old AP. With the IAPP procedure, handover reassociation can be invoked without the involvement of IP network layer.

2.3 Handover Methodology

2.3.1 Handover Strategies

With a well engineered interworking architecture, handover across heterogeneous wireless networks can be made possible. Considering current widespread deployment of cellular networks, it is reasonable to assume that a MH is within the coverage range of at least one base station at all times. The dimensions of a base station’s coverage depend upon various factors such as network type, transmission power and so forth. Therefore, a key issue for both network and mobile user is to reach a decision as to which network would be selected, and how handover should be handled when a link transfer is necessary. In an interconnected heterogeneous wireless infrastructure, the coverage of different wireless networks may be overlapped in some service areas as illustrated in Figure 2.1.

Handover Control

As wireless networks evolve towards the 4th Generation (4G) wireless systems consisting of individual heterogeneous wireless networks, the complexity of handover process will increase. Handover control is about which entity should be in charge of controlling handover procedure. The scheme whereby network controls handover is called Network-Controlled Handover (NCHO). Similarly, in Mobile-Controlled Handover (MCHO), mobile user exercises control. The third approach whereby network controls handover but with assistance from mobile user on measurements of radio links is called Mobile-Assisted Handover (MAHO).

In the Network-Controlled Handover (NCHO), a base station BS monitors the signal strength and quality of a MH. If the measure deteriorates below a certain threshold, the network arranges a handover of the MH to a new BS. In this case, the network is in charge of making handover decision for the MH. The network may ask its controlled BSs in vicinity to make measurements for the MH, and then chooses a new BS being capable of providing the best quality of service. In addition, with the coordination of a central network entity, the overall network load can be intelligently distributed among the BSs. Nevertheless, a fast moving MH could incur heavy signalling traffic for exchanging link measurement data. Without adequate radio resources at BSs to make frequent measurements of neighbouring links, the handover execution time could be in the order of seconds according to [30], which is highly undesirable. Therefore, this type of handover control is unsuitable for a network with densely populated mobile users due to the associated delay. The NCHO was used in the first generation analog systems such as Advanced Mobile Phone Service (AMPS).

The Mobile-Assisted Handover (MAHO) is a variant of the network-controlled handover. In the MAHO, a network gathers the link-related information provided by a mobile user, who may be instructed to measure the signal strength of nearby BSs. The network makes handover decision by taking into account mobile supplied measurement. It decides whether a handover should be made, and to which BS. The MAHO is widely used in the second generation mobile systems such as GSM.

The two schemes mentioned earlier use a centralised approach, in which one single entity located at the network end, makes handover decision for mobile users. In contrast, the Mobile-Controlled Handover (MCHO) uses a decentralised approach. The MCHO is employed by both the European DECT and the North American PACS air interface protocols [30]. In the MCHO, a MH is completely in control of handover process. The MH keeps examining radio link quality including signal strength and interference levels on all the available channels. A handover is initiated by the MH when the radio link quality of the serving BS drops below a certain threshold. Since the MH is unaware of other mobile users, it simply triggers a handover to a selected BS, the one with the strongest signal strength (RSS). This type of handover control allows for faster handover decision and is effective in reducing handover latencies for high mobility inside micro-cellular systems [31].

A number of publications [31-33] have provided insight into various handover control strategies for next generation wireless networks. *Zhu et al.* developed a policy-based two-element model in [32] to analyse policy-based handover control. Two conceptual elements: Policy Enforcement Point (PEP) and Policy Decision Point (PDP) were introduced. According to whether two elements are located in the same entity (e.g. a network node or a MH), two possible handover control architectures are identified: 1) The PEP and PDP located in BS/AP for NCHO/MAHO; 2) The PEP and PDP located in MH for MCHO. *Calvagna et al.* summarise the advantages and disadvantages of the two different approaches in [31]. *Aguiar et al.* demonstrate in [33] how the network-controlled and mobile-controlled approaches can be implemented for the future all-IP-based 4G networks using the knowledge obtained from practising Daidalos project. Both approaches were proved to be viable in terms of scalability and QoS support if an appropriate architecture like Daidalos is implemented.

Generally, the suitable handover control scheme for an integrated wireless infrastructure is coupled with a specific integration architecture. The NCHO allows optimised resources management to be applied in the network, but may need extensive collaboration between network domains. This makes it harder, if not impossible, to deal with inter-operator handover. The MCHO ensures that handover decision is made timely towards maximising service quality by taking only individual needs into

consideration. This is due to the fact that MCHO-enabled mobile user is always in a position of discovering neighbouring networks irrespective of their trust relations in between each other. However, the sum of every individual user's convenience will hardly result in an efficient global network resources management [31]. Therefore, in the tightly coupled integration architecture, the NCHO and MAHO are the most desirable solution from a network operator's perspective. In contrast, the MCHO is preferred for the loosely coupled and peer-network integration architecture, in which less collaboration between network operators may be available or information exchange would incur hefty signalling traffic.

2.3.2 Handover Algorithms

Once a handover control scheme is determined for a specific integrated network, relevant handover procedure can be executed. Handover procedure in a heterogeneous environment is more complex than that in a homogeneous wireless network, which is single technology based. As a rule of thumb, the switching of underlying access technologies should be kept transparent to higher layer applications of a mobile user during a handover.

Figure 2.5 shows the handover procedure when the MCHO is applied in the integration of 3G PLMN and WLAN networks.

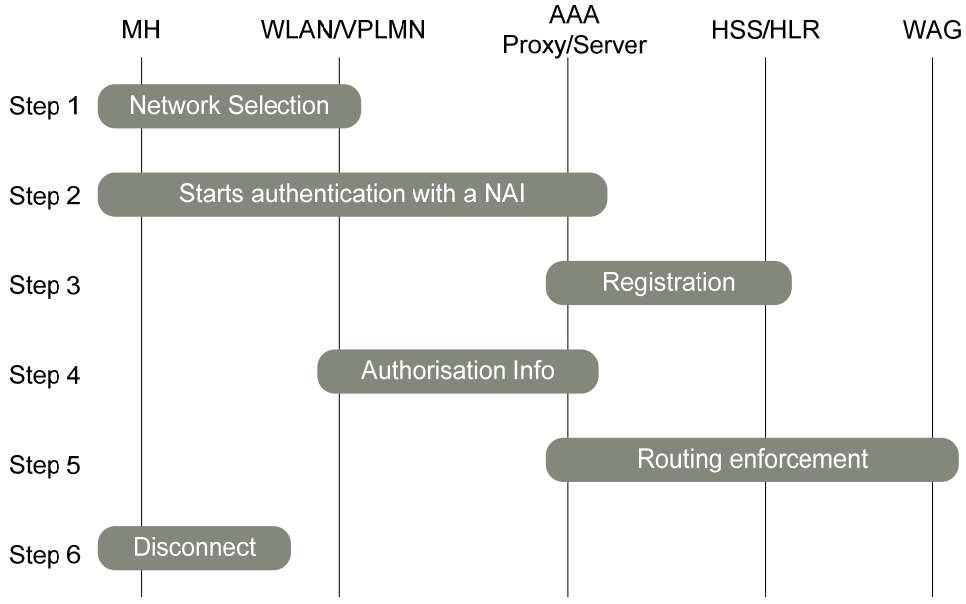


Figure 2.5 Handover procedure in the integration of 3G PLMN and WLAN networks

When a MH is moving into a WLAN's radio coverage, it selects an Access Point (AP) to establish the WLAN connection. The network selection is determined by the implemented handover decision algorithms at the mobile handset (Step 1). Then, the MH starts to get itself authenticated with the AAA server using its Network Access Identifier (NAI) (Step 2). The NAI may be modified and resent if the visited WLAN is unable to route the authentication request to the AAA server. The AAA messages may have to be routed via several AAA proxies, which could reside in the third-party networks. Once the MH's identity is verified, the AAA server registers the MH's 3G AAA server to the Home Subscriber Server (HSS) (Step 3) and sends the connection authorization information to the WLAN for tunnel establishment (Step 4). The WLAN stores the keying materials and tunnelling attributes for later communication with the MH. The routing enforcement (Step 5) at the WLAN Access Gateway (WAG) is optional for the MH to use 3G packet-based services. The WAG is on the path between the MH and the Packet Data Gateway (PDG). After the MH associates with the new network, in this case, the WLAN, it may explicitly indicate to the previously attached network to perform disconnection (Step 6).

During a handover across heterogeneous wireless networks, once the target network is determined, the following procedure is defined by the interworking mechanisms including mobility protocols, authentication methods, integration architecture and so forth. Handover algorithms are employed to determine the target network for handover, and make the corresponding decision. This is driven by the introduction of "Always Best Connected" (ABC) concept [34] in mobile service provisioning.

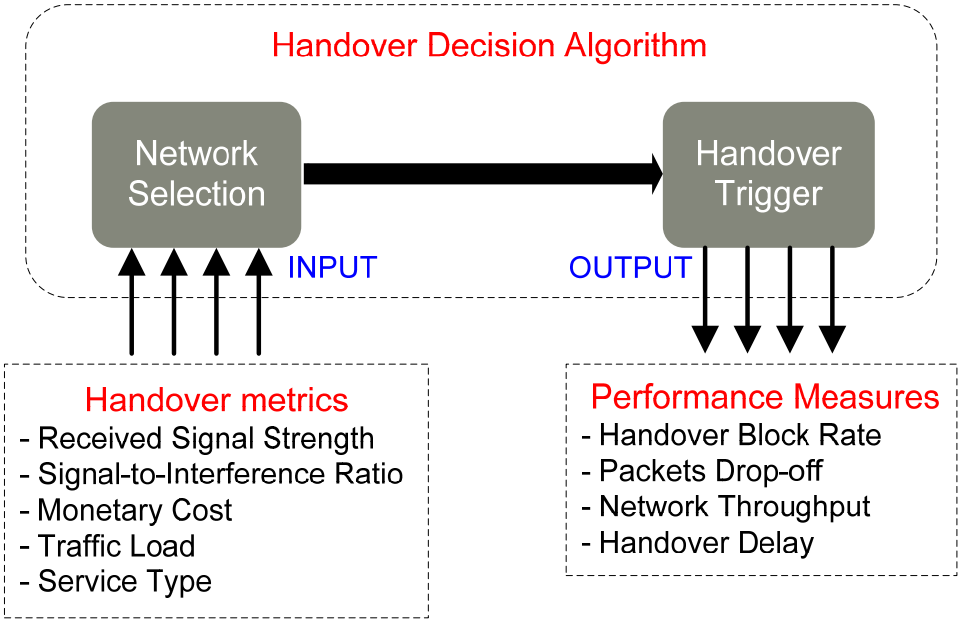


Figure 2.6 Block diagram of handover algorithm

Basically, a handover algorithm deals with two essential tasks in a handover: *network selection* and *handover triggering*. Network selection determines WHERE a mobile user would be switched to when a handover is necessary. Practically, it is usually complemented by a separated handover triggering process, which decides WHEN the switching action to the selected network should be triggered. These two processes are seen as two consecutive steps in a handover as shown in Figure 2.6. Firstly, the handover executor selects the best available network [34] for attachment. The network selection could be based on the periodic evaluation on a series of handover metrics of access networks. The executor can be located in a mobile terminal as required by MCHO, or in a BSC for NCHO. Then, the handover executor triggers the switching to the selected network at an appropriate time. The handover initiation time is carefully calculated to avoid unnecessary back and forth handover between two points of attachment, which is also known as ping-pong effect.

Chapter 3

HANDOVER SECURITY

3.1 Introduction

A variety of handover related issues, which lie in two categories: mobility engineering and handover methodology have been discussed in Chapter 2. These research issues have been raised in the literature mainly to address the heterogeneities arising from the interworking of different wireless networks. However, they are incapable of addressing seamless mobility when multiple network operators need to collaborate to extend service coverage for their subscribers. For example, a mobile user roaming globally may be unable to efficiently deal with a growing number of network operators. The demand for seamless handover in a multi-operator environment creates new research challenges.

In recent years, the consumer market has witnessed the success of some new access network technologies such as IEEE 802.11 WLAN (better known as Wi-Fi). Other upcoming technologies, e.g. IEEE 802.16 (also known as WiMax), are being commercially deployed. The increased affordability and other benefits such as high bandwidth of these new access methods are encouraging both wireless Internet Service Providers such as SpeedNet Services and retailers such as Starbucks to deploy their Wi-Fi services in airports, train stations, hotels and so forth. Traditional cellular network operators such as T-Mobile are beginning to provide Wi-Fi services. According to a British consulting firm, Analysys Consulting Ltd. (Cambridge), the current cost of transferring 1 Mb over an IEEE 802.11 network is between 0.2 and 0.4 eurocent, compared with between 3 and 38 eurocents for GPRS networks [35]. Till 2007, 66,921 Wi-Fi hot spots have been deployed in the U.S., up 56% from the previous year [36]. Endorsed by the certification programs such as Wi-Fi ZONE from Wi-Fi Alliance [37], Wi-Fi deployment will continue to grow. This situation is further underpinned by the

technology advances, e.g. Wi-Fi mesh architecture [38], which enables the delivery of Wi-Fi services in citywide areas. The global Wi-Fi sharing scheme *FON* [39] allows any person to share his Wi-Fi broadband access at home/work by creating a network mesh. This makes it possible that individual Wi-Fi owner (user), so-called *Foneros* in FON, offers services to others as a public Wi-Fi operator.

On the other hand, the limited reach of Wi-Fi propagation and the high cost of installing and maintaining a wired network backhaul connection have limited Wi-Fi network's deployments to homes, offices, public hot spots, and some wide-area hot zones [38]. It is still difficult to provide high speed wireless access to mobile users beyond metropolitan scale using single access technology. This determines that the global Wi-Fi service can be served jointly by a considerable number of Wi-Fi operators, due to its financial appealing.

Considering the large installed base of Wi-Fi hot spots, it would be an economical choice to integrate these high-bandwidth-enabled access points with legacy cellular networks than build a new network from scratch in regards to providing seamless service. The co-existence of multiple network operators and the interworking of different wireless networks is expected to be the prevailing model [40]. Consequently, the interoperability between multiple network domains belonging to different network operators becomes a key issue to achieve seamless mobility in global scale. New research challenges would arise from working with a growing number of network operators and their versatile trust relationships in between each other.

The interoperability requires the proper trust relationships to be established between network domains belonging to different operators. This trust relationship may be enabled in a form of either a direct peer-to-peer roaming agreement or an indirect roaming agreement brokered through a third party. From the perspective of technical implementation, establishing, maintaining and verifying trust relationships translate into Authentication, Authorisation and Accounting (AAA) related operations. In a multi-technology and multi-operator scenario, the overall security solution needs to balance the specific AAA mechanisms implemented in each interconnected heterogeneous network. Therefore, security enhancements would be necessary for efficient handover across heterogeneous wireless networks.

3.2 Authentication, Authorisation and Accounting (AAA) in Wireless Systems

3.2.1 Concepts of AAA

Authentication, Authorisation, and Accounting (AAA) defines a framework for controlling access to distributed network resources. It provides a means of enabling trust relationships between heterogeneous network domains, and thus can be utilised to facilitate network interoperability. With the proper mechanisms in use, they can reduce the vulnerabilities of a network operator and its subscribers to various security threats such as eavesdrop, identity theft and so forth. In this section, each of the "A"s in the AAA will be described.

Authentication

Authentication is the process in which an entity proves its identity to another party, for example, by showing photo ID to a bank teller or entering a password on a computer system. This process is broken down into several methods which may involve something the user knows (e.g. password), something the user has (e.g. card), or something the user is (e.g. fingerprint, iris, etc). The acts of providing proof and verifying the authenticity of the identification presented are the two acts of authentication.

In a wireless network, a mobile user wishing to gain access to a network presents its identity along with a set of credentials. The credentials are then used by the network to verify that the mobile user is actually what it claims to be. A set of messages can be exchanged between the mobile user and the network for identity verification. Two authentication key exchange models are presented here to show the available authentication message exchange approaches.

The two-party authentication exchange model is used when the two peers interact with each other through a direct line of communication without the involvement of any middle nodes such as gateways or proxies [40]. Many key exchange mechanisms such as Internet Key Exchange (IKE) [41] and Wireless Encryption Protocol (WEP) [42] can

be employed to exchange a set of request/response messages directly between two entities to establish a security association.

As the network grows in size, the two-party authentication exchange model has been extended to a three-party authentication exchange model, which is more scalable. This created functional separation between "authenticator" and "authentication server". Intuitively, three main entities would be involved in the three-party authentication exchange model as shown in Figure 3.1:

- Supplicant: a supplicant is an entity at one end of a point-to-point link that is being authenticated by an authenticator attached to the other end of that link [17]. The supplicant can be an IEEE 802.11 client station (STA) or a UMTS terminal.

- Authenticator: an authenticator is an entity at one end of a point-to-point link that facilitates authentication of the entity attached to the other end of that link [17]. This entity often appears as Network Access Server (NAS) in the AAA model that will be discussed later.

- Authentication Server: an authentication server has the real authority and maintains user credentials database. It determines, from the credentials provided by the supplicant, whether the supplicant is authorised to access the services provided by the authenticator [17].

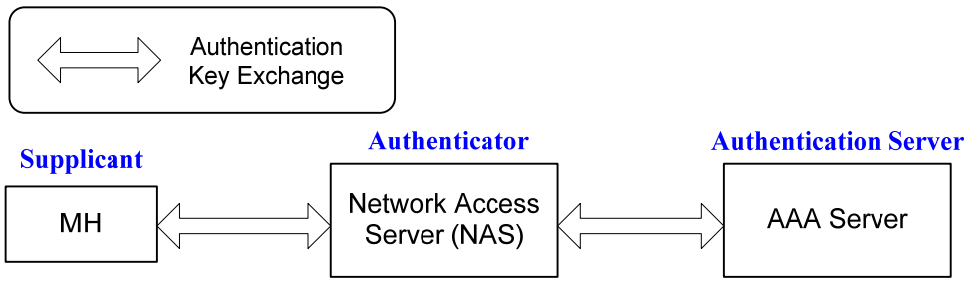


Figure 3.1 The three-party authentication key exchange model

A separation between the role of "authenticator" and the role "authentication server" is very suitable for deploying a large-scale network that consists of multiple domains. Inevitably, the point-to-point authentication exchange is required between the supplicant and the authentication server. By taking part in the exchange of authentication key, the

authenticator controls communications into and out of the wired network, and acts as a protocol dividing point [40]. As a result, the communications between the authenticator and the authentication server can use a standard protocol, e.g. Extensible Authentication Protocol (EAP) [18], to carry authentication messages. The protocols for communications between the authenticator and the supplicant may vary according to the type of access technology involved. For example, IEEE standard 802.1x [17] can be employed for a IEEE 802.11 WLAN to do a four-way handshake exchange between a STA and an Access Point (AP). With the mediating of an authenticator, the authentication across heterogeneous wireless systems is supported.

Authorisation

Authorisation is the process of granting a particular privilege for the access to a service or information based on a user's presented credential. While authentication attempts to establish a level of confidence that a certain thing holds true, authorisation decides what a user is allowed to do. For example, a mobile user purchases a pre-paid SIM card that is supposed to include the credits for 60 mins phone calls. Every time the user requests to make a phone call, the network must check to see whether there is sufficient credit left on the user's account before allowing the user to connect. The decision on authorisation may be restricted by a number of factors: e.g. key lifetimes, Service Set Identifier (SSID) restrictions, called-station-ID restrictions suggested for IEEE 802.11 in [43].

A simple authorisation process is described as follows. Upon receiving the request for access attachment, the network operator first consults an authorisation server that holds user profiles. A good example of the authorisation server is the Home Location Register (HLR) in a GSM network. Then, the network makes a decision on whether the user is authorised to use the service it has requested. Authorisation requests are processed after the user has been authenticated. It makes sense that both authentication and authorisation functions could be implemented in the same server.

The authorisation example shown above assumes that the access request is made through a network that is administered by the same network operator. In a more complex case, the mobile user may be served by a foreign network operator. This

requires that the two operators have a trust relationship in between each other, and have agreed on a charging scheme. When a user requests access through the network of a foreign operator, the foreign operator forwards the access request to the user's home network. The user's authorisation profile maintained by its home operator will then be consulted for further actions. The foreign operator must ensure that the access provided to the user is within the established roaming agreement with its home operator. The final authorisation decision is made by the home operator, and the results will be sent back to the foreign operator. In this case, the services and the trust relationships are taken into account in the authorisation.

Accounting

Accounting measures the resources a mobile user consumes during access, and makes the corresponding bill according to the service agreement between the mobile user and the network operator. This can include the amount of connection time or the amount of data usage a user has incurred during a session. Accounting is carried by keeping records of usage information, and involves a number of activities including auditing and reporting.

- Auditing: the act of verifying the correctness of an invoice submitted by a network operator, or the conformance to usage policy, security guidelines, and so on [40].

- Reporting: the act of providing a 'trail' in the event that the system is compromised or found faulty.

3.2.2 Authentication in Cellular Networks

One-way Authentication Method

In GSM, authentication takes place in one direction. The network always verifies the identity of a mobile user, but the mobile user is unable to verify the identity of the network. One-way authentication is widely employed in the 2G wireless systems, e.g. GSM. GSM security is addressed in three aspects: authentication, confidentiality and anonymity. Authentication is carried by network operator through verifying the user's knowledge of a subscriber authentication key K_i that is stored in both the

Authentication Centre (AuC) and the Subscriber Identity Module (SIM). Confidentiality and anonymity on the radio path is provided by encrypting data streams between mobile user and access network. However, GSM was not designed to protect against active attacks on the radio path, because they would require an attacker to masquerade as a GSM network [44].

In a GSM network, authentication is often involved in many system operations such as mobile registration, mobile handover and so forth. To initiate an authentication process, the AuC in the home network generates a 128-bit random number (*RAND*). This random number *RAND* will be sent to the Mobile Host (MH) as a challenge. Then, both the MH and the AuC produce a 32-bit signed response *SRES* by applying a vendor-specific *A3* algorithm, $SRES = A3(RAND, K_i)$. The MH sends its *SRES* to the HLR for verification. The HLR checks whether the *SRES* from the MH is identical to what it has obtained from the AuC. Alternatively, the *SRES* and the *RAND* could be provided to the VLR in advance. In this case, the SRES comparison can be done locally at the VLR on the visited network.

If the MH is accepted for access, an encryption key K_c is generated by another vendor-specific algorithm *A8*, $K_c = A8(RAND, K_i)$. The K_c is produced by the MH and the AuC separately. Then, the AuC sends its copy of K_c to the visited network. The visited network applies an *A5* encryption algorithm to cipher the data streams between the BS and the MH. The ciphering $data^* = A5(K_c, F_n)\ XOR\ data$ will be applied on both directions. F_n is a 22-bit frame number, and 114-bit data are used as input.

Although GSM provides many security measures to protect against some typical attacks such as eavesdropping, unauthorised access and masquerade, it is still vulnerable to false base station attacks [44]. As the costs of mobile base equipments reduce greatly and such kind of attacks become easier to implement, GSM mobile users would inevitably be at a risk when roaming globally.

Mutual Authentication Method

Having backward compatibility with GSM [19], Universal Mobile Telecommunications System (UMTS) defines many new security features. One of the most important enhancements is mutual authentication, which consists of two integral parts: *subscriber authentication* and *network authentication*. With subscriber authentication, a serving network verifies the identity of a subscriber. Furthermore, the subscriber can corroborate that it is connected to an authorised serving network by conducting network authentication.

The authentication procedure of a UMTS network is executed in two stages: 1) the distribution of Authentication Vectors (AVs) from the home network to the serving network; 2) the Authentication and Key Agreement (AKA) procedure between the MH and the serving network.

The provision of AVs is invoked by sending an authentication data request with the identity of the MH, e.g. International Mobile Subscriber Identity (IMSI), to the home network. After receiving an authentication data request, the home network generates an ordered array of *n* AVs. Each AV, also called a quintet, consists of five components: a random number *RAND*, an expected response *XRES*, a cipher key *CK*, an integrity key *IK*, and an authentication token *AUTN*. Each quintet is generated at the AuC as follows:

$$XRES = f2_K(RAND)$$

$$CK = f3_K(RAND)$$

$$IK = f4_K(RAND)$$

The authentication token *AUTN* is represented as:

$$AUTN = SQN[\oplus AK] \,||\, AMF \,||\, MAC - A$$

where the sequence number *SQN* is generated by the home network, and maintained for each individual user. The " $\oplus$ " is bit-wise exclusive or operation. The optional anonymity key *AK* is produced by $f5_K(RAND)$. The authentication management field *AMF* and the message authentication code for authentication *MAC-A* are two values

included in the *AUTN*. The *MAC-A* is generated by executing the *f1*: $MAC\text{–}A = f1_K(RAND, SQN, AMF)$.

Table 3.1 UMTS security algorithms [45]

Algorithms	**Functions**	**Location**
f0	the random challenge generating function	AuC
f1	the network authentication function	AuC and USIM
*f1**	the re-synchronisation message authentication function	AuC and USIM
f2	the user authentication function	AuC and USIM
f3	the cipher key derivation function	AuC and USIM
f4	the integrity key derivation function	AuC and USIM
f5	the anonymity key derivation function	AuC and USIM
*f5**	anonymity key derivation function for the re-synchronisation message	AuC and USIM

Figure 3.2 gives an example of how the authentication and key agreement AKA between the MH and the serving network is conducted on a UMTS network. After the radio connection is set up between the MH and the Radio Network Controller (RNC) (*STEP 1*), the serving network identifies the MH and requests its Temporary Mobile Subscription Identity (TMSI). If no AV is available for this user, the VLR sends an authentication data request to the MH's home network (HLR/AuC) (*STEP 2*). The AuC generates and sends an array of AVs to the VLR according to the aforementioned AV distribution procedure (*STEP 3*). Upon receiving the AVs, the VLR sends a user authentication request with the *RAND* and the *AUTH* to the MH (*STEP 4*). The MH computes the anonymity key *AK* using the *f5* algorithm and retrieves the sequence number $SQN = (SQN \oplus AK) \oplus AK$. Then, the MH computes the *MAC* using the *f1* algorithm with the *SQN*, *RAND* and *AMF* as inputs. The computed *MAC* is compared with the *MAC-A* included in the received *AUTN*, and this verifies the network. If they are identical, the received sequence number *SQN* will be further checked. If the sequence number is within the correct range, the MH produces a response

$RES = f2_K(RAND)$, and sends it back to the serving network (*STEP 5*). The serving network verifies the identity of the MH by comparing the user's response *RES* with the expected response *XRES*. If the two values are identical, the MH will be accepted for accessing the network.

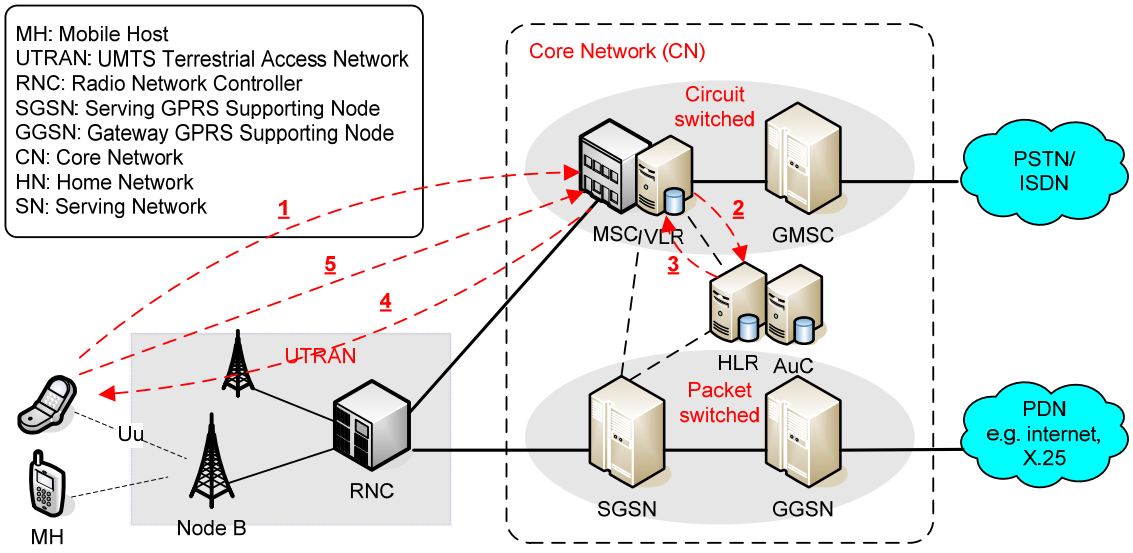


Figure 3.2 Authentication procedures in UMTS network

3.2.3 Authentication in IEEE 802.11

IEEE 802.11 standard [46] has defined two security services: the authentication service and the Wired Equivalent Privacy (WEP) mechanism. Both of the services have been classified as pre-RSNA (Robust Security Network Association) security mechanisms, and proved to be vulnerable [42]. To eliminate the security flaws of its ancestor 802.11, a new standard IEEE 802.11i [42] has been developed. In addition to providing the legacy security services of 802.11, e.g. WEP, 802.11i enhances key management and encryption algorithms by incorporating IEEE 802.1X [47], a port-based network control mechanism.

IEEE 802.1X defines a means of authentication and authorisation at link layer for IEEE 802 Local Area Network (LAN). As described in the 802.1X standard [47], both the supplicant and the authenticator have a Port Access Entity (PAE), through which the authentication between the two parties can be performed. The PAEs operate the algorithms and protocols associated with the authentication mechanisms. The

authenticator PAE exchanges 802.1X messages with the supplicant PAE through the uncontrolled port before the supplicant is authenticated. The exchange of data traffic is allowed through the controlled port after the supplicant is authenticated successfully. The 802.1X utilises Extensible Authentication Protocol (EAP) [18] to provide a variety of authentication mechanisms. The EAP does not have an addressing mechanism and has its messages encapsulated over EAP Over LAN (EAPOL) protocol between the supplicant and the authenticator.

The 802.11i defines two classes of security algorithms for IEEE 802.11 networks: Robust Security Network Association (RSNA) and pre-RSNA. A Robust Security Network (RSN) is a security network that allows the creation of robust security network associations, RSNAs, between all stations [48]. The RSNA security comprises two security algorithms: IEEE 802.11 entity authentication and WEP. The 802.11i standard suggests that pre-RSNA methods that have already been included in IEEE 802.11 [46] will be implemented to aid migration to RSNA methods. The key management defined for RSNA authentication will be presented here for further elaboration on the enhanced authentication mechanism in 802.11 WLAN.

When the IEEE 802.1X authentication is used, the supplicant PAE initiates the authentication to the authenticator by sending an EAPOL-Start message to the authenticator. As shown in Figure 3.3, the authenticator replies with an EAP-Request/Identity to obtain the user's identity. The user then sends back an EAP-Response/Identity containing its identity in response to the received EAP identity request. Upon receiving the EAP Response, the authentication PAE needs to deliver the EAP response message to the authentication server. The communications for authentication between the authenticator and the authentication server can be achieved using the AAA protocols like Remote Access Dial In User Service (RADIUS, RFC 2865 [49]). The authenticator encapsulates the EAP-Response/Identity message in a RADIUS Access-Request message, and sends it to the RADIUS authentication server. Multi-round authentication message exchanges will be needed to verify the identities of both EAP entities (the supplicant and the authentication server as shown in Figure 3.3). The verification can be carried out by means of Extensible Authentication Protocol-Transport Level Security (EAP-TLS, RFC 2716 [50]) protocol, which is outside the

scope of this thesis. If the authentication is successful, some keying materials such as Pairwise Master Key (PMK) will be delivered to the authenticator from the authentication server.

A four-way handshake follows the 802.1X EAP authentication to negotiate the pairwise cipher suites for the local transmission to the AP. The authenticator issues an Authenticator Nounce (ANounce) in an EAPOL-Key message sent to the supplicant. The ANounce is essentially a random or pseudo-random value. After receiving the EAPOL-Key, the supplicant generates a Supplicant Nounce (SNounce). By using a Pseudo-Random Function (PRF) algorithm with the ANounce, SNounce, PMK, and other information as inputs, the supplicant derives a Pairwise Transient Key (PTK). The supplicant then sends an EAPOL-Key message containing the SNounce and Message Integrity Code (MIC) (*Note: MIC is a cryptographic digest used to provide integrity service.*) back to the authenticator. The authenticator uses the same PRF algorithm to derive the PTK. The PTK is a session key shared between the supplicant and the authenticator. Later, the authenticator can start the group key handshake for configuring a Group Temporal Key (GTK) on the supplicant to protect the broadcast/multicast messages.

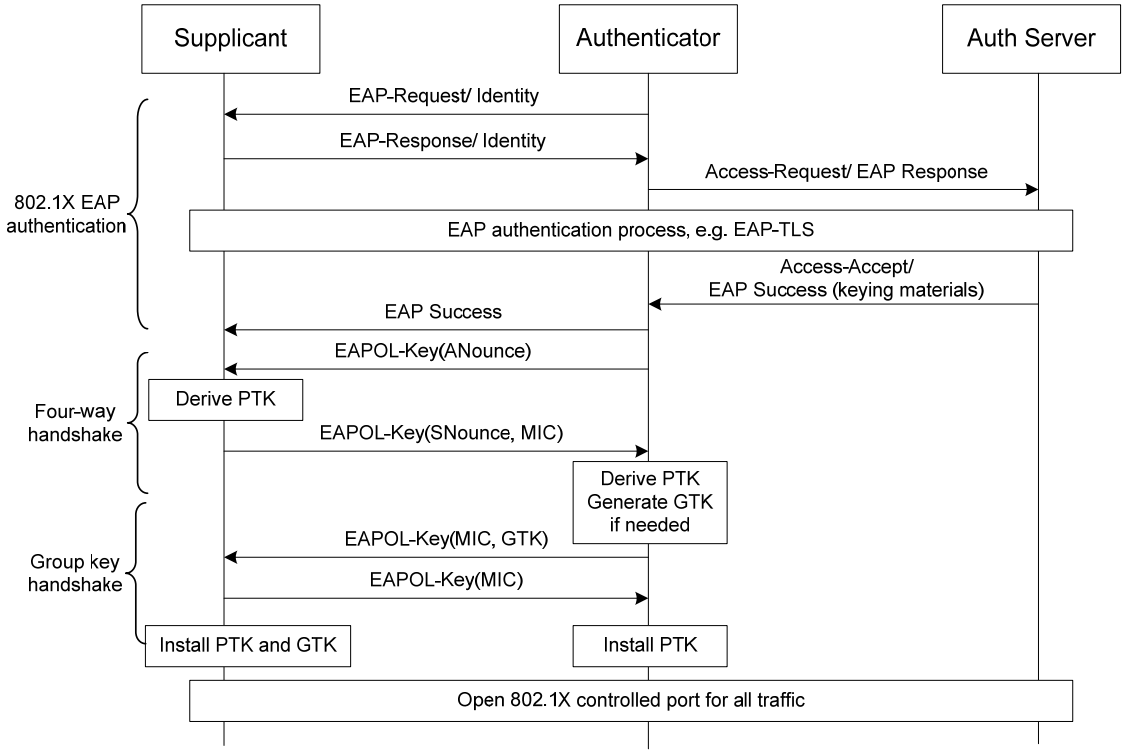


Figure 3.3 Example: an exchange of RSNA authentication messages

The derivation of PTK for session protection is described below. The pseudo-random function PRF uses the PMK and the nonces provided by the two entities, and generates a PTK.

$$PTK = PRF - X(PMK, "pke", Min(AA, SPA) \| Max(AA, SPA) \| Min(ANounce, SNounce) \| Max(ANounce, SNounce))$$

where the output *PRF-X* can be 384 or 512 bits depending on the confidentiality algorithms utilised, and AA and SPA represents the MAC addresses of the authenticator and the supplicant respectively. The addresses are converted to positive integers first, and then compared for the Min and Max operations. The pairwise key expansion "pke" is a fixed character string. The PTK is split into three portions: EAPOL-Key Confirmation Key (KCK), EAPOL-Key Encryption Key (KEK), and Temporal Key (TK) shown in Figure 3.4. The KCK and KEK are used by 802.1X to provide data origin authenticity and confidentiality respectively in the four-way handshake and group key handshake. As indicated in Figure 3.4, the derivation of the corresponding key value is conducted using the *L(Str, F, L)* function, which extract bits *F* through *F+L-1* from *Str* starting from the left. All the temporal keys can be refreshed to prevent key reuse. This provides dynamic key distribution that significantly enhances the security provided by WEP [48].

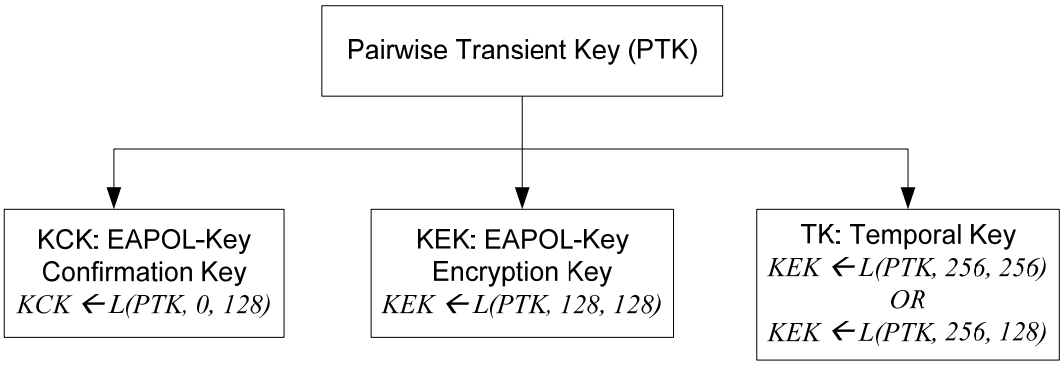


Figure 3.4 Pairwise transient key structure

3.3 Security for Efficient Handover Across Heterogeneous Networks

The emergence of new access technologies has led to a variety of authentication solutions. As wireless networks are evolving towards an integrated architecture, the demand for the standardised security solutions in support of efficient handover across heterogeneous wireless technologies has increased. Current AAA solutions for wireless networks are proposed for individual wireless network, but lack a generic approach for the interworking of heterogeneous wireless networks [51]. Therefore, they need to be extended to build more general AAA services to facilitate seamless handover when heterogeneous systems are involved.

3.3.1 Generic AAA Architecture

In a heterogeneous environment with multiple network domains belonging to different operators, each administrative domain may have an AAA server for managing its own subscribers. To enable interoperability between those heterogeneous domains, their network operators need to cooperate with each other as discussed in Sec. 3.1. A more advanced form of network interoperability - seamless handover, requires that a change of serving network operator that often results in a new trust relationship to be established (see Sec. 3.1) can be kept transparent to mobile users. Therefore, a generic AAA architecture that supports efficient AAA services across heterogeneous networks of different operators becomes the key to the success of an all-IP-based heterogeneous wireless infrastructure.

Figure 3.5 shows a Mobile Host (MH) roaming case, in which the MH roams to disparate networks belonging to different operators. It is assumed that at least one AAA server resides in each network for providing AAA related services. When the MH hands over to a foreign network, its authentication process involves the Foreign AAA server (FAAA) on that network. The MH must be authenticated by the corresponding FAAA to verify its access privileges established through its Home AAA server (HAAA). The FAAA may communicate with the MH's HAAA for authorisation policies. When the FAAA and the HAAA belong to different operators, a trust relationship between the two entities must be present for interoperability. This is referred to as an explicit mutual

trust relationship between two trusted parties [52]. Such a trust relationship can be established through shared security keys or dedicated communication channels. Maintaining trust relationships between different AAA servers for seamless roaming presents additional security requirements.

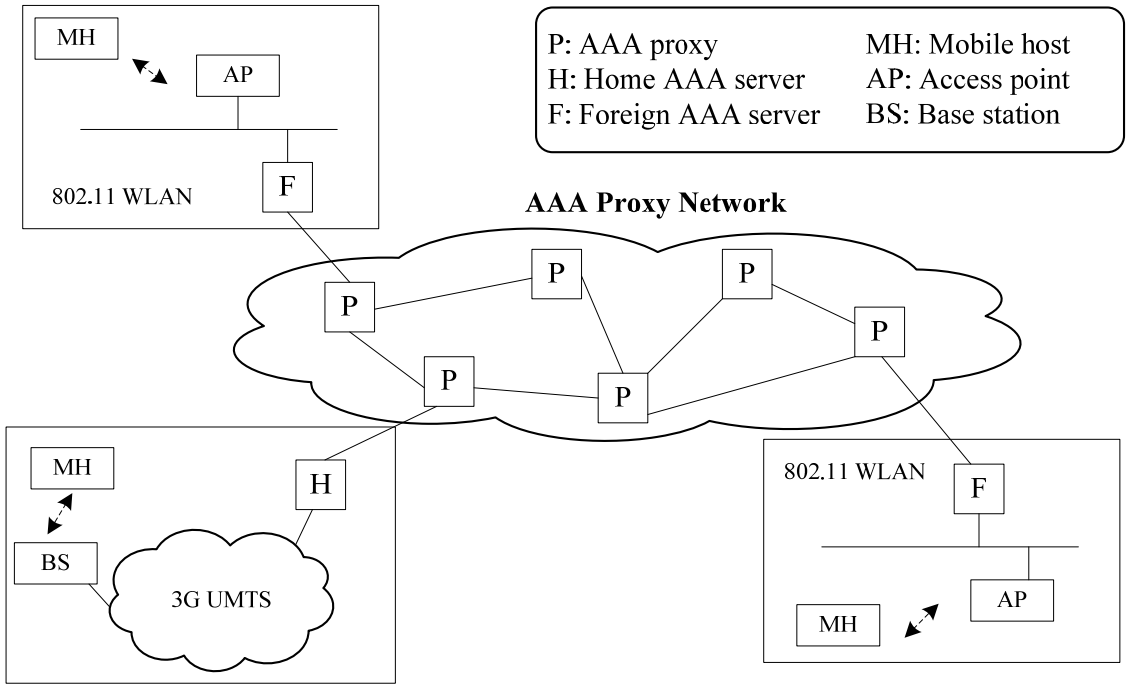


Figure 3.5 A mobile host roaming case in a heterogeneous multi-operator environment

To enable roaming capability for its subscribers, a network operator must establish roaming agreements with other network operators. The most straightforward approach is to have a pairwise agreement with each network operator. However, this approach leads to $O(N^2)$ overhead when N operators are involved in the interoperation. To reduce the number of roaming agreements required, an AAA proxy network is introduced. The AAA proxy network consists of a number of interconnected AAA proxies that hold security associations between each other and may exchange AAA messages for authentication of roaming mobile users. In a roaming case, an AAA message exchange between the FAAA of a visited network and the HAAA may pass through one or more AAA relay hops. With an AAA proxy network, instead of peering with every operator, each operator establishes a roaming agreement with one AAA proxy on the AAA proxy network, which acts as the third-party trust broker. This can

effectively reduce the number of roaming agreements required for interoperation from $O(N^2)$ to $O(N)$ as stated in [53].

In Figure 3.5, three different types of AAA entities are presented: Foreign AAA server (FAAA), Home AAA server (HAAA) and AAA proxy. An AAA server may play these different roles in different contexts. For example, the AAA server in a network may negotiate the identity verification with an external authority (e.g. HAAA) when dealing with a roaming mobile user, and in this case, acts as a FAAA. When processing authentication requests from its own subscriber residing in either its own network or a foreign network, the same AAA server makes authorisation decision and acts as a HAAA. When it has been interconnected with a series of AAA servers with relevant trust associations in place, this AAA server may serve as an AAA proxy for relaying AAA messages from/to other networks. RFC 2903 [54] proposed such a generic AAA architecture that facilitates interoperability between peered AAA servers via a standard AAA protocol.

Figure 3.5 indicates a security model that is applied to the presented AAA architecture. It is important to identify all the trust relationships needed for such an AAA architecture. With the concept borrowed from the social science literature, there is no clear consensus on the definition of trust in distributed computer networks [55]. A trust relationship is enforced on top of a number of security features that are enabled via a security association between two entities. Therefore, it is interpreted as security association in some literatures [53, 56, 57]. A security association between entities X and Y is defined as the combination of the entities' identity information (e.g. Network Access Identifier, NAI), some forms of cryptographic keys (e.g., public keys, preshared symmetric keys), and information on cryptographic algorithms to use in order to authenticate and/or protect data in transit between X and Y [53].

Figure 3.6 shows all forms of trust relationship that need to be present on a generic AAA architecture. First, the MH has a trust relationship $TR_{MH,HAAA}$ with its HAAA, which means that the MH belongs to its home network. Second, the FAAA and HAAA have to trust each other for service roaming; otherwise, they can not exchange AAA messages to pass authentication requests. The trust relationship $TR_{FAAA,HAAA}$ between the

FAAA and HAAA can be established through shared keys, and is governed by a legally binding roaming agreement [58]. Third, a transitive trust relationship $TR_{MH,FAAA}$ is required to make local resources that are under the control of the FAAA in the visited network accessible to the MH. This trust relationship to be established during a handover attachment is denoted as the implicit trust relationship [52], in contrast to the two explicit trust relationship $TR_{MH,HAAA}$ and $TR_{FAAA,HAAA}$.

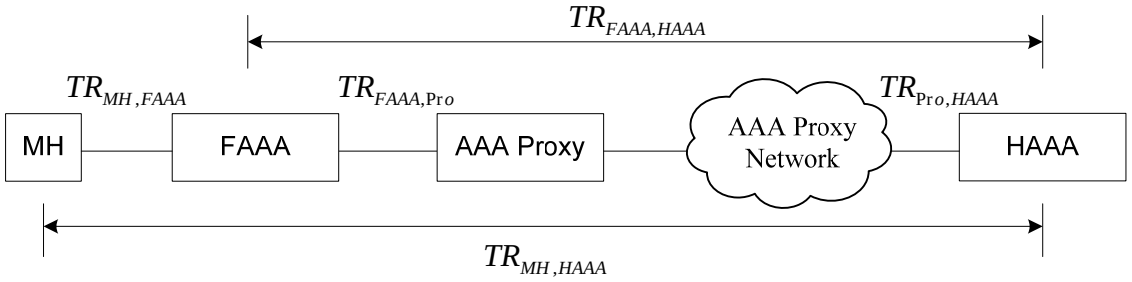


Figure 3.6 Trust relationships on the generic AAA architecture

3.3.2 Access, Authorisation, and Accounting (AAA) Protocols

Modern wireless networks rely on the three-party authentication model (shown in Figure 3.1) as well as a proxy-chaining architecture, where a network access server (as a pass-through) at the edge of a network interacts with a back-end AAA server (for authentication and authorisation) through AAA protocols to conduct access control. Figure 3.7 shows where AAA protocols may play a role in a generic AAA architecture.

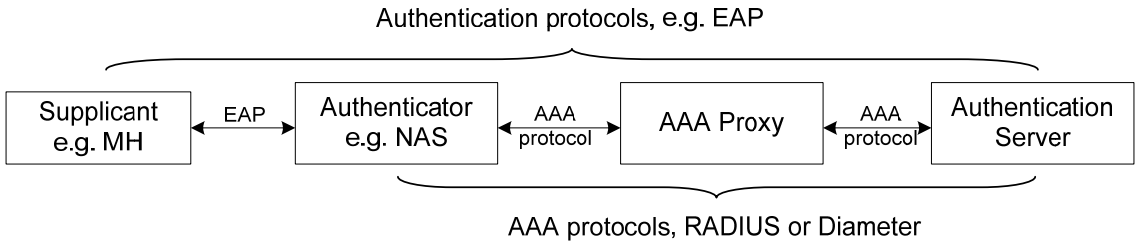


Figure 3.7 AAA protocols in a generic AAA architecture

The authentication protocols such as Extensible Authentication Protocol (EAP) [18] provide a generic authentication framework. The EAP does not perform authentication,

but provides a mean for the communications between a supplicant and its authentication server in a challenge-response manner. The specific authentication methods such as EAP-TLS [50] can be easily introduced by extending the EAP framework. The message flow of the EAP negotiations between the supplicant and the authentication server in the three-party authentication model is described in Sec. 3.2.1. The transmission of EAP messages on the client side (between the supplicant and the authenticator) can run directly on link layer without requiring a network layer protocol. This section would focus on the AAA protocols that carry EAP messages for authentication and key exchanges on the AAA server side (between the authenticator and the AAA server).

Remote Access Dial-In User Service (RADIUS)

Remote Access Dial-In User Service (RADIUS) is a client/server protocol that enables a remote Network Access Server (NAS, as a RADIUS client) to communicate with a central RADIUS server to authenticate users and authorise their access to the requested resources. A RADIUS client is responsible for passing user information in the form of requests to the designated RADIUS server, and waits for a response from the server. The RADIUS server is responsible for receiving user requests, authenticating users, and then returning all configuration information necessary for the RADIUS client to deliver service to the user. The RADIUS server can acts as a proxy client to other RADIUS servers in a proxy chaining architecture that will be discussed later.

The type of RADIUS message is identified by the *Code* field in a RADIUS message shown in Figure 3.8. According to the RADIUS specification (RFC 2865 [49]), eight messages are defined. The *Access-Request*, *Access-Accept*, *Access-Reject*, and *Access-Challenge* are of particular interest to this study. RADIUS carries information (e.g. specific authentication, authorisation, and configuration details) in the form of attributes, which are of variable length. New attribute values can be readily added to extend the functionality of RADIUS server to interact with other entities.

Transactions between RADIUS client and server are protected through the use of a shared secret. In addition, any user passwords sent over RADIUS has to be encrypted. This is done by utilising the RSA Message Digest Algorithm 5 (MD5). As indicated in Figure 3.8, an authenticator field of 16 octets is added to all RADIUS messages. This

value is used to authenticate the reply from the RADIUS server, and is used in the password hiding algorithm. The IETF specification for RADIUS [49] suggests that the response authenticator that is included in Access-Accept, Access-Reject, and Access-Challenge messages can be calculated using one-way MD5 hash:

ResponseAuth = MD5(Code, ID, Length, Request Authenticator, Attributes, Shared secret)

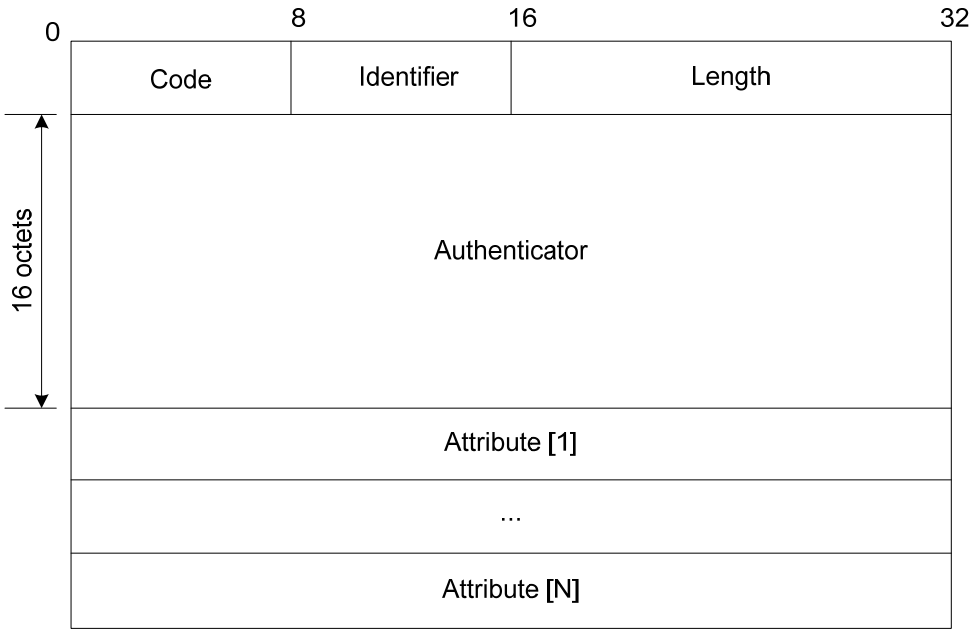


Figure 3.8 RADIUS message format

The transmission of EAP messages over a AAA protocol can be supported by another specification - RADIUS support for EAP (RFC 3579, [59]). The EAP-RADIUS framework allows EAP messages to be embedded inside RADIUS attributes. Two new attributes, EAP-Message and Message-Authenticator, have been introduced in the EAP-RADIUS specification [59] for such purpose. The basic mechanism of carrying EAP messages over RADIUS is explained as follows. The EAP request from a RADIUS server to a supplicant is included in a RADIUS Access-Challenge message by encapsulation. The NAS decapsulates the RADIUS Access-Challenge, and obtains the EAP request, which is then sent to the supplicant through link layer protocols. The EAP response can be delivered to the RADIUS server using a RADIUS Access-Request message in the same manner.

In a roaming case, authentication, authorisation and accounting packets are routed between a NAS and an AAA server through one or more AAA proxies, which constitutes a proxy-chaining architecture as shown in Figure 3.5. With proxy-chaining, two or more administrative entities are allowed to open their networks to each other's dialled-in users for roaming [49]. The benefits of proxy-chaining for roaming has been summarised in six aspects: scalability improvement, authentication forwarding, capabilities adjustment, policy implementation, accounting reliability improvement, atomic operation in IETF specification RFC 2607 for Proxy Chaining and Policy Implementation in Roaming [60], and are supported by other proposals such as Wireless Shared Key Exchange (W-SKE) [53].

In RADIUS, the procedures for proxy chaining are defined to forward AAA packets between a NAS and a RADIUS server through a number of proxies as shown in Figure 3.9. The NAS generates a request and sends it to Proxy 1. Proxy 1 examines and forwards the request to Proxy 2. Proxy 2 then forwards the request to the RADIUS Server. Both Proxy 1 and Proxy 2 may modify the attributes in the packet since proxies are allowed to implement some local policies. On the reverse path, the RADIUS Server generates a reply and sends it to Proxy 2. After receiving the reply, Proxy 2 matches it with the request it had sent, and forwards the reply to Proxy 1. Proxy 1 checks the reply for matching, and forwards the reply to the NAS.

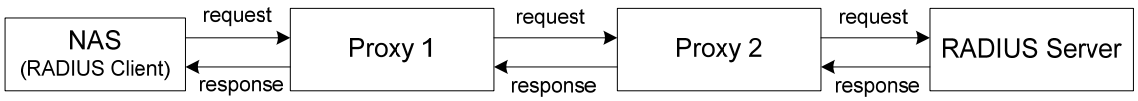


Figure 3.9 Proxy chaining in RADIUS (RFC 2607 [60])

The choice of which server receives the forwarded request is based on the authentication "realm". The authentication realm can be the realm part of a NAI. A RADIUS server can function as both a forwarding server and a remote server: serving as a forwarding server for some realms and a remote server for other realms according to the RADIUS specification [49]. As this indicates, the roaming relationship path (e.g. the path to the next proxy towards the RADIUS server) is determined by the network

access identifier. Most RFCs do not specify the routing procedure along the roaming relationship path when using RADIUS [40].

On the AAA routing path, each proxy can implement its local policies by modifying attributes when forwarding the RADIUS messages. This can be done without providing any notifications, although it risks of being misused by external parties and undetected by the end entities. RADIUS utilises a shared secret between a proxy and a remote server to protect hop-by-hop AAA transmissions instead of end-to-end security between the NAS and the RADIUS server. This would result in a number of security threats such as message editing, attribute editing, replay attacks, connection hijacking and so forth, as stated in the IETF specification [60]. Lacking auditability and transmission-level security features makes RADIUS-based roaming susceptible to fraud perpetrated by the roaming partners themselves.

Diameter Protocol

As the successor to RADIUS, the Diameter protocol has been developed to provide a series of enhancements in response to new requirements on failover, transmission-level security, reliable transport, agent support, capability negotiation, roaming support and so forth, as described in IETF RFC 2989 [61]. Diameter provides an upgrade path for RADIUS.

The Diameter base protocol is defined in RFC 3588 [62] to provide the minimum requirements needed for an AAA protocol. The concept of “Application” is introduced in the Diameter base protocol. A Diameter application is a protocol based on the Diameter base protocol. For example, the interactions of a Diameter server with a NAS for authentication and authorisation is considered an application for Diameter, and is defined in a separate specification RFC 4005 [63]. The Diameter applications such as Mobile IPv4 (RFC 4004), Network Access Server Requirements (NASREQ) (RFC 4005), and EAP (RFC 4072) applications are defined to extend the base protocol by adding new commands and attributes.

Diameter is a peer/peer protocol, where both the client and server can issue request or response in a transaction. In contrast, RADIUS mentioned earlier is a client/server

protocol, where requests are always initiated by the client, while responses (e.g. challenge or accept/reject) are sent by the server.

In Diameter, all data are delivered in the form of Attribute Value Pairs (AVPs). These AVP values can be used by the Diameter protocol itself, and applications that employ Diameter. The Diameter base protocol supports the introduction of new AVPs so as to make Diameter extensible. Instead of using message type (seen as *Code* field in a RADIUS packet), Diameter defines the concept of "*Command*", which is assigned for each command request/answer pair and determines the action to be taken for a particular message. The commands are distinguished by *Command Code* field in the Diameter message as shown in Figure 3.10. *Application ID* is four octets and is used to identify to which application the message is applicable for. *Hop-by-Hop Identifier* is used to match requests and responses on a hop, which is denoted as a *Connection* to be discussed later in Diameter. The sender must ensure that the hop-by-hop identifier is unique on a given connection at any given time [62]. Another field, *End-to-End Identifier*, is used to detect duplicate messages. The originator of an answer message must ensure that the end-to-end identifier value of the message is the same as the value found in the corresponding request. Diameter AVPs carry specific authentication, authorisation, accounting, routing, and security information for the Diameter transactions. The Diameter base protocol defines a large number of AVPs, e.g. Origin-Host AVP, Origin-Realm AVP, Destination-Host AVP and Destination-Realm AVP.

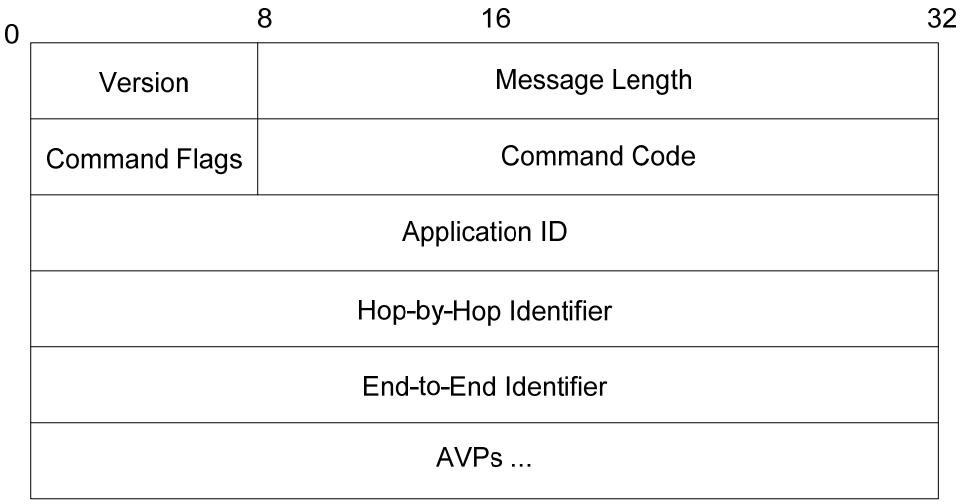


Figure 3.10 Diameter message format

With various types of Diameter messages, AAA requests/responses can be transported within Diameter. Diameter can be run on both Transmission Control Protocol (TCP) and Stream Control Transmission Protocol (SCTP) transport protocols, in contrast to RADIUS which relies on User Datagram Protocol (UDP).

The end to end transmission path between a client and a Diameter server is identified as a session, which is a logical concept at the application layer. A session is processed by end entities and is identified by *Session-ID AVP* in Diameter. A session is established through a number of individual connections, which is a transport level connection between two peers, and used to send and receive Diameter messages. In the example shown in Figure 3.11, two peer connections are established between the client and Diameter server. The user session X spans from the client NAS to the Diameter server crossing an agent. As noted in the Diameter base protocol [62], there is no relationship between a connection and a session, and Diameter messages for multiple sessions are all multiplexed through a single connection.

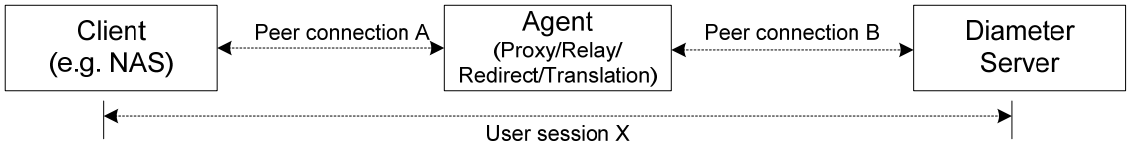


Figure 3.11 Diameter connections and sessions

Unlike the RADIUS protocol, message routing is explicitly defined in the Diameter specification. This effectively addresses the limitations of RADIUS, e.g. vulnerability to attack from external parties, and thus facilitate more secure and scalable roaming in a multi-operator environment. Diameter routing is conducted through the use of "*Peer Table*" and "*Realm-Based Routing Table*", the latter of which is also known as *Realm Routing Table*. A realm is an administrative domain used for routing purposes, and the concept of it is originated from the NAI specification [64]. Peer table is used in message forwarding, and referenced by realm routing table. Each Diameter node keeps a peer table that maintains an entry for each of its peers. Each entry in the peer table contains information on the host identity of the peer, the state, whether a peer entry was statically configured or dynamically discovered, the expiration time for the dynamically discovered peer entry and whether TLS is enabled for communications. Realm routing

table is consulted by Diameter agents to find the message destination or the next AAA hop that may reside in other realms.

A Diameter node can process a request message locally, or forward the message according to the final destination of that message. The *Local Action* value of the realm routing entry in a Diameter node determines how a message with a specific Destination-Realm AVP is processed. The Diameter base protocol [62] defines four types of actions that can be imposed on a message: *Local*, *Relay*, *Proxy* and *Redirect*. It is noted in [62] that Diameter agents must support at least one of the Local, Relay, Proxy and Redirect modes of operation, but do not support all modes of operation.

Diameter defines relay, proxy, and redirect agents, and requires that agents maintain transaction state, which is used for failover purposes.

- Relay Agents: accept requests and route messages to other Diameter nodes based on the Destination-Realm of messages. Relay agents can manipulate Diameter messages through inserting and removing routing information without modifying any non-routing AVPs;

- Proxy Agents: route messages using the realm routing table as relay agents do. However, they can modify messages to apply local policies, and add new AVPs to Diameter messages prior to routing;

- Redirect Agents: do not route messages, but simply return an answer with the information necessary for Diameter agents to communicate directly, without modifying messages.

As discussed in Sec. 3.2, the EAP three-party authentication model is considered as a standard method to accomplish access control through an AAA protocol in modern wireless networks. The Diameter NAS application specification (NASREQ), describing the interaction between NAS and Diameter server is standardised in IETF RFC 4005 [63]. Along with the Diameter EAP application (RFC 4072) [65], it supports the EAP authentication through a NAS with a Diameter server. The Diameter NAS application defines a number of commands and AVPs for authentication and authorisation. The NAS and Diameter server utilises these commands to conduct operations such as re-

authentication and RADIUS/Diameter protocol interactions. The mechanism of Diameter support for EAP is similar to what RADIUS does for EAP, by encapsulation of EAP messages.

Chapter 4

DYNAMIC NEIGHBOUR TRUST INFORMATION RETRIEVAL FOR GLOBAL ROAMING

4.1 Problem Definition

As discussed in Chapter 3, the maturity of new access technologies such as IEEE 802.11 (also known as Wi-Fi) and their widespread deployment would see a large number of small independent Wi-Fi operators co-existing with the cellular network operators. High speed Wi-Fi services can be provided with small radio coverage. Cellular technologies such as Universal Mobile Telecommunications System (UMTS) can provide wide radio coverage, but allow limited data rates. The growing demands for ubiquitous access are encouraging both independent Wi-Fi operators and big cellular operators to collaborate to enable seamless roaming across their heterogeneous wireless networks so as to maximise returns on investment.

Two interconnected networks belonging to different operators must satisfy three prerequisites to enable seamless roaming between each other. First, an appropriate interworking architecture must be in place to integrate heterogeneous network resources. An interworking architecture provides a "hard" platform for network integration [66]. This is underpinned by interworking signalling, mobility protocols and so forth. Second, a mobile's Quality of Service (QoS) during a handover needs to be optimised through some "soft" mechanisms, such as network selection, handover triggering and performance optimisation. QoS awareness ensures that handover operations can be kept transparent to upper layer applications. A number of papers [4, 13, 67, 68] have addressed the first two prerequisites discussed in this paragraph.

The third prerequisite on trust relationship requires that a mobile user's home network must have a trust relation with the mobile's visited network. This is usually available by enabling a roaming agreement between the two operators. A mobile user can roam to a visited network that has a roaming agreement with its home network due to security considerations. The co-existence and collaboration of a large number of operators inevitably result in complicated network trust relationships that are subject to changes in the roaming agreements over time. In order to access resources of global operators, a mobile user would have to deal with new trust-related challenges during global roaming.

Research issues related to trust relationship for seamless roaming are being addressed. *Shin et al.* [52] suggested that three kinds of trust relationship are essential for secured access in a roaming scenario: 1) The explicit Trust Relationship (TR) between a mobile user and its home network that is established through service subscription; 2) The explicit TR between a visited network and the mobile's home network, which can be set up by roaming agreements between the two network operators; 3) The implicit TR between the mobile user and its visited network, which is transitive and derived from the two explicit TRs. These relationships are shown in the Figure 4.1 below.

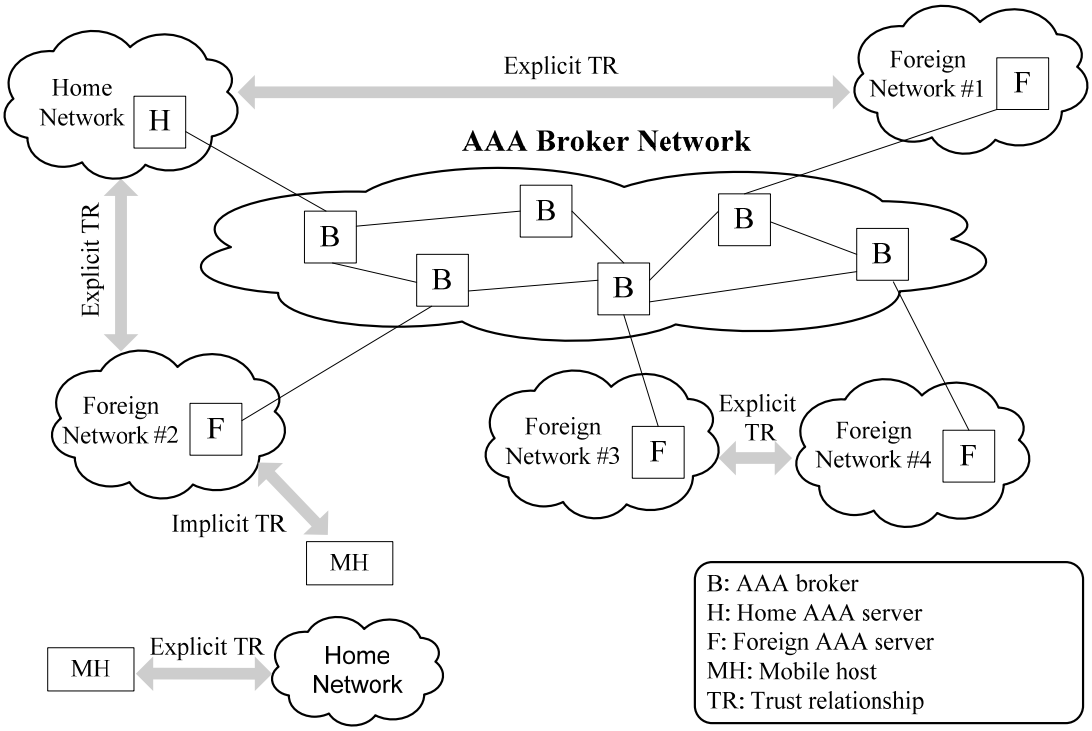


Figure 4.1 Trust relationships in a multi-operator environment

Figure 4.1 illustrates a multi-operator environment, in which multiple heterogeneous networks belonging to different network operators are interconnected through an AAA broker network. The AAA broker network is responsible for intermediating Authentication, Authorisation and Accounting (AAA) messages [53]. Current research has mostly focused on quick establishment of implicit TR so as to reduce handover delay using a variety of fast authentication mechanisms [69-71]. All these solutions are based on the assumption that an explicit TR between a mobile's home network and its selected network already exists, and does not need verification during a handover. This assumption is reasonable only when a small number of network operators are available for network selection, and their trust relationships information can be easily made available to a mobile user. The mobile user can hold a network identifier list to assist it with network selection, as defined by 3GPP's specification on system interworking [72]. However, such a solution lacks the scalability of supporting a large number of operators, and the flexibility of reacting to changes to network trust relationships. The current multi-operator support is largely provided as an add-on rather than an integral part of roaming solution to be effective. As a result, a mobile user's unawareness of the second explicit TR may lead to unnecessary handover attempts and affect handover performance in a multi-operator environment [73]. A mobile user requires additional network trust information in a handover to support global roaming.

In this chapter, a dynamic trust information retrieval scheme is proposed that overcomes the above mentioned problem. The proposed scheme called Dynamic Neighbour Trust Information Retrieval can provide a mobile user with ample network trust information about surrounding Points of Attachment (POA).

4.2 Network Trust Correlation

The basic idea behind the proposed scheme is as follows. The network trust relation between two networks can be validated when a mobile subscriber of a network successfully handover to another network. In a handover, the AAA server on the mobile's home network communicates with both the old POA (oPOA) and the new POA (nPOA) for AAA related services. The home AAA server can thus be used as an intermediate for exchanging information. Taking advantage of the home AAA server,

the information about the implicit TRs of the two neighbouring networks with a mobile user can be exchanged. This is applicable even when there is no direct trust association between the two neighbouring networks. By analysing handover history of a large number of mobile users, a network can dynamically obtain rich network trust information of its neighbours.

In a peer-to-peer direct roaming, one network can be directly interconnected with another network via an interworking gateway [20]. Alternatively, a roaming broker network [53] can be utilised for assisting roaming, which supports collaboration between two networks being enabled through roaming agreements. In this chapter, it is assumed that the roaming broker network is to be used for global roaming.

Every time a mobile user attaches to a POA, it has to establish an implicit TR with the visited POA. This can be done by getting the mobile user authenticated to its home network. A roaming broker network is used for intermediating AAA messages when the two networks are indirectly interconnected. Figure 4.1 illustrates how an implicit TR between a Mobile Host (MH) and a visited POA can be established using an efficient re-authentication method during a handover. The fast re-authentication has been specified in the Extensible Authentication Protocol Method for the 3rd Generation Authentication and Key Agreement (EAP-AKA) [74], and applied in 3GPP's specification [75]. The MH sends an EAP/Identity to the new POA (nPOA) to request attachment. The nPOA forwards this identity to the MH's home AAA server (HAAA). The HAAA will respond to a recognised identity by providing a set of security credentials as a challenge to the MH. The HAAA and MH exchange their security credentials and verify each other's identity in a set of round trips. During this process, better known as mutual authentication, the visited network (e.g. nPOA) plays an important role as shown in Figure 4.1. In summary, a mutual relation between a visited network and a mobile's home network must exist if the mobile wants roam to the visited network.

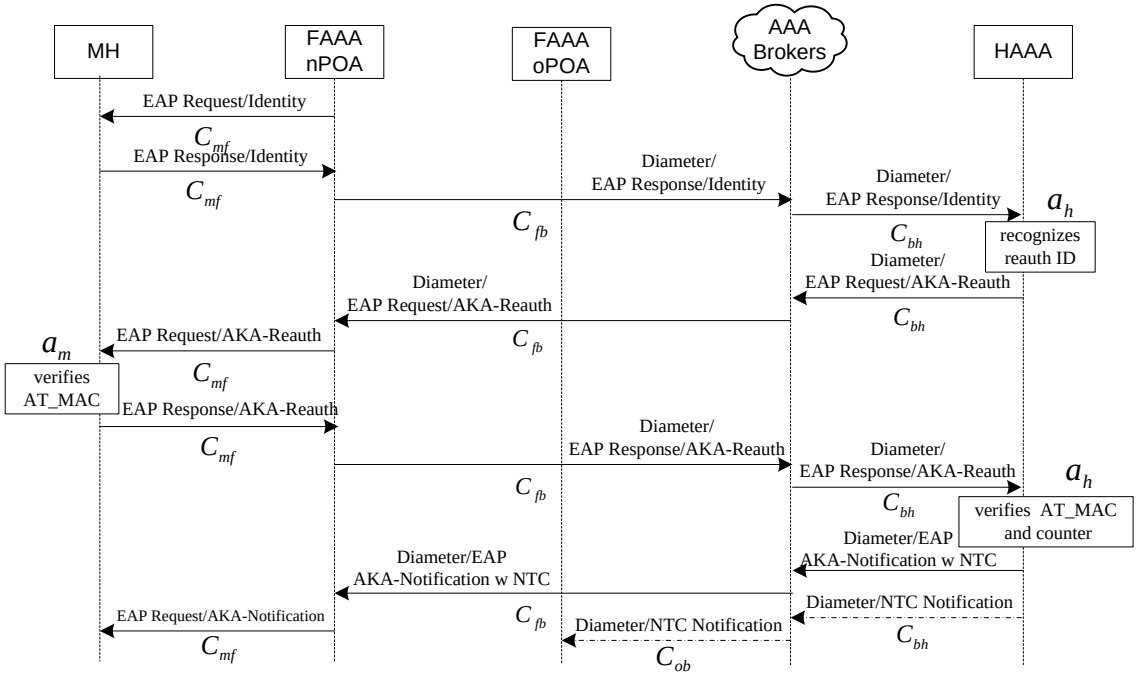


Figure 4.2 EAP-AKA re-authentication in a handover

To abstract such a mutual relation between two networks, a new term called *Network Trust Correlation* (NTC) is introduced. The NTC refers to the explicit TR between a mobile's home network and its visited network, which is required for roaming. From a mobile user's perspective, one network appears to be a service domain of another network if their NTC is present. The NTC reflects network trust relationship and can be built through a roaming agreement between the two networks. Therefore, the NTC is subject to changes of roaming agreements to some extent.

The re-authentication process of Figure 4.2 demonstrates that the absence of the NTC between the target network and the mobile's home network would be a preliminary indicator for the handover failure in a roaming scenario. When a mobile user is in roaming, the required NTC may have an impact on handover performance. Therefore, it becomes necessary to have a quantitative NTC model for studying the NTC's impact on roaming. In its simplest form, the NTC can be modelled as an on/off switch, which indicates the existence of an explicit TR. In this section, a more intelligent approach is presented. The proposed NTC model is based on the roaming broker network as shown in Figure 4.3. The third-party entities such as AAA brokers involved in intermediating

AAA services are considered. It is assumed that AAA services intermediated by too many third parties may be less reliable. An applicable approach of modelling the NTC is to have the NTC correlated to the number of Trust Association Hops (TAH) on the AAA path between two networks. For example, the AAA path from the visited network to the MH's home network covers two TAHs as illustrated in Figure 4.3. The TAH count would be the right parameter to show how a network is trust- associated with another.

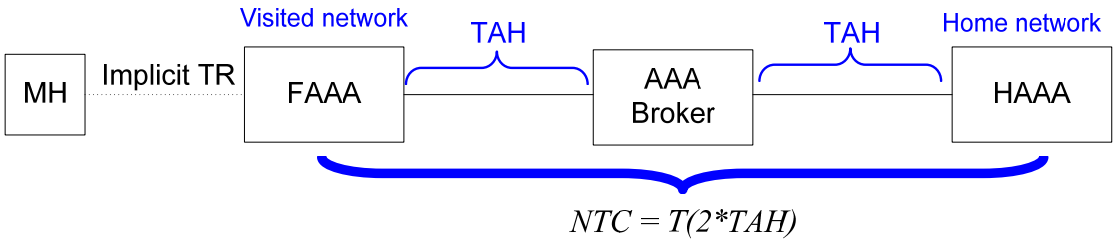


Figure 4.3 A network trust correlation model

Therefore, the NTC between any pair of networks can be quantified as:

$$NTC = h(i) \qquad \text{(Equation 4.1)}$$

where h means the number of trust association hops and i represents the distance between two networks.

4.3 Trust Information Retrieval and Distribution

Using the trust association hop based NTC model, the NTC between two networks can be promptly quantified when the authentication messages are exchanged. Because authentication is required in every handover attempt, the proposed scheme makes use of handover authentication process to obtain the NTC data between a pair of networks.

Basically, the NTC data retrieval can be performed every time a mobile user attaches to a visited network in a handover. Because the authentication request passes through the visited network, e.g. the oPOA as shown in Figure 4.4, and would be processed at the HAAA, both the visited network and the HAAA have the knowledge of their NTC data. The NTC data retrieval is conducted during the handover. The distribution of a

network's NTC data to its neighbouring network is initiated by the HAAA after a mobile user hands over from one network to another. Figure 4.4 illustrates an example of the NTC data retrieval and distribution process. The MH attaches to the oPOA, and gets the NTC data between the oPOA and the HAAA retrieved in *Step 1*. The HAAA retained the cached NTC data with the oPOA for future use. Then, the MH moves into the radio coverage of the nPOA, and triggers the handover (*Step 2*). During the handover, the authentication messages are exchanged between the MH and the HAAA via the nPOA. Thus, the HAAA retrieves the NTC data with the nPOA (*Step 3*). With the NTC data for both the oPOA and the nPOA, the HAAA initiates the distribution of the NTC data to the corresponding party (*Step 4* and *5*). More specifically, the NTC data for the oPOA would get distributed to the nPOA while the NTC data for the nPOA would be released to the oPOA. As a result, following each successful handover, an exchange of the NTC data can be done between arbitrary pair of adjacent networks.

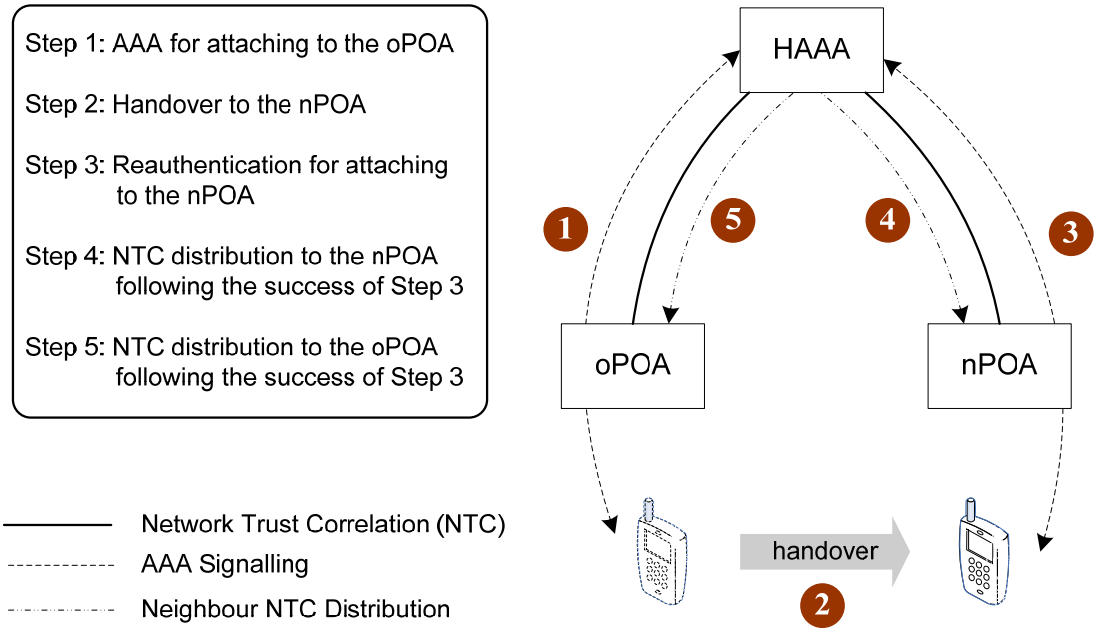


Figure 4.4 The proposed neighbour NTC exchange process

As a MH moves across several networks, a number of the NTC data exchanges would be performed. In Figure 4.5, the cell is used to abstract a network that is adjacent to six neighbouring networks. The trajectory of the MH1 would sequentially have the NTC exchanges $\{E_{AB}, E_{BC}, E_{CD}, E_{DE}\}$ to be done. In addition, another mobile user, MH2

moving in a different trajectory at the same time would contribute to the NTC exchanges $\{E_{FE}, E_{EG}, E_{GD}, E_{DH}\}$ as shown in Figure 4.5. In a confined area, as a large number of mobile users move across network boundaries, each network would have a great chance of learning the NTC data about its neighbouring networks. In this chapter, a network's neighbour NTC pattern is regarded as being established when it has full knowledge of every neighbouring network's NTC data. For example, network D needs two more NTC exchanges $\{E_{BD}, E_{DI}\}$ to obtain its neighbour NTC pattern.

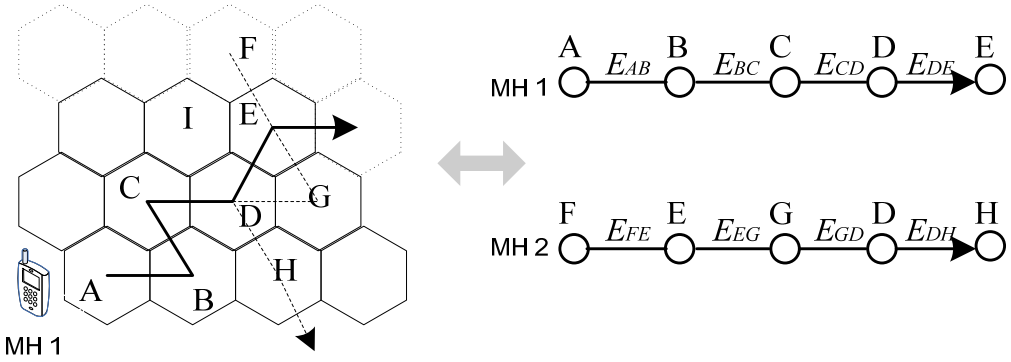


Figure 4.5 Mobile hosts' trajectory and the NTC data exchanges

Once a network has established its neighbour NTC pattern, it can provide such trust information to the attached mobile user to assist its network selection. Instead of relying on statically cached operator information in mobile terminal, trust information on demand makes sure that a roaming mobile user can deal with any operator irrespective of the complexity and versatility of network trust relationship.

4.4 Implementations

The NTC exchange between a network and its adjacent networks relies on the handover of mobile users between the networks. In a handover, a mobile user's network selection is influenced by a variety of factors related to its visited networks, such as placement of POA, signal strength, and radio propagation. These factors are often beyond the mobile's home operator's control. However, the home operator can still decide when and where the NTC distribution should occur, since the NTC distribution is always initiated by the AAA server.

The mobile user that is involved in the NTC retrieval and distribution may work in two different operation modes: Active Operation Mode (AOM) or Power Save Mode (PSM). The former refers to as the state of a mobile terminal that carries an active connection. The latter denotes the state, in which a mobile terminal's transmit and receive (RxTx) activities are reduced to save energy consumption. Accordingly, two NTC retrieving implementation methods corresponding to the two different operation modes were proposed. The proposed schemes are expected to be applied for different application scenarios rather than a contender to each other.

4.4.1 Active Operation Mode NTC

Mobile users working in active operation mode can get involved in the NTC retrieval and distribution process when performing handover. In this thesis, this process is referred to as the Active Operation Mode NTC (AOMN). Network operator can define its own policies to specify how its NTC data with a certain network should be updated as its subscribers roam to that network. In this case, the network operator reactively triggers the NTC process when its AAA servers are processing handover requests.

Apart from initiating the NTC process reactively along with a handover, the network operator can select and request its subscribers to perform handover for triggering the NTC process actively. For example, a network operator may instruct one of its subscribers carrying low priority services, e.g. web browsing, to switch to another network so as to collect and distribute the NTC data. The AOMN process can be called either when no NTC data for an area is available or the NTC data for that area has expired. Instead of selecting a network based on the rule of "always best connected" [34], the mobile users being selected for making the AOMN would sacrifice their QoS to discover the neighbour's trust pattern.

The AOMN process is quite similar to a standard handover except having two additional NTC distribution steps as shown in Figure 4.4. The signalling cost of the AOMN consists of two parts: 1) the signalling cost related to a mobile's standard handover, which results in the NTC data retrieved. 2) the signalling cost related to the NTC distribution following a handover. The parameters to be used to represent each procedure are shown in Figure 4.2.

C_{mf}: the cost of AAA signalling incurred between the Mobile Host (MH) and the Foreign AAA (FAAA);

C_{nb}: the cost of AAA signalling incurred between the nPOA FAAA and its associated AAA broker entry;

C_{bh}: the cost of AAA signalling incurred between the Home AAA (HAAA) and its associated AAA broker entry;

C_{ob}: the cost of AAA signalling incurred between the oPOA FAAA and its associated AAA broker entry;

μ: the average cost of signalling incurred on each AAA hop across the AAA broker network;

a_m: the cost of the AAA related processing at the MH's terminal;

a_h: the cost of the AAA related processing at the HAAA;

According to the re-authentication procedure of Figure 4.2, the handover part signalling cost C_H of the NTC retrieval can be represented as:

$$C_H = 5 \cdot C_{mf} + 4 \cdot C_{nb} + 4 \cdot C_{bh} + a_m + 2 \cdot a_h + 4\mu \cdot h_{nb}(i) \qquad \text{(Equation 4.2)}$$

in which $h_{nb}(i)$ represents the trust association hops that the AAA message originated from the nPOA traverses on the AAA broker network in handover. It is assumes that the signalling overheads for the AAA message's transmission between the nPOA FAAA and its associated broker, and its transmission between the HAAA and its associated broker, are the same as what is required for transmitting AAA message between two AAA brokers. Thus, when $C_{fb} = C_{bh} = \mu$, the handover part signalling cost C_H can be simplified as:

$$C_H = 5 \cdot C_{mf} + 4\mu \cdot [2 + h_{nb}(i)] + a_m + 2a_h \qquad \text{(Equation 4.3)}$$

The NTC distribution process begins once the identity of the MH is successfully verified by the HAAA. Because the NTC distribution to the nPOA (C_{hn}) can be included in the EAP-AKA notification sent by the HAAA, it would avoid any additional costs. Thus, the NTC distribution part signalling cost (C_N) is equal to the transmission cost (C_{ho}) of the one-way NTC data to the oPOA. Thus, we can get:

$$C_N = C_{bh} + C_{ob} + \mu \cdot h_{ob}(j) = \mu \cdot [2 + h_{ob}(j)] \qquad \text{(Equation 4.4)}$$

where $h_{ob}(j)$ denotes the trust association hops that an AAA message has to pass through from the HAAA to the oPOA FAAA.

From Equation 4.3 and 4.4, total signalling cost related to the NTC retrieval and distribution process is

$$C = C_H + C_N = 5C_{mf} + \mu \cdot [10 + 4h_{nb}(i) + h_{ob}(j)] + a_m + 2a_h \qquad \text{(Equation 4.5)}$$

4.4.2 Power Save Mode NTC

With the appropriate operator policies in place, the active operation mode NTC, AOMN process can be implemented to trigger the NTC exchange in a timely manner. This makes sure that a network can obtain an accurate picture of its neighbouring networks' network trust information. However, the disadvantages of the AOMN are that it requires the involvement of a mobile user carrying active user sessions which may result in compromised QoS.

For this reason, an alternative method named as Power Save Mode NTC (PSMN) process was proposed. The PSMN makes use a group of mobile users in power save mode to trigger the NTC exchange between neighbouring networks. Instead of relying on actual handover, Location Update (LU) and Paging (PA) procedures of mobile terminals in power save mode are utilised. Location update LU has been proposed to keep track of a mobile user's locations by registering to the network about the current location of the mobile user. The accuracy of location data obtained during the LU is determined by the defined update interval [76]. Therefore, theoretically, if a fine-grained LU scheme is applied, a mobile user in power save mode is able to perform the

LU through two adjacent networks sequentially. Therefore, using the NTC retrieval process shown in Figure 4.4, a network would have a great chance of being able to exchange its NTC data with a neighbouring network.

Paging PA is the process of determining the exact location of a mobile user by polling individual locations of an area. IP paging has recently been demonstrated in [77] as a practical method for mobile terminals in IEEE 802.11 power save mode. With a paging controller that exerts full paging control, data packets being addressed to the mobile user arrive at the paging controller, and get buffered. The data packets are distributed to the access routers associated with a paging area at the selected intervals, at which the mobile user activates its terminal and listens to paging related signalling. Along with the LU, the PA provides an effective means of pushing the needed network trust information to mobile users.

The proposed PSMN procedure is shown in Figure 4.6. The NTC retrieval and distribution can be initiated by a network operator, which the network trust pattern of an area is incomplete or has expired. The HAAA may page its mobile user in power save mode to start the PSMN. On reception of the PSMN request, the mobile user awakes its network interface card, and issues a location update request to the HAAA at the cost of C_{LU}. Then, the mobile user registers to the paging controller at the cost of C_{PSM} for receiving paging related signalling when entering dormant state. The mobile user keeps in dormant state for a period of NTC Update Interval (NUI) before taking next location update. The NTC update is referred to as the location update that is performed by a mobile user in power save mode for retrieving NTC data. The NUI is represented as:

$$T_{NUI} = k \cdot \rho \cdot T_{BI} \qquad \text{(Equation 4.6)}$$

where T_{BI} means Beacon Interval (BI) of local point of attachment. T_{NUI} is a multiple (k) of the mobile's RxTx activity period $\rho \cdot T_{BI}$. Note that the NUI is a parameter pre-determined by the operator policy, and is outside the scope of this thesis.

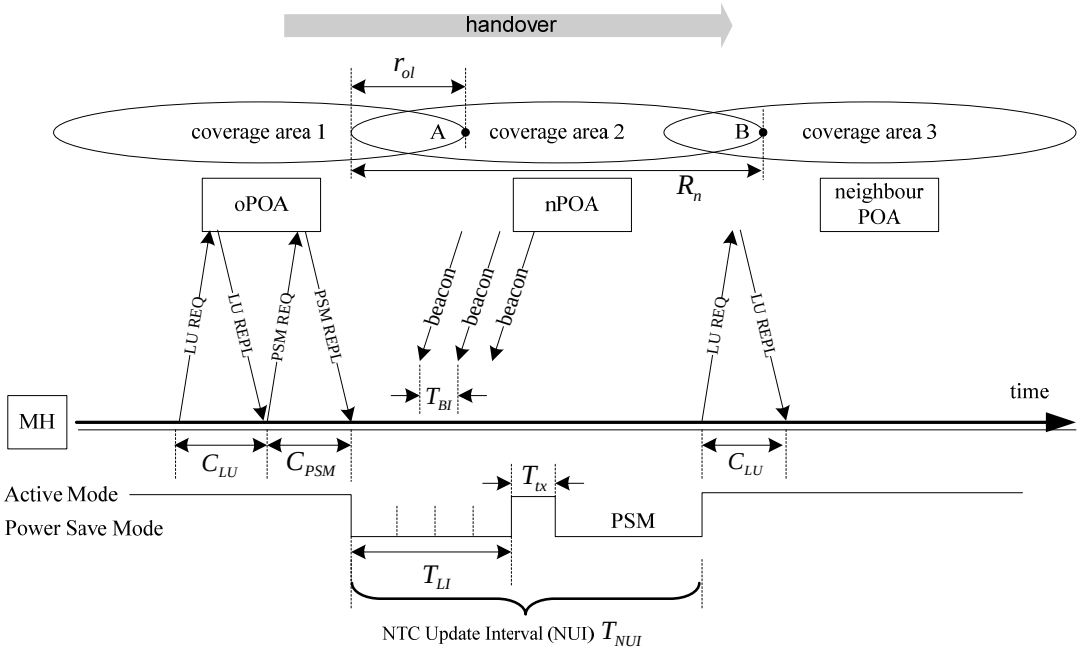


Figure 4.6 The power save mode NTC procedure

According to the PSMN procedure shown in Figure 4.6, theoretically, the NUI interval should be as short as possible to maintain a high degree of accuracy of neighbouring relationship between two networks. However, practically, the frequency of the related LU has to be balanced against the signalling overheads and the power consumption of a mobile terminal.

The maximum NUI is explained as follows. In the worst case, a mobile user has to renew its oPOA location as its residence time at the oPOA has just elapsed at time A as shown in Figure 4.6. To get an accurate picture of the neighbouring relationship between the oPOA and nPOA, the mobile user has to perform another location update through the nPOA before time B. It is assumed that the radio coverage of the two POAs is partially overlapped. In a handover, overlapped radio coverage may have the effect of probable decrease of a mobile user's residence time on a network. The variation in residence time caused by radio overlap effect is denoted as t_{ol}, which fluctuates according to an exponential probability distribution with the mean of u.

$$T_{nui} <= R_n - r_{ol} - C_{LU} - C_{PSM} \qquad \text{(Equation 4.7)}$$

where R_n represents the maximum residence time of the mobile user in the nPOA without handover. The location update cost and the power save mode registration cost largely depend on the signalling process taking place in networks. The NTC update interval may change over different access networks.

Taking different approaches, the PSMN and AOMN methods are intended to complement each other in facilitating network trust information distribution among neighbouring networks. The AOMN that relies on actual handover event can produce the most accurate information about neighbouring relationship. In the AOMN, the mobile users sacrifice their QoS and have the attached network obtain its neighbour network trust pattern in return. In contrast, the PSMN derives neighbouring relationship from analysing location update records. The mobile users selected for participating in the PSMN sacrifice their power consumption instead.

4.5 Performance Evaluation

A random walk model [78] as shown in Figure 4.7 is introduced to simulate a multi-operator environment. It is assumed that the radio coverage of every network (denoted as a *cell* in the model) is hexagonal shaped, and has six neighbouring cells. A mobile user has equal probability of moving to any of the six neighbouring networks. The cell residence time follows a Gamma distribution. The rings as marked in Figure 4.7 are used to group cells. Each ring of cells is surrounded by the neighbouring outer ring of cells, and is also adjacent to an inner ring of cells. The outmost ring of cells is referred to as boundary ring or Ring *n*, and assumed to be in an absorbing state. A mobile user entering the boundary ring will remain in that ring at all times. This can be explained by the nomadic behaviour of a mobile user. Moreover, it is assumed that each network belongs to an independent operator to make the handover task more challenging. Thus, whenever a mobile user leaves a cell, a NTC exchange between two adjacent cells along with a handover can be expected. The objective of the proposed NTC process is to have every network to obtain its neighbour network trust pattern. According to the procedure of Figure 4.4, this requires that a network has to exchange the NTC data with all of its 6 neighbouring networks at least once.

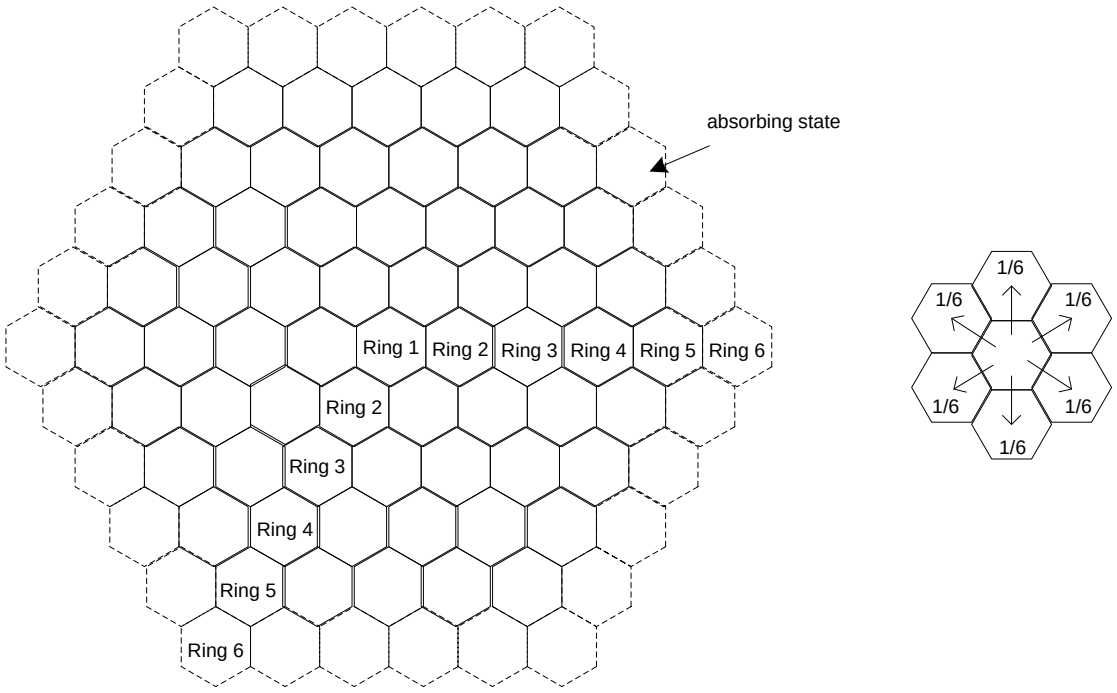


Figure 4.7 The hexagonal random walk model

Using the method proposed in [79], the expectation of the transition probabilities for the mobile user moving within Ring *n* can be derived. The state transition follows a Markov chain as shown in Figure 4.8.

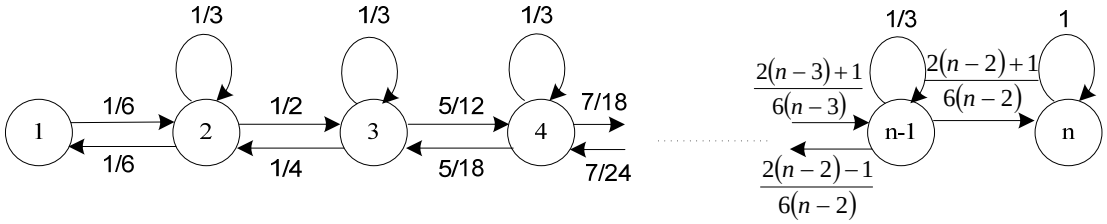


Figure 4.8 Transition probabilities for the random walk model

Its transition probability matrix is denoted as *P*. Let $S_{i,n}$ represent the number of cell crossings the movement takes for its first entrance into boundary ring given that $X_0 = i$. Random variable $S_{i,n}$ is known as the first passage time from *i* to *n* [80]. $q_{i,n}^{(m)}$ is used to represent the probability mass function for $S_{i,n}$. Thus, we get:

$$q_{i,n}^{(m)} = P\{S_{i,n} = m\} = P\{X_m = n, X_{m-1} \neq n, \ldots, X_1 \neq n \mid X_0 = i\} \qquad n=1, 2, \ldots$$

$q_{i,n}^{(m)}$ can be derived recursively as follows:

$$q_{i,n}^{(1)} = p_{i,n}$$

$$q_{i,n}^{(m)} = \sum_{k=0,k\neq n}^{\infty} p_{i,k} q_{k,n}^{(m-1)} = \sum_{k=0}^{\infty} p_{i,k} q_{k,n}^{(m-1)} - p_{i,n} q_{n,n}^{(m-1)} \qquad n=2, 3, \ldots$$

The matrix form of the equation can be represented as:

$$Q^{(m)} = PQ^{(m-1)} - PQ_d^{(m-1)} \qquad \text{n=2, 3, …} \qquad \text{(Equation 4.8)}$$

where $Q^{(m)} = \{q_{i,n}^{(m)}\}$ and Q_d denote a diagonal matrix formed by the diagonal elements of Q. We can also get $Q^{(1)} = P$. The cumulative first-passage-time probability is denoted as:

$$q_{i,n} = \sum_{m=1}^{\infty} q_{i,n}^{(m)} \qquad \text{(Equation 4.9)}$$

Thus, we can get the mean number of cell crossings of Ring *n*:

$$E[S_{i,n}] = \sum_{m=1}^{\infty} m \cdot q_{i,n}^{(m)} \qquad \text{(Equation 4.10)}$$

where $E[S_{i,n}]$ is the expectation of the number of cell crossings the mobile takes for its first entrance into boundary cells.

The cumulative fist-passage-time probabilities are employed to study the movement of a mobile user in a region. It is observed that the mobile user will eventually be absorbed at the boundary ring (with $q_{i,n} = 1$) irrespective of its initial position. The mathematical analysis of the ring sojourn time is in accordance with the intuitive perception that the mobile users closer to the boundary ring will move out of the region first. For the region of 6 rings, on average, a mobile user will end up being absorbed at the boundary ring after 50 cell crossings as shown in Figure 4.9.

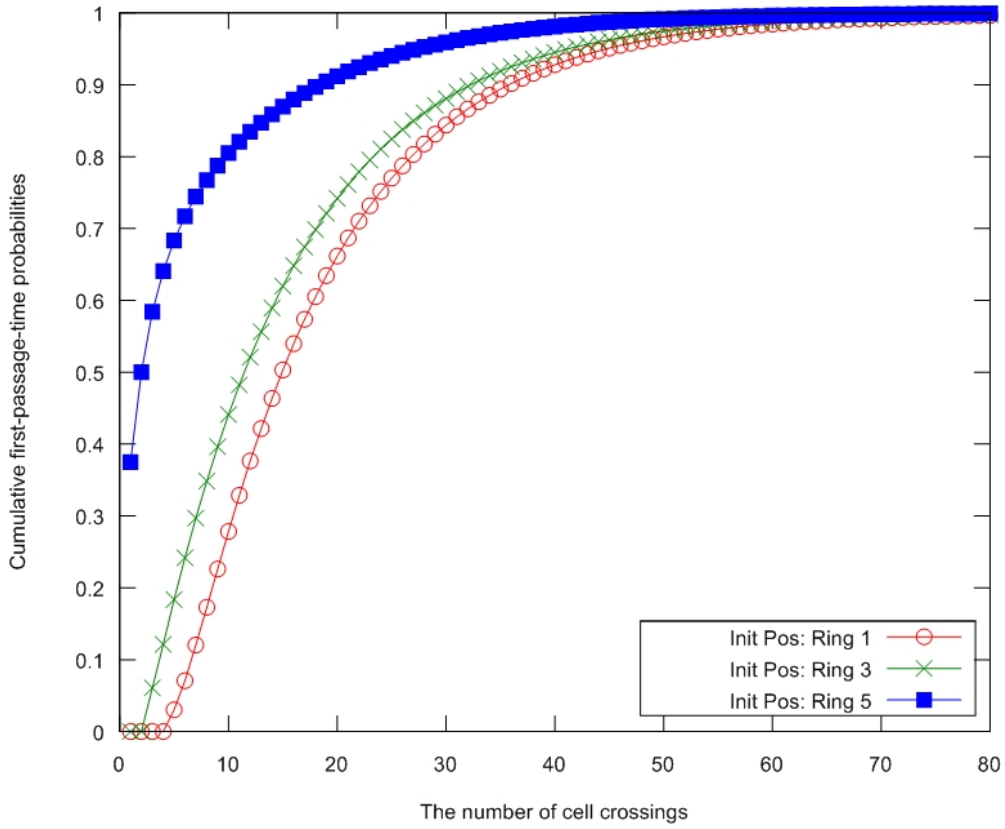


Figure 4.9 The cumulative first-passage-time probabilities of entering into boundaries [n=6]

With the transition matrix shown in Figure 4.8, Equation 4.10 can be used to calculate the mean number of cell crossings incurred by single mobile user in the region. As discussed early, the NTC exchange relies on a mobile user's cell crossing. Therefore, the number of cell crossings shows how many NTC exchanges may have been performed. In Figure 4.10, the mean number of cell crossings incurred (before being absorbed at the boundary) is compared with the total number of neighbour edges, which is the minimum number of cell crossings required for establishing the complete network trust pattern in the region and is determined by the size of rings. As the size of rings increases, the number of neighbour edges grows much faster than the number of cell crossings a mobile user can conduct. Larger is the region (measured by n), less contribution can a mobile user make to the building of the network trust pattern. From Figure 4.9 and Figure 4.10, it is found that the contribution of single mobile user to the NTC distribution is quite limited due to its mobility pattern.

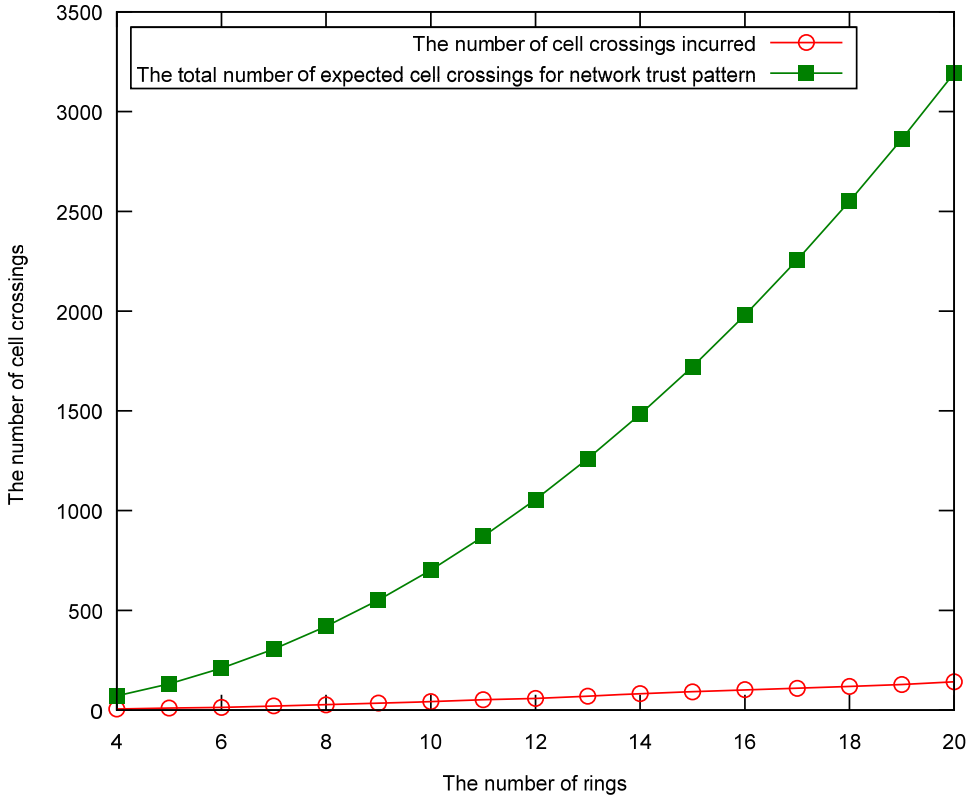


Figure 4.10 The mean number of possible cell crossings

In this study, it is assumed that a mobile user's cell residence time follows the Gamma distribution with the Mean Cell Residence Time (MCRT) determined by $\alpha \cdot \theta$. Mobile users are uniformly distributed among the networks in the region when the simulation is initiated. Mobile users are independent to each other in regards to their mobility patterns.

4.6 Simulation Results

The simulation has been run to evaluate how the proposed NTC distribution scheme performs in a multi-operator environment. 500 runs were conducted for each simulated scenario for averaging the results. In the simulation, two aspects were focused: 1) how a mobile user can contribute to the NTC distribution; 2) how the anticipated NTC distribution would have an impact on networks.

A new parameter, named as the NTC Exchange Finish Ratio (EFR) is introduced, which represents the completion status of the NTC distribution among the neighbouring networks in a region. The NTC EFR is defined as the ratio of the total number of the

NTC exchanges that have been performed between the adjacent networks. As proved earlier, a group of mobile users are required to participate in the NTC distribution to obtain complete network trust pattern. For practical implementation, it is assumed that a NTC EFR of 95% is a good indicator that the network trust pattern in a region has been effectively formed. In this simulation, the size of rings as shown in Figure 4.7 was varied to get a comparative result. The experimental results show that the minimum number of mobile users required for establishing network trust pattern for the region in size: 4, 8, 12 (rings) is 132, 169 and 196 respectively. Figure 4.11 shows that the NTC EFR tends to stabilise at a certain value no matter how long the observation on the NTC exchange has lasted. The stabilisation of the NTC EFR under a fixed number of mobile users can be explained by the rationale behinds Figure 4.9 and Figure 4.10 that the contribution of single mobile user to the NTC distribution is very limited. By comparing the NTC EFRs under the different numbers (2, 5, 10) of mobile users, it is found that an effective way of increasing the NTC EFR is to get more mobile users participate in the NTC distribution. More mobile users are involved, higher the NTC EFR is.

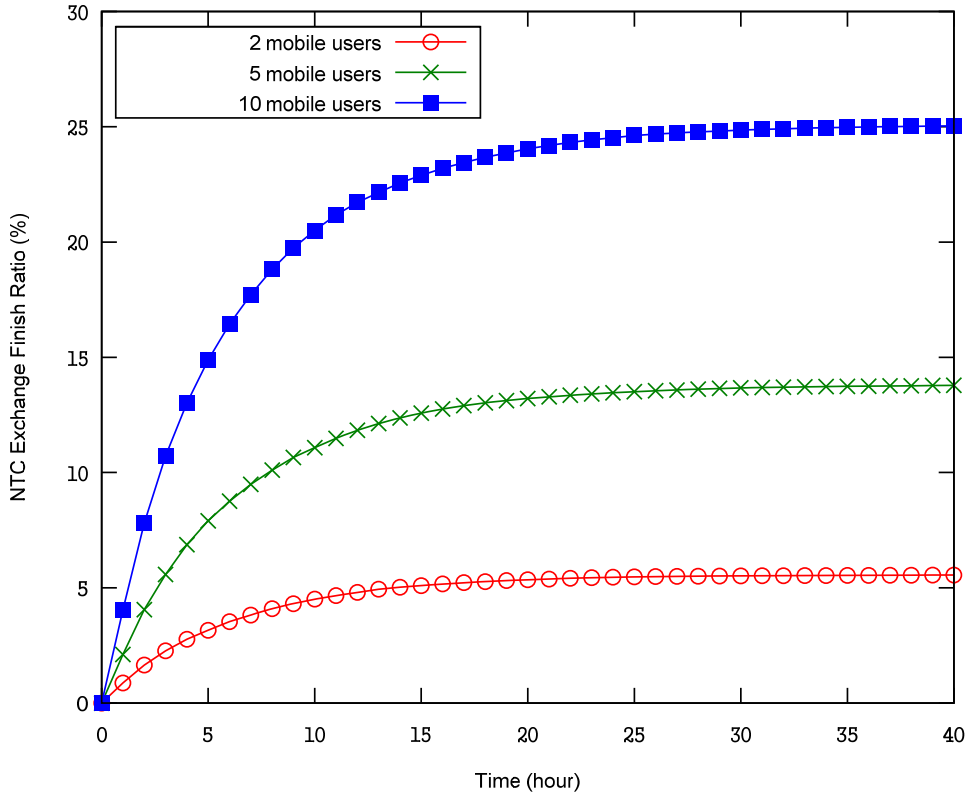


Figure 4.11 NTC exchange finish ratio vs. time [Size of rings=6, MCRT=30mins]

The simulation moves on to further the evaluation of the exact number of mobile users required for establishing complete network trust pattern. As aforementioned, a NTC EFR of 95% is assumed to be sufficient to have the NTC data exchanged between most of the neighbouring networks. The simulation results show that the minimum number of mobile users needed for the region in size: 4, 8, 12 (rings) is 132, 169 and 196 respectively. This proves that a complete network trust pattern in a region can be established using the proposed NTC distribution process. In a small region with 4 rings of cells, no less than 132 mobile users are required according to the assumed mobility pattern. The size of rings has an impact on the number of mobile users required for the NTC distribution. Generally, larger is a region, more mobile users the NTC distribution demands. The model of Figure 4.7 indicates that larger a region is, more independent networks are included.

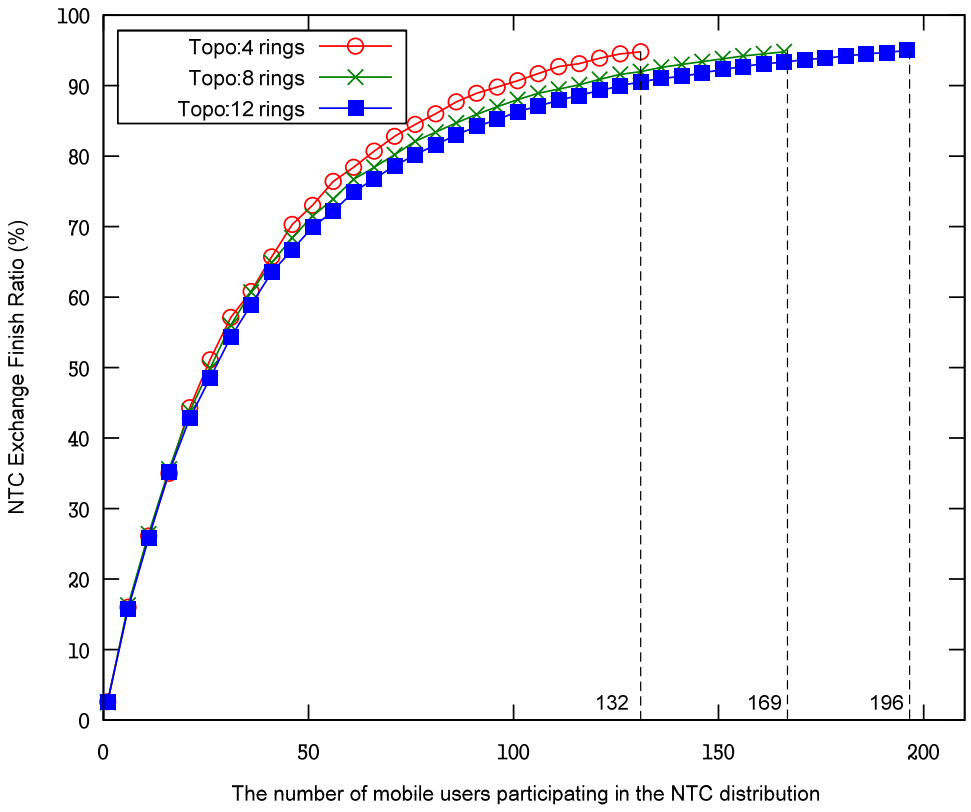


Figure 4.12 Number of mobile users vs. NTC exchange finish ratio

Figure 4.12 demonstrates the formation of network trust pattern at the macro level. For the benefits of real deployment, the issue was studied at the micro level. This is done by examining the number of mobile users required per network/cell. In the experiment, the

size of the simulated region was changed from 4 to 18 rings. Figure 4.13 shows the simulation results. Basically, when a region grows in size, less mobile users would be needed on a cell basis. For example, a region consisting of 6 rings needs a minimum of 2.5 mobile users per cell to establish complete network trust pattern. In contrast, when a region includes more than 12 rings of cells, a mobile user per cell would be sufficient. This demonstrates that the overheads of making use of mobile users for building network trust information can be easily controlled when the networks are expanded in scale.

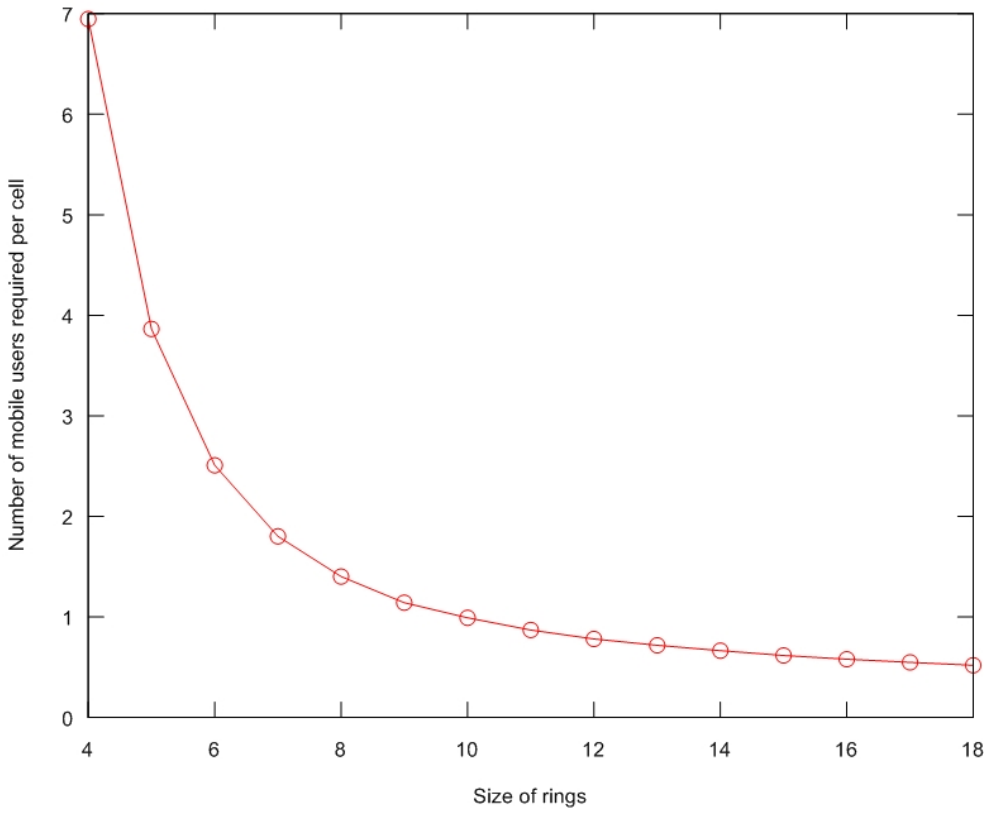


Figure 4.13 Size of rings vs. Number of mobile users required per cell [MCRT=30mins]

Since the proposed NTC distribution needs additional transmissions of data packets from the HAAA to access networks, its signalling costs are a great concern. In the simulation, the signalling cost of the handover plus the NTC distribution (NTC HO) was compared with that of the standard handover process (Std HO). It is assumed that a standard handover involves the re-authentication procedure as described in [74]. The NTC distribution is performed during every handover. Alternatively, the intelligence can be introduced: the NTC distribution is executed when no valid NTC data is

available to avoid repetition. This intelligent method is referred to as the NTC distribution with enhancement (NTC with EN). Figure 4.14 shows that the NTC distribution along with handover generates 15.2% more signalling overhead than the standard handover. However, with an enhancement of checking-before-distributing NTC data, the signalling cost of the NTC distribution can be much reduced. Comparing with the standard handover, the enhanced NTC distribution incurs a 3.7% additional signalling overhead. The low-cost approach of the proposed scheme makes sure that its large scale deployment would not generate excessive signalling flooding networks.

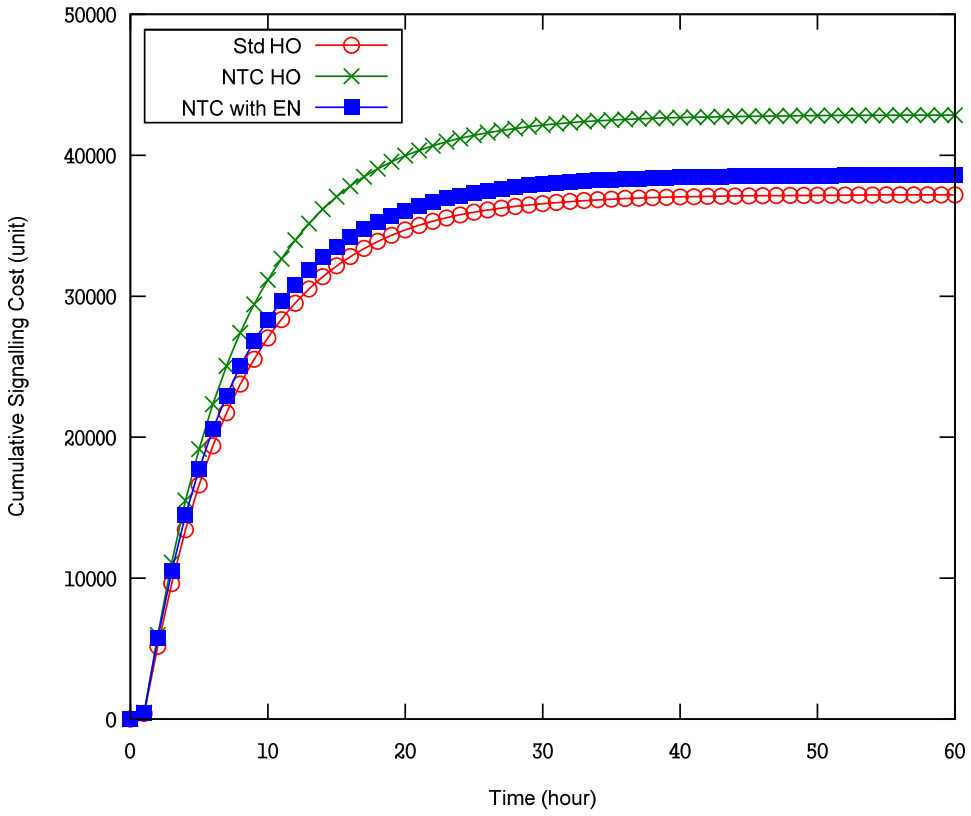


Figure 4.14 Cumulative signalling cost vs. time [MCRT=30mins]

Another concern in a real deployment is how long the proposed scheme would take to build up a complete network trust pattern, particularly when new access networks are added. For such a purpose, a new parameter called NTC Pattern Construction Time (PCT) is introduced. The NTC PCT is defined as the time taken from retrieving the first NTC data to having the NTC data of every network in the region retrieved and distributed between the networks in the region. The number of mobile users per cell was varied from 2 to 5 in the region of 6 rings, and from 1 to 5 in the region of 10 rings in

the experiment. These values were particularly selected because of the impact of the size of a region on the number of mobile users required for the NTC distribution as shown in Figure 4.13. Figure 4.14 shows that the NTC PCT for the region is cut down when the number of mobile users increases. The effect of adding more mobile users for facilitating the NTC distribution is the most prominent when the number of mobile users per cell is above the minimum number needed. It is observed the NTC PCT when the mobile users were having different MCRTs. Generally, mobile users with high MCRT require less time to get the NTC data distributed. This means that the mobility pattern of a mobile user plays a role in determining the NTC PCT. However, the role of the MCRT in determining the NTC PCT tends to be less important when the number of mobile users per cell increases to a certain extent, e.g. 2 for the region of 10 rings, and 3.2 for the region of 6 rings as shown in Figure 4.15.

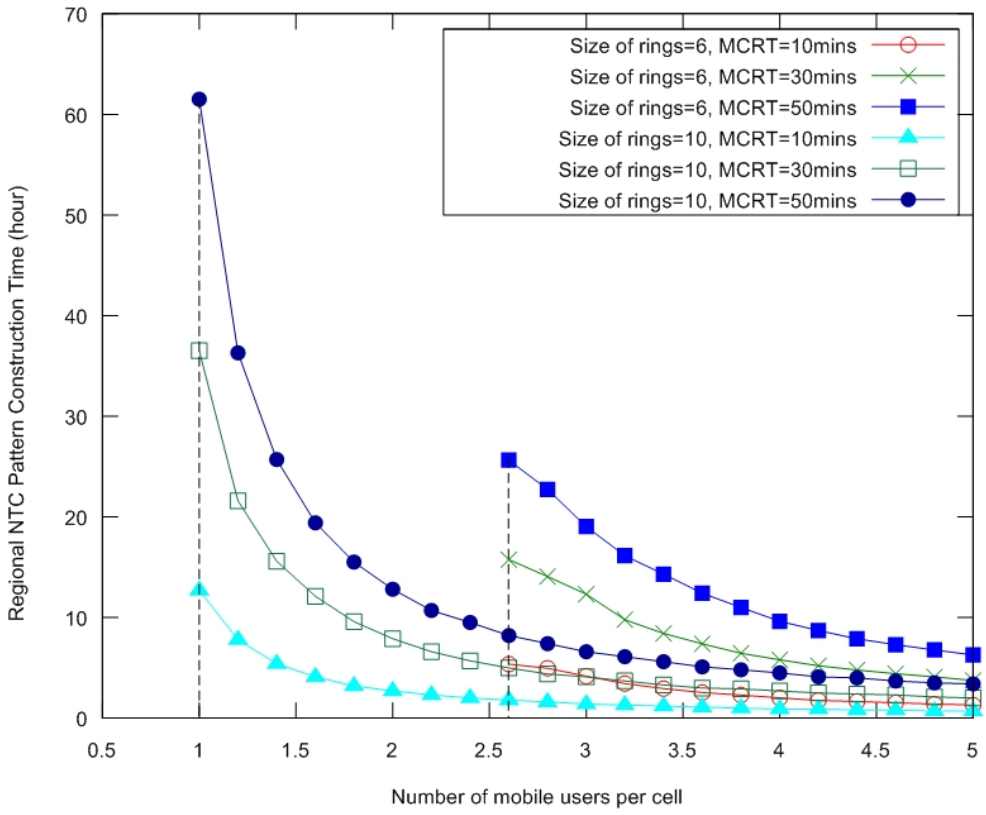


Figure 4.15 Number of mobile users per cell vs. regional NTC pattern construction time

Next, the experiment moves on to evaluate the accuracy of the PSMN in the NTC distribution. As discussed early, mobile users in a dormant state can be chosen to participate in the NTC distribution. However, under the PSMN, Equation 4.7 has to be

satisfied to make sure that the neighbouring relationship is correctly obtained. That means that the NTC update interval is restrained by a threshold, which is jointly determined by a user's mobility pattern and the extent of the radio overlap. The skip of a particular cell during the NTC update may result in non-adjacent cells being derived as being neighboured. The NTC exchange error rate represents the ratio of the NTC exchanges wrongly conducted to the total NTC exchanges expected for the topology. Figure 4.16 shows that the maximum allowable NUI varies as the MCRT changes. Generally, the NTC exchange error rate increases when the NTC update interval is widen. According to the simulation results, the maximum allowable NUI is 4 mins for MCRT=10mins, 14 mins for MCRT=30mins, and 32 mins for MCRT=50mins. The results suggest that the NUI should be adjusted according to the mobility pattern of the NTC participants in the PSMN.

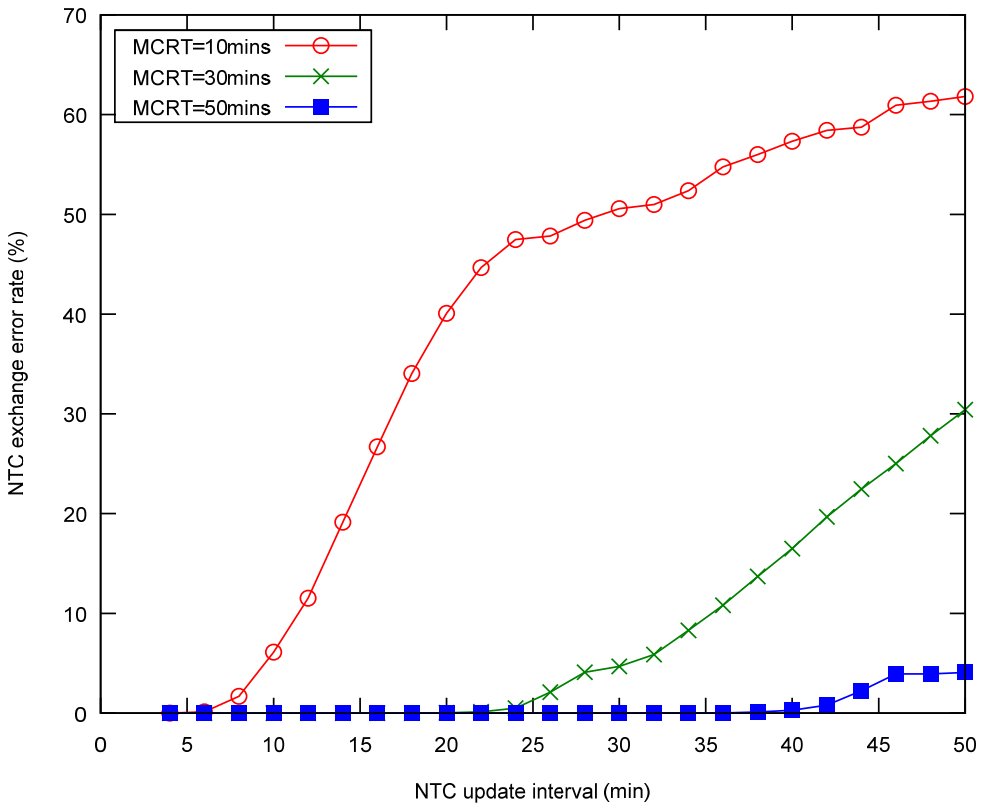


Figure 4.16 NTC update interval vs. NTC exchange error rate

4.7 Conclusion

In this chapter, a network trust information retrieval scheme for global seamless roaming is presented. To the best of the author' knowledge, it is the first of its kind that addresses the seamless roaming problem in a multi-operator environment. The proposed scheme makes use of a mobile user' mobility and handover process to exchange network trust information between neighbouring networks without any direct communication. In this way, a network can build up its neighbour network trust pattern. As a result, a mobile user would no longer need to carry network trust information for roaming with it all the time, e.g. resorting to a locally cached network identifier list. The network trust information related to neighbouring networks can be dynamically provided by the serving network.

To evaluate the performance of the proposed scheme, a series of simulations have been conducted. The simulation results show that the proposed trust information retrieval scheme can function in a cost-effective manner, generating an additional signalling overhead of 3.7% on networks compared with the standard handover signalling cost. Moreover, it was shown that one mobile user per network would be sufficient to establish a complete network trust pattern in most cases when the proposed scheme is widely deployed in networks. For real deployment, this makes sure that its implementation would not be a burden for networks in operation. It was found that the time needed for constructing neighbour trust pattern can be speeded up by increasing the number of mobile users involved.

Chapter 5

TRUST ASSISTED HANDOVER ALGORITHM FOR RELIABLE HANDOVER

5.1 Introduction

Homogeneous wireless networks generally employ a centralised handover approach [31, 81, 82], e.g. Network-Controlled Handover (NCHO), in which network-end components make handover decisions. However, in Next Generation (NG) Heterogeneous Wireless Networks, it is widely envisioned that the decentralised Mobile-Controlled Handover (MCHO) would be a better choice [31, 81].

In recent years, there have seen a number of papers [82-85] investigating how handover decisions should be made in heterogeneous wireless networks. It is commonly proposed [83-85] that multiple handover metrics, such as Received Signal Strength (RSS) [86], service type, and bandwidth, shall be considered in the handover decision process so as to be "always best connected" [34]. The notion of "always best connected" means that a mobile user is always connected through the best available device and access technology at all times [34]. At present, handover decision processes tend to adopt a cost-based approach, which employs multiple handover metrics for the sake of Quality of Service (QoS). The current approaches have been specifically designed to address the heterogeneity challenges of handover [13] arising from the interworking of multi-technology networks. However, in NG Wireless Networks, disparate network domains may belong to multiple network operators, and rely on their roaming agreements to collaborate between each other [87]. Roaming agreements are subject to changes with time, and this can subsequently affect trust relationships between the interconnected networks.

The state-of-the-art handover approaches discussed in Chapter 2 such as Dynamic Vertical Handover Algorithm with Network Elimination [83], Network Selection Scheme using Analytic Hierarchy Process (AHP) [84] and Policy Enabled Handover [88], [83, 84, 88] take into account the QoS capability of candidate networks. These handover approaches assume that any discovered network can be a candidate network for handover, and is available for providing services at any time. Unfortunately, this is seldom the case in a multi-operator environment, which has versatile network trust relationship. A network with good QoS may be inaccessible for a mobile user at all times due to roaming agreements implemented between operators. Unavailability of network trust information at mobile terminals may result in unnecessary handover attempts, and thus increase handover latency.

In order to resolve trust-related handover problems in a multi-operator environment, 3GPP has specified methods for manual and automatic network selection in its latest specification for 3GPP-WLAN interworking *TS 24.234* [72]. It is proposed that a user-controlled or operator-controlled network identifier list can be employed by a mobile user to assist its network selection during a handover. For example, the order of network selection could follow the order in the identifier list at the mobile terminal. However, the scheme of storing accessible networks in a static data file ignores the dynamics of network trust relationships, which may be a great challenge as stated in Chapter 4. Moreover, the versatile network trust relationships may be further complicated when a vast number of small and independent network operators collaborate in a global roaming environment.

For handover in a multi-operator environment, without consideration on network trust relationships, the current approaches [72] appears more likely as an add-on rather than an integral part of handover solution.

5.2 Related Work

Handover algorithms are employed to deal with two main things during a handover: *network selection* and *handover triggering*. Network selection determines the most appropriate network, to which a mobile user can be switched and get the best QoS in a handover. Handover triggering initiates the switching of a mobile user to the selected

network. During a handover, it is required to determine a handover triggering time, which indicates to the handover decision maker when the mobile's handover operation should be conducted. The handover triggering time is carefully calculated to avoid unnecessary back and forth handover between two networks having similar conditions.

In voice-oriented wireless networks such as GSM, handover is conducted for maintaining the user's telephony voice sessions when a Mobile Host (MH) moves across different points of network attachment. A variety of metrics such Received Signal Strength (RSS), Bit Error Rate (BER) and Signal to Interference Ratio (SIR) have been employed to decide the two key things: network selection and handover triggering time.

Traditionally, the RSS or received power measurements from surrounding points of network attachment are used as the primary metric to decide which network is selected and when handover is executed. When a MH is in motion, the RSS of its serving network P_{old} is constantly compared with a predefined RSS threshold P_{th}. When P_{old} is less than P_{th}, the ongoing call session is switched to the newly selected network assuming that it can provide stronger signal strength P_{new} for the MH. For example, a handover would be performed at time *t1* when the better signal can be received from the new network as shown in Figure 5.1. This RSS threshold-based approach can be simplified as:

$$P_{new} > P_{th} > P_{old} \qquad \text{(Equation 5.1)}$$

However, the RSS of a point of network attachment may fluctuate in a certain range with time due to its radio propagation. The decision based on the single RSS metric may lead to the so-called ping pong effect, which sees a MH conducts a number of handovers between two points of attachment over a short period of time. One way to eliminate the ping-pong effect is to persist with a point of attachment for as long as possible with the minimum degradation of quality of service. This can be done by introducing additional metrics such as hysteresis margin [82], dwell timer [89], traffic load [90] and so forth. For example, the RSS measurements can be employed jointly with a hysteresis margin *H* as shown in Figure 5.1. The handover is made at time *t2*

when the RSS P_{new} of the new point of attachment is larger than the P_{old} by a hysteresis margin H.

$$P_{new} > P_{old} + H \qquad \text{(Equation 5.2)}$$

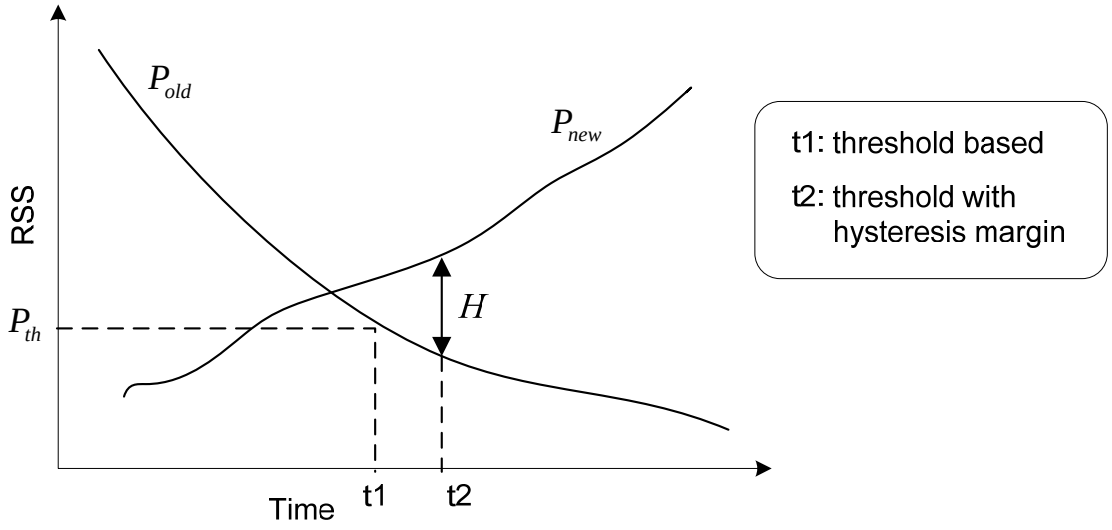


Figure 5.1 RSS-based handover decision algorithm

Therefore, in the scenario of Figure 5.1, the handover occurs at time *t1* if a single RSS metric is used. The handover triggering time is delayed to time *t2* when a RSS hysteresis margin is added. Other handover metrics can be used as well for more complex scenarios, such as Microcell/Macrocell overlay [1].

In packet-oriented wireless networks, handover decision has to deal with more things. A variety of services being carried on IP packets may have different requirements related to QoS. The quality of the wireless channel in a handover is no longer the only concern. A number of handover metrics such as bandwidth, service type, monetary cost and so forth would have to be considered in handover decision making for wireless data services.

The cost-based Multicriteria Handover Algorithms (MHOA) such as Dynamic Vertical Handover Algorithm [83] and Policy-Enabled Handover Algorithm [88] were proposed to handle multiple handover decision metrics that are required for optimising data services in heterogeneous wireless networks. The basic idea behind the MHOA is easy

to understand. The mobile user uses multiple handover metrics as input, and makes decision based on the comparison of the costs of a set of candidate networks using a cost function model. The weight for each metric is determined based on the contribution of that metric to network selection.

Basically, the MHOA can be generalised as follows:

$$C_n = \sum_i w_i \cdot N(m_i) \quad \text{where } (\sum w_i = 1) \qquad \text{(Equation 5.3)}$$

where C_n means the cost of the *n*th candidate network. m_i and w_i represent the *i*th handover metric and its weight in all handover decision metrics. The parameter m_i is normalised to $N(m_i)$ using a normalisation function, which ensures the sum of different metrics is of the same magnitude or order.

Wang et al. first presented the concept of policy-enabled handover in [88], in which three handover metrics: network bandwidth (B_n), power consumption (P_n) and monetary cost (C_n) were considered in handover decision making for the overlapped heterogeneous networks. These parameters are normalised using logarithm so that all of the parameters can be included in the same cost function f_n, which is the normalised cost. Mobile user can change the handover policies by adjusting the weight of each parameter in the cost function. However, *Wang*'s policy-enabled cost function had three metrics, which are limited and insufficient for a comprehensive evaluation of candidate networks.

In the Dynamic Vertical Handover Algorithm with Network Elimination [83], McNair et al. developed a two-dimensional handover cost function that took into account new handover metrics including service type, network conditions, system performance, user preference and so on. In one dimension, the function uses the different types of services requested by the user. In another dimension, it represents the cost to the network according to specific handover parameters [83]. The general form of the cost function is represented as:

$$f^n = \sum_s \sum_i w_{s,i} \cdot p_{s,i}^n \qquad \text{(Equation 5.4)}$$

where $p_{s,i}^{n}$ represents the cost in the *i*th parameter to carry out service *s* on network *n*, and $w_{s,i}$ represents the weight assigned to using the *i*th parameter to perform services. The vertical handover algorithm [83] is actually an extension of the policy-enabled handover [88] by considering both network constraints and network selection requirements of services.

Song et al. applied the Analytic Hierarchy Process (AHP) matrix in [84] to determine the weight assigned to handover metric in the cost function. The AHP procedure was introduced to structure a complex problem as a decision hierarchy of independent decision elements. The quantitative weights of decision elements are calculated by doing pairwise comparison of their relative importance. The comparison results constitute an AHP matrix, whose eigenvector determines the element's appropriate distribution to its parent, and can be transformed into the final weights.

5.3 Problem Definition

In homogeneous wireless networks, NCHO or Mobile-Assisted Handover (MAHO) mechanisms are the most widely used handover control approaches. They are implemented at the network end for controlling handover operations. As the handover control entity, the network is in a position of ensuring that any selected point of network attachment is accessible in a handover. In homogeneous wireless network, this is often an integral feature supported by network. Network controllers such as Mobile Switching Centre (MSC) in GSM are able to communicate with both serving Base Station (BS) and candidate BS for relevant trust information.

Furthermore, the number of network operators involved in network selection is quite limited due to the prohibitive costs of deploying mobile cellular services. Therefore, the security check on the existence of network trust relationship has been fundamentally integrated within the handover decision process initiated by networks.

However, in heterogeneous wireless networks, with a reduced cost of ownership, new access technologies such as IEEE 802.11 Wireless Local Area Network (WLAN) can be readily deployed by independent network operators. New access networks are expected to interwork with current mobile cellular networks and may need to be interconnected

with each other. As such, mobile terminals need to deal with both the heterogeneities of networking technologies and the multiplicities of network operators. Network selection coupled with trust relationship would become a prominent issue for roaming mobile users.

The network selection issue in heterogeneous wireless networks has raised a lot of concerns in the 3GPP specifications [72, 91] for Public Land Mobile Network (PLMN) and WLAN interworking. In 3GPP, it is proposed that network identity lists that are either user controlled or operator controlled can be used for network selection by mobile users. However, the 3GPP solutions provide very limited support to the multiplicity of network operators in NG wireless networks. State-of-the-art handover algorithms such as the MHOA, aims at optimising QoS by mitigating technology heterogeneity caused by the interworking. Unfortunately, they have not taken into account complex network trust relationships between multiple network operators, and thus are ineffective in resolving the multiplicity problem.

Handover failure occurs in two possible cases when a mobile user attempts to attach to its selected network in a handover. In the first case (Case A), handover authentication can fail in the absence of roaming agreements between the selected access network and the home network of the roaming user. This can be promptly detected by the AAA functions at visited networks. In contrast, in the second case (Case B), the HAAA may refuses access to the user-selected network, especially if a fine-grained AAA policy is applied for security reasons. In this case, a long AAA delay would be expected because the HAAA is involved in the validation. In both cases, the mobile user would be forced to reselect another network for roaming related access. Obviously, the mobile user would likely have missed the best time for triggering handover. Long interruption in connections would incur as a consequence.

Generally, for the mobile-controlled handover, the unawareness of the network trust relationship between a candidate network and the home network can cause unnecessary handover attempts. This inevitably increases signalling overheads and results in a long latency in a handover. Thus, the MHOA may become inefficient and has limited scalability in a multi-technology and multi-operator environment. Current handover approaches take into account the heterogeneities of access technologies of interworking

in order to achieve the best QoS, but the multiplicities of network operators have been ignored.

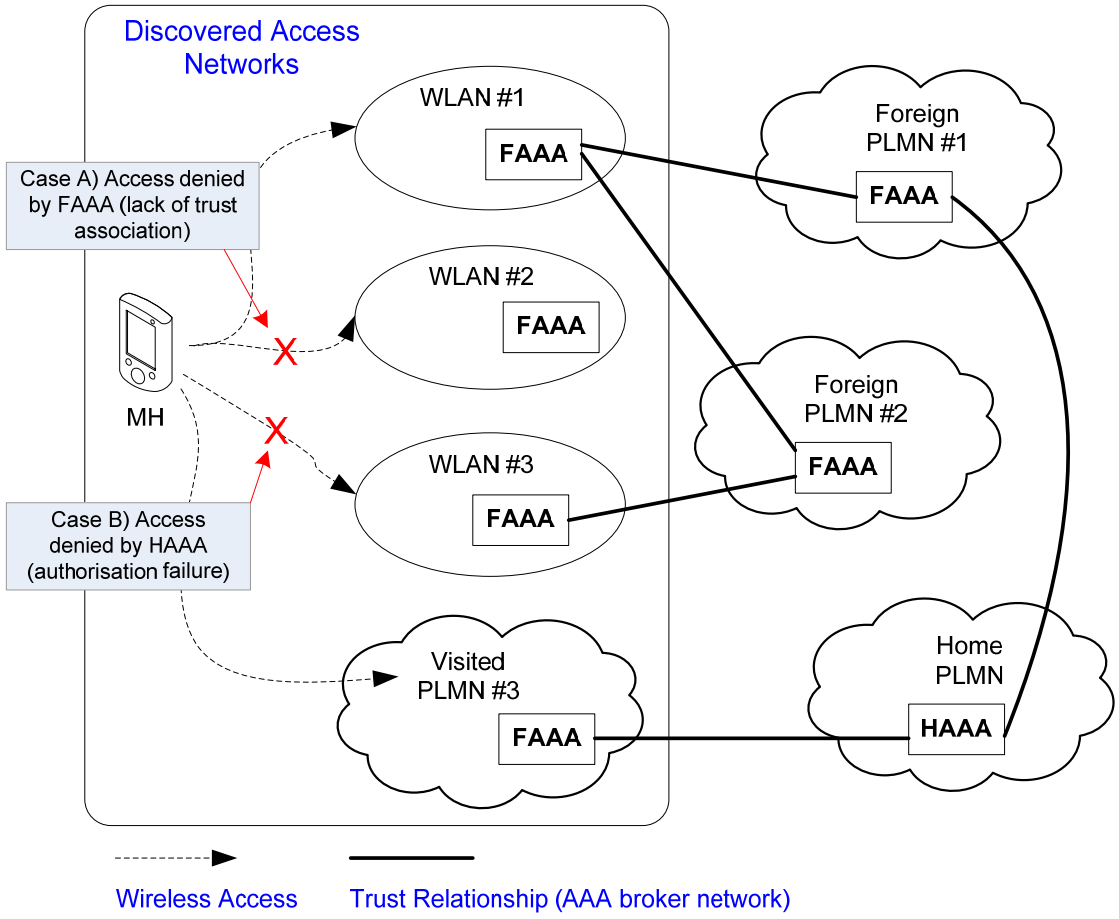


Figure 5.2 Network selection problem in a PLMN-WLAN interworking

A UMTS-WLAN interworking scenario is illustrated in Figure 5.2 as an example. A MH moves into the radio coverage of four access domains: WLAN#1, WLAN#2, WLAN#3 and VPLMN#3, which belong to different operators. It is assumed that the MH has been equipped with dual interfaces for UMTS and WLAN access. WLAN#3 and VPLMN#3 are directly associated with HPLMN via an interworking gateway. WLAN#1 has an indirect trust association with HPLMN, which is enabled by the third-party VPLMN#1. All the authentication requests originated from WLAN#3, foreign PLMN#3 and WLAN#1 can be sent to HPLMN directly or through an AAA proxy. However, WLAN#2 is isolated from other networks in terms of trust association. Therefore, WLAN#2 has actually been excluded from providing services to the roaming

users of HPLMN, because no trust association with HPLMN can be found. QoS-driven network selection schemes may initially choose WLAN#2 as the best candidate, but later realise that mobile users are not allowed to access the selected domains.

According to the above analysis, the handover requirements in a NG multi-technology and multi-operator network are summarised as follows. Firstly, handover algorithms process should not depend on underlying access technologies to deal with their heterogeneities in the interworking. Handover algorithms should be scalable over a variety of access technologies and allow the introduction of new access technologies. Secondly, network trust relationship between a candidate access network and a mobile's home network is the prerequisite for successful handover. Therefore, in a multi-operator environment, it should be checked with network selection before any handover decision is made. Thirdly, handover algorithms should be capable of accommodating various handover metrics such as signal strength, network conditions and so forth so as to guaranteee QoS.

5.4 Trust-Assisted Handover Decision Algorithm

Chapter 4 has presented a neighbour network trust information retrieving scheme (named as Neighbour Trust Correlation, NTC), which is based on the analysis of a large number of mobiles' handover history. With the NTC scheme, the network trust information of neighbouring candidate networks to a mobile's home network can be made available to an access network. Thus, it is assumed that Points of Attachment (POA) at access networks have knowledge of their neighbour network trust patterns.

In this section, with the input from the NTC scheme, a Trust-Assisted Handover Decision Algorithm (THOA) is proposed for making handover decision in a multi-operator and multi-technology environment. The proposed THOA algorithm uses the NTC data of the candidate POAs as an additional input, and is expected to be used at the mobile terminals. Because it can learn related network trust information before conducting network selection, a mobile user is able to check whether it is possible to access a candidate network before selecting that network and initiating the handover. This would avoid unnecessary handover attempts in the NG multi-technology and multi-operator network. The basic ideas behind the THOA are as follows. The serving

POA provides the NTC data (network trust information) of its neighbouring POAs to the attached mobile user using the active connection. The mobile user stores this received network trust information about the surrounding POAs for future use (for example, when it moves out of the radio coverage of the serving POA.). During a handover, both network trust relations and QoS are considered. This enables a much more reliable handover strategy when a mobile user is dealing with multiple network operators.

To bring trust awareness to the THOA, a new metric named *trust coefficient* (τ_n) is introduced. The trust coefficient normalises the NTC data derived through the NTC model as described in Chapter 4. The value of trust coefficient is expected to be inversely proportional to the numerical value of the NTC obtained from the NTC model such as the Trust Association Hop (TAH) based on the one shown in Figure 4.3. The mobile user MH maintains a local network trust coefficient table for network selection, instead of a 3GPP specified user controlled or operator controlled identifier list [72]. The update of this trust coefficient table is largely determined by the MH's location. Table 5.1 is illustrated as an example of the trust coefficient table that can be built for the scenario of Figure 5.2. An entry can be created for each network that is involved in the neighbour NTC pattern regardless of the access technology it is based on.

Table 5.1 An example of network trust coefficient table for network selection

Network ID/ Operator ID	Relationship to HPLMN	Trust Coefficient
HPLMN	--	1
WLAN#1	via VPLMN#1	0.5
WLAN#2	no trust relation with HPLMN	0
WLAN#3	via PLMN#2, WLAN#1 and PLMN#1	0.2
PLMN#1	directly connected to HPLMN	0.8
PLMN#2	no trust relation with HPLMN	0
PLMN#3	directly connected to HPLMN	0.8

With the candidate network's NTC information, the MH can work out which candidate networks should be avoided during the network selection because of lack of trust relationship due to low trust coefficiency. If two access networks appear to have the same amount of NTC with the home network (e.g. WLAN#3 and PLMN#1 in Figure

5.2), other handover deciding criteria such as bandwidth, network latency and so forth, can be compared for the sake of better connection QoS.

Mathematically, τ_n is used to represent trust coefficient. The cost function of the THOA is derived as follows:

$$f_n = \frac{C_n}{\tau_n} = \frac{\sum_i w_i \cdot N(m_i)}{\tau_n} \qquad \text{(Equation 5.5)}$$

in which C_n represents the QoS evaluation part of the THOA, which takes into account multiple handover metrics. The conditions of C_n has been explained in (Equation 5.3). A flow chart of the THOA process is demonstrated in Figure 5.3. In the network discovery phase, the MH discovers all the available POAs. Thereafter, it retrieves their corresponding trust coefficient values from its local network trust coefficient table. Those networks having a positive non-zero entry in the table (with $\tau_n > 0$) are regarded as within the MH's circle of trust. The *circle of trust* is defined as a group of networks that have established trust associations with the MH's home network. In the handover decision process, the discovered POAs within the circle of trust are selected for handover. The handover cost of attaching to each POA within the circle of trust will be calculated according to (Equation 5.5). Their handover costs are compared, and the one having the minimum value of f_n is selected as the next POA. If two POAs appear to provide the same QoS as a result of the multicriteria cost evaluation (having the same value of C_n), the one with the higher value of τ_n is selected. This justifies that home network is always superior to foreign networks in regard to network selection when their network conditions are the same.

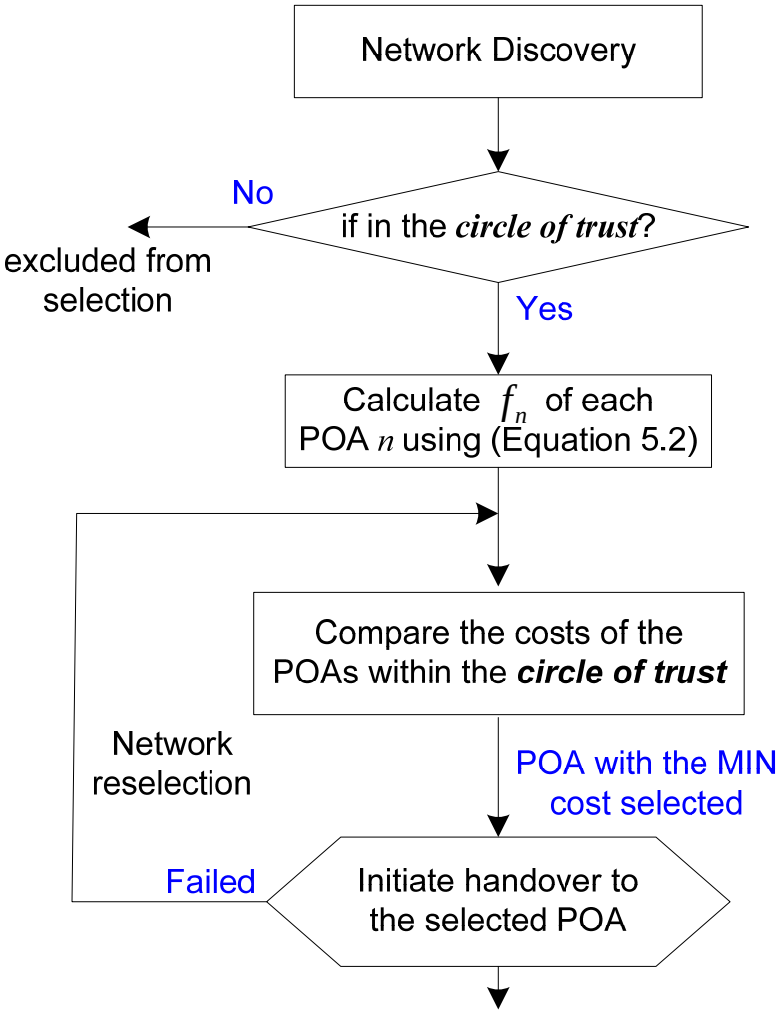


Figure 5.3 Trust-assisted handover algorithm flow chart

Trust coefficient τ_n is intended to make the candidate networks having weak trust relationship to home network less desirable in regards for network selection.

Next, the derivation of the trust coefficient function is explained for the NTC model of Figure 4.3. h_n represents the Trust Association Hop (TAH) count between a visited network and a mobile's home network. Intuitively, as the TAH count h_n grows, τ_n is expected to decrease so that the handover cost of switching to network n is increased according to (Equation 5.5). Generally, a handover algorithm should be encouraged to select those networks "close" to a mobile's home network (measured by trust relationship) in most cases. So, the values of τ_n for low hop count should be closer to 1 such that the handover costs f_n of the networks closer to the MH's home network still

remain low. This will make those networks having close trust relationship with the MH's home network remain competitive in network selection. As soon as h_n reaches a predefined TAH threshold, τ_n should drop rapidly to zero and thus exclude network *n* as a candidate for handover. This is justified by the fact that too many third parties involving in intermediating AAA messages may become inefficiency and weaken the reliability of handover. The predefined TAH threshold is denoted as Effective TAH Scope (ETS, H_{eff}). The ETS indicates how network trust relationship may have an impact on handover related AAA. Based on the above analysis, it was found that the reciprocal of inverse tangent function can provide the logic for the trust coefficient function. Figure 5.4 shows the curves of the proposed trust coefficient function, compared with a reciprocal of an exponential function.

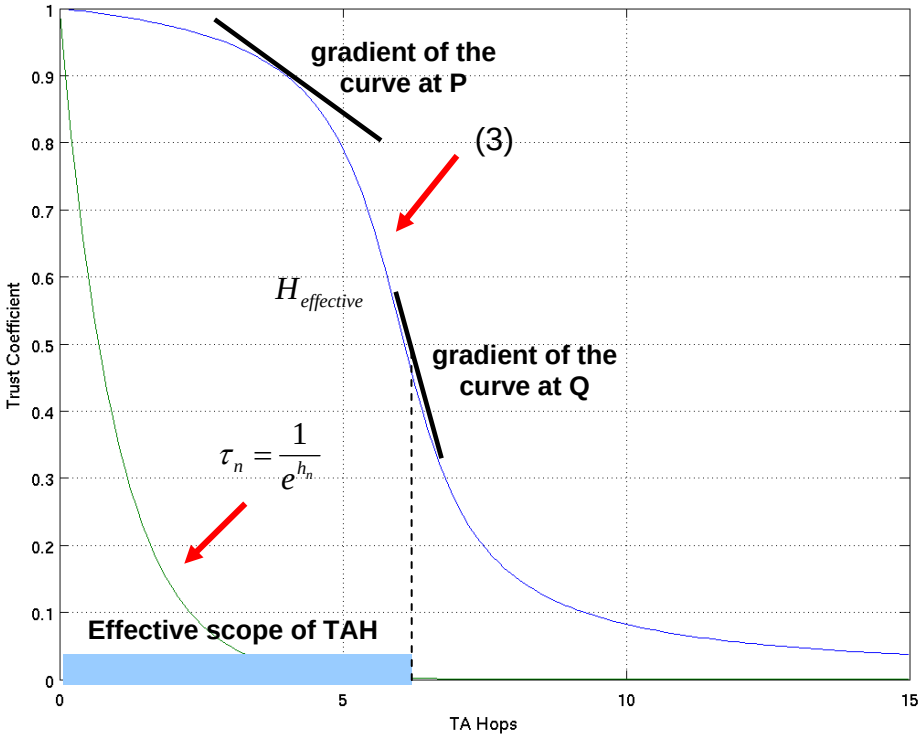


Figure 5.4 Trust coefficient function (with $H_{eff} = 6$)

The trust coefficient function is represented as follows:

$$\tau_n = d \cdot \left[\frac{\pi}{2} - \arctan\left(h_n - H_{eff}\right) \right] \quad \text{(Equation 5.6)}$$

where H_{eff} represents the effective TAH scope, within which τ_n is made large enough to keep network *n* competitive in network selection. *d* is a constant determined by H_{eff} (when $h_n = 0$, we let $\tau_n = 1$). H_{eff} can be dynamically adjusted by handover policies set by home network operator. Thus, handover AAA delay can be "programmed" to suit different services. According to (Equation 5.6), home network is always fully trusted ($\tau_n = 1$ with $h_n = 0$). As h_n increases, handover cost f_n would steadily increase. Once h_n reaches H_{eff}, τ_n would decrease dramatically, and effectively make the handover cost f_n of network *n* too large to be considered for network selection.

Using (Equation 5.6), Equation 5.5 can be further represented as:

$$f_n = \frac{\sum_i w_i \cdot N(m_i)}{d \cdot \left[\frac{\pi}{2} - \arctan\left(h_n - H_{eff}\right)\right]} \quad \text{(Equation 5.7)}$$

This solution assumes that the MH has to reauthenticate to its home network prior to the association with new POA, e.g. AP of 802.11 WLAN, in a handover. In some specific interworking cases, this may possibly be avoided by forwarding the cached AAA contexts from the old POA to the new POA. However, in a heterogeneous wireless network, especially a multi-operator environment, it can not always be guaranteed that any pair of adjacent access networks has a trust relationship established between each other. Moreover, as stated in [53], there is no standard state transfer protocol that could be used to achieve such a transfer functionality in a secure way in IP networks. Therefore, for the inter-operator handover, it is reasonable to assume that authentication occurs in every handover attachment.

5.5 System Analysis and Model

5.5.1 System Analysis

Handover delay is the essential performance concern for handover in heterogeneous wireless networks. In this study, *handover delay* is defined as the time interval between

the moment a mobile user loses its connection with the old POA (oPOA) to the time it receives the first packet from the new POA (nPOA). To analyse handover delay, a system model is developed.

In a typical "break-before-make" handover scenario of the interconnected WLAN-UMTS network, handover delay comprises of the following major components:

Movement Detection (MD): Movement detection delay (t_{MD}) is the period of time taken on deciding whether a mobile user has moved to a different network. Due to the asymmetry of move-in and move-out handover scenarios [89] in the interconnected WLAN-UMTS network, WLAN→UMTS and UMTS→WLAN handover mean different movement detection latencies. In Mobile IPv4 specification [21], two algorithms have been specified with regard to movement detection. In the first method, a mobile user checks agent advertisement lifetime (ADF) periodically. Upon the expiration of ADF, it assumes that it has lost contact with the agent router. In the second approach, the network prefix of the advertisements from the nPOA is compared with the network prefix of the oPOA. A change in network prefix would imply that the mobile user may have moved to another subnet. In this study, it is assumed that the first method is applied to WLAN→UMTS handover and handover between homogeneous networks. While, for UMTS→WLAN handover, t_{MD} is equal to 0 because discovering a WLAN nPOA would not cause transmission interruption if a second interface is employed.

Network Selection (NS): Network selection delay (t_{NS}) is denoted as the time required for network solicitation (L2+L3) (α) in absence of the valid router advertisements from the nPOA, plus handover decision making time (β) at mobile terminals. Mobile terminals with multiple network interfaces [92] can be enabled to conduct network solicitation using a standby interface, while carrying data traffic on the primary interface simultaneously. To simplify the analysis, it is assumed that handover occurs only between heterogeneous wireless networks. When multiple interfaces are simultaneously used [92], network solicitation can be completed preliminarily ($\alpha = 0$). While, β is often a constant determined by the computing power of mobile terminals. Therefore, t_{NS} is believed to be able to remain unchanged because handover decision can be made independent of network discovery.

Address Configuration (AC): Address configuration delay (t_{AC}) is incurred when a mobile user obtains a topologically correct address for local access from visited networks. t_{AC} depends on the specific mobile IP implementation and the type of Care-of-Address (CoA) [21] to be used.

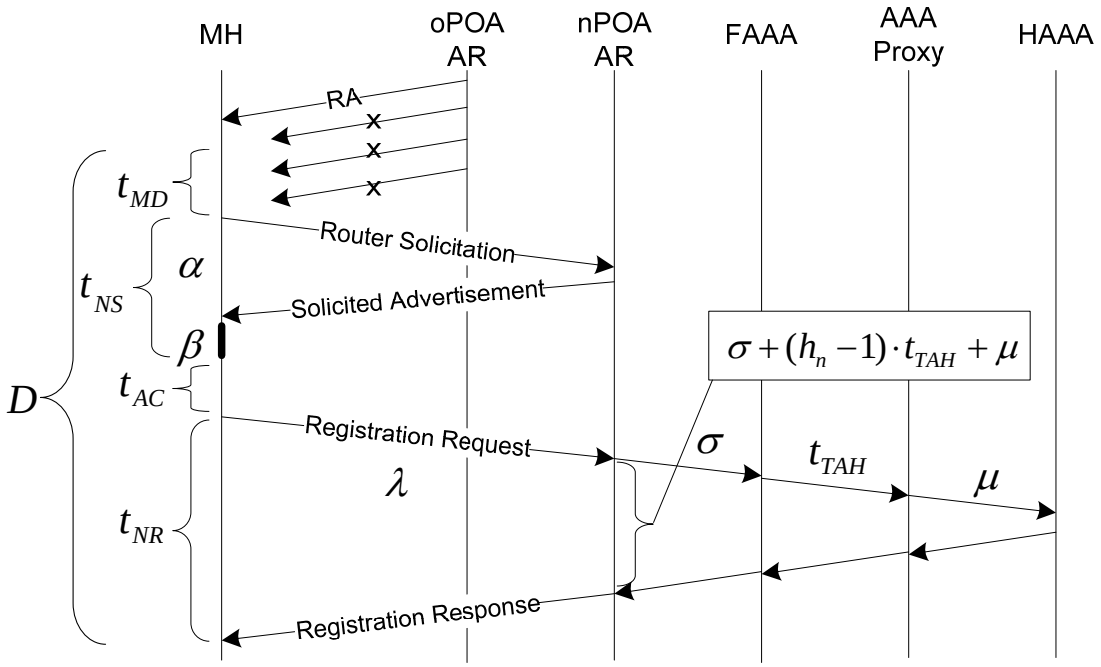


Figure 5.5 An Analysis of delay in WLAN→UMTS handover

Network Registration (NR): Network registration delay (t_{NR}) is the period of time taken on establishing a mobility mapping between a mobile user's CoA and its home address plus the delay on the AAA processing delay. Mobile user to nPOA Agent Router (AR) Round Trip Time (RTT) (λ), nPOA AR to foreign AAA server FAAA RTT (σ), FAAA to FAAA RTT (t_{TAH}), and FAAA to home AAA server HAAA RTT (μ), and AAA processing at HAAA (t_{AAA}) may combine to contribute to t_{NR}. It is reasonable to assume that the signalling delay between AR and local AAA server can be neglected ($\sigma = 0$) due to much less signalling cost for internal communications compared with external ones. The MH-AR RTT largely depends on accessing technologies, and thus it may result in different values for WLAN (λ_{WLAN}) and UMTS (λ_{UMTS}). Based on the above analysis, the network registration delay function can be represented as follows:

$$t_{NR} = \lambda + \sigma + (h_n - 1) \cdot t_{TAH} + t_{AAA} \qquad \text{(Equation 5.8)}$$

In the roaming AAA architecture, HAAA and FAAA are actually similar entities with common AAA functions and implementations, but play different roles for a particular mobile user. They exchange their roles for different roaming users. For this reason, the same AAA signalling delays between a FAAA and an AAA proxy, and between an AAA proxy and a HAAA can be expected in the AAA broker network [8]. Equation 5.8 can thus be simplified:

$$t_{NR} = \lambda + h_n \cdot t_{TAH} + t_{AAA} \qquad \text{(Equation 5.9)}$$

Based on the above analysis and Equation 5.9, handover delay incurred in each handover attempt is derived as follows:

$$\begin{aligned} D_j &= t_{MD} + t_{NS} + t_{AC} + t_{NR,j} \\ &= t_{MD} + \beta + t_{AC} + \lambda + h_{n,j} \cdot t_{TAH} + t_{AAA} \end{aligned} \qquad (j = 1, 2 \ldots)$$

where j means the jth handover attempt. Thus, for Case A (lack of trust association) described in Figure 5.2, the trust relationship verification by the HAAA can be done locally at the visited network by the FAAA. The incurred delay for all the related operations is represented as t_{auth}. The handover delay function for the "no trust association" scenario is given below:

$$D = t_{MD} + \sum_{j}^{m} (t_{AC} + \lambda + t_{auth}) \qquad (j = 1, 2 \ldots m) \qquad \text{(Equation 5.10)}$$

For Case B (authentication failure) shown in Figure 5.2, the following handover delay function applies:

$$D = t_{MD} + \beta + \sum_{j}^{m} (t_{AC} + \lambda + h_{n,j} \cdot t_{TAH} + t_{AAA})$$

Because the handover decision computing time β is rather small compared to the total signalling delay, it is omitted in this analysis.

$$D = t_{MD} + \sum_{j}^{m} (t_{AC} + \lambda + h_{n,j} \cdot t_{TAH} + t_{AAA}) \qquad \text{(Equation 5.11)}$$

5.5.2 System Model

The simulation scenario is designed as such. A MH will be trying to complete a series of handover in a heterogeneous wireless environment with overlapped radio coverage. The MH makes efforts to make sure it's always best connected when network conditions change. There are a total of 10 POAs in a proposed hotspot. five are UMTS POAs and the rest are WLAN POAs. The MH retrieves network condition information of each POA from its periodic advertisement being broadcast. It can associate with one POA at any time.

The POA may establish a trust relationship with the MH's home network or have it disabled by the operator's roaming policies. All the POAs share a common AAA proxy network to the MH's home network. The TAH h_n vectors for the UMTS POAs and the WLAN POAs are $h_{1\ldots5}^{UMTS} = (1, 3, 5, 7, 9)$ and $h_{6\ldots10}^{WLAN} = (2, 4, 6, 8, 10)$ respectively. The POA(s) having their trust associations to the MH's home network disabled is(are) randomly selected. Every POA is assumed to be able to supply 3 handover metrics: *available bandwidth*, *network latency*, and *packet loss*, the weights of which are quantified by Analytic Hierarchy Process (AHP) presented in [84]. The time between these metric values being changed varies according to an exponential probability distribution with the mean of *u*. The change in the value of the metrics follows a Markov chain with the below transition probability matrix. The 4×4 matrix below shows the transition probability matrix for a metric with four possible values.

$$\begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.3 & 0.4 & 0.3 & 0 \\ 0 & 0.3 & 0.4 & 0.3 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

The available bandwidth vector of the UMTS POAs is [32, 64, 128, 384, 768, 1024, 2048, 3600] kbps, while the WLAN POAs have [1, 2, 4, 6.5, 8, 11] Mbps. For the network latency factor, [20, 60, 100, 150, 210] is given to the UMTS POAs and [10, 30,

60, 120, 180] is assigned to the WLAN POAs. Both UMTS and WLAN have the same values for the packet loss vector $\left[10^{-4}, 5\times10^{-4}, 10^{-3}, 5\times10^{-3}, 10^{-2}\right]$.

The parameters for the simulation are illustrated in Table 5.2. When the network conditions and the trust relationships between network domains are changed, various handover decision scenarios can be composed. The simulation results will be shown in Sec. 5.6.

Table 5.2 Simulation parameters

Parameter	UMTS (ms)	WLAN (ms)
Move detection t_{MD}	[0, 360]	[0, 210]
Network discovery α	0	0
HOA computing β	0	0
Address configuration t_{AC}	30	30
MH-to-AR RTT λ	330	150
TAH signalling t_{TAH}	30	
Home AUTH t_{AAA}	80	80
Local AUTH t_{auth}	60	60

As aforementioned, handover delay is the primary performance concern in the interworking of heterogeneous wireless networks. The simulation has been designed to focus on the impact of network trust relationship on handover delay, because of the initiatives of this study as stated in Sec. 5.3. QoS in handover is examined in order to evaluate how the proposed trust-assisted handover approach may influence upper layer services.

A new parameter named trust density is defined to determine the accessibility of access networks. *Trust density* refers to the fraction of the total trusted POAs that have established trust association with a roaming mobile user's home network. If the trust density in a hotspot has the value of 1, a mobile user would be able to attach to any discovered POA.

Load balance factor (LBF) is another parameter proposed to evaluate how handover incurred load may be distributed among all the available POAs. Upon accepting handover request, a POA needs to open its local resources to the attached mobile user. This would place a certain amount of burden $l = u \cdot \rho$ on its systems. u denotes the

number of successful handovers. A uniform cost ρ is assumed to be related for processing the handover request and assigning local resources. Ideally, handover load should be evenly distributed. The expression for load balance factor is:

$$L = \frac{(u_{\max} - u_{\min}) \cdot N}{\sum_{a=1}^{N} u_a} \quad \text{(Equation 5.12)}$$

where $u_{\max} = \max(u_1, \ldots, u_N)$ and $u_{\min} = \min(u_1, \ldots, u_N)$. N denotes the number of available POAs for accepting handover. According to Equation 5.12, $L = 0$ is the most desirable, because if it is the case, all the POAs process an equal number of handover requests. A high LBF value means that some POAs play dominant role in taking handover requests, while other POAs rarely serve roaming mobile users.

5.6 Simulation Results

5.6.1 Handover Delay

The scenarios with a random distribution of trust density among the 10 POAs are proposed. The 10 POAs are within the range of the MH. 20 runs of simulation are conducted for each concerned parameter to investigate its impact on handover performance. In this analysis, trust density, AAA failure, Effective TAH Scope (ETS), and round trip time RTT are of major concerns. In each simulation run, the POAs that have their trust association with the MH's home network disabled are randomly selected. Intuitively, it makes sense to assume that at least one UMTS POA should be available to the MH in order to maintain user sessions. Furthermore, it is assumed that another WLAN POA needs to be present so that a handover can be measured in the simulation. The trust density of the scenarios changes in value from 0.2 to 1. The THOA was assigned different values of ETS (H_{eff}) in each simulation run when being compared with the MHOA represented by Equation 5.3.

Figure 5.6 clearly demonstrates that the THOA is able to make roaming mobile users unaffected by the trust pattern of visited networks with regard to handover delay. In contrast, roaming mobile users employing the MHOA experience longer handover delay,

especially in a low trust density scenario (A large portion of the POAs has no trust association with mobile user's home). If the THOA takes a median ETS value of 8, an implementation of the THOA would see a reduction of 35% of the handover delay. The much improvement is due to the fact that the THOA can always avoid unnecessary handover attempts by checking accessibility of networks before triggering handover. Such a preliminary check makes sure that each handover attempt would be effective. While, the MHOA would have to deal with handover reselection(s) if an unsuccessful handover takes place. Apparently, more handover attempts lead to longer handover delay, and more signalling overheads. On the other hand, the changes in the ETS of the THOA can result in different handover delay. Smaller handover delay is observed when the THOA is given a low value of ETS. A decreased ETS means the network selection scope is narrowed down.

In another experiment, the load balance factor LBF is examined. It is found that an increased LBF was observed when the ETS was reduced in the sake of faster handover as illustrated in Figure 5.7. Intuitively, a larger LBF means that the handover requests are not evenly distributed to the available POAs. This is true because more access networks are regarded as being "far away" from the MH's home network during handover decision making. Some of the networks would thus be excluded from network selection. However, in favour of a fair balance of load, the THOA with a high value of ETS (e.g. ETS=15) can achieve the similar performance as the MHOA in regards to the LBF. The comparison is shown in Figure 5.7.

The level of trust association has an interesting impact on handover performance as shown in both Figure 5.6 and Figure 5.7. In this simulation, the level of trust association is determined by two parameters: the probability of no trust association (p_A) and the probability of authorisation failure (p_B). The latter determines if the mobile user's home network would grant access through the selected foreign network. When the THOA with a smaller ETS is applied, the LBF appears to be affected by the trust density of access networks. The data line of the THOA with ETS=4 in Figure 5.7 shows this effect. However, the MHOA shows a fairly smooth LBF curve due to its unbiased network selection process. The THOA tends to give access networks "close" to mobile user's

home network in trust relation higher priority. The THOA can perform well in achieving load balance when the ETS is made suitably large.

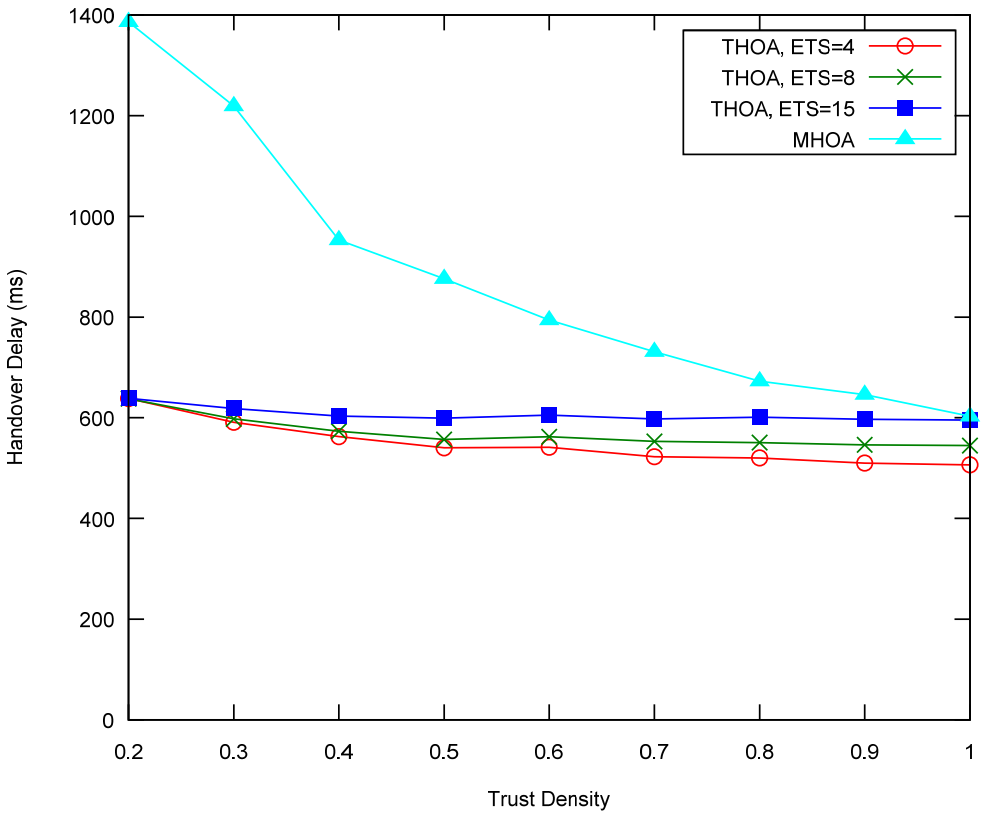


Figure 5.6 Handover delay vs. Trust density [$p_A = 50\%$]

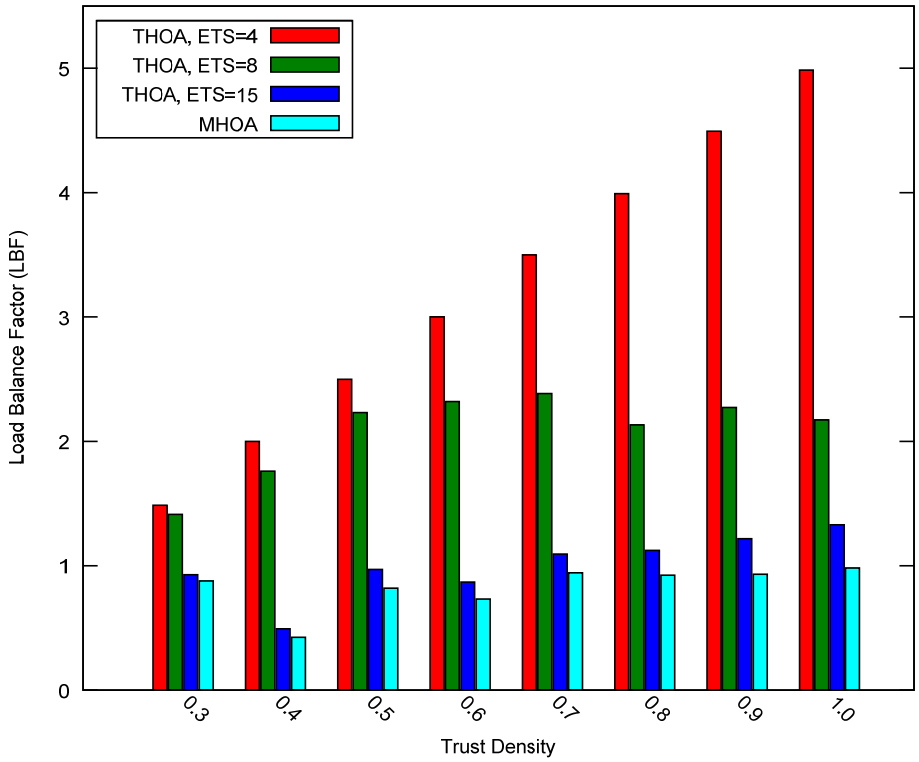


Figure 5.7 Load balance factor vs. Trust density [$p_A = 50\%$]

It has been observed that the level of trust association between home network and foreign networks can have certain effect on handover delay if the MHOA is applied. Figure 5.9 shows that the handover delay grows from 1070ms to 1255ms, when the probability of no trust association (trust density equals 0.3) drops from 1 to 0 in the MHOA applied case. The results suggest that reducing the trust ambiguity in the AAA relationship between networks may improve handover performance in low trust density scenarios. In comparison, the same experiment was conducted using the THOA. Figure 5.8 shows the experimental results of applying the THOA. The THOA apparently mitigates the side effect of implementing the fine-grained AAA policies (e.g. a detailed classification of subscribers for authorisation) in home networks.

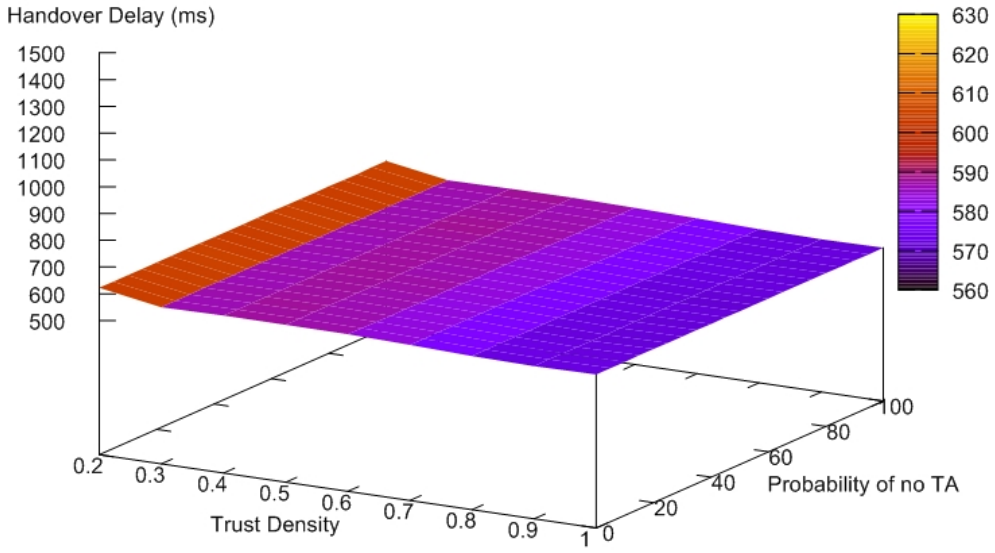


Figure 5.8 Handover delay vs. Trust density vs. Probability of no TA [THOA, ETS=10]

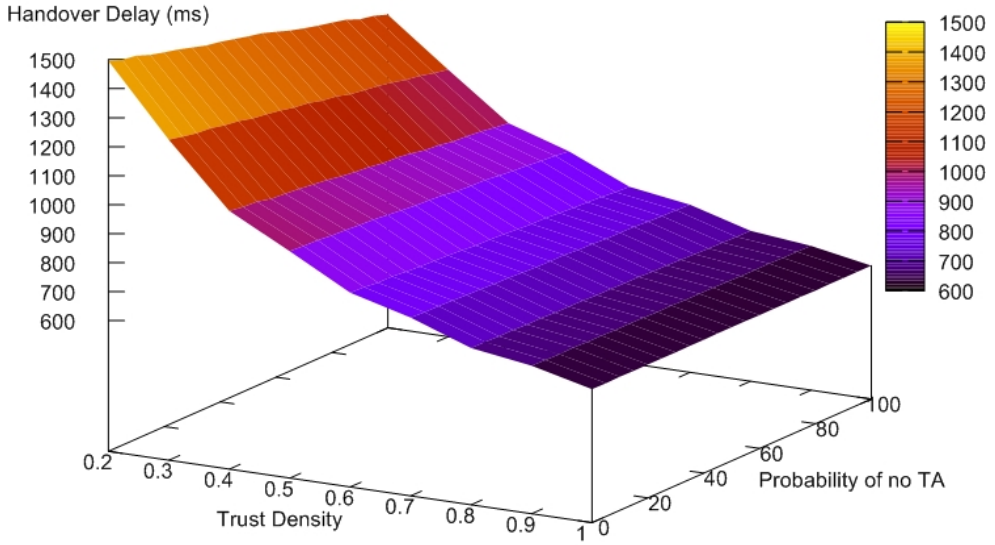


Figure 5.9 Handover delay vs. Trust density vs. Probability of no TA [MHOA]

5.6.2 Impact of Effective TAH Scope

In this analysis, the effect of adjusting the THOA ETS on handover delay and load balance factor is measured. In the simulation, three typical handover scenarios have been considered: low trust density (0.3), medium trust density (0.6) and high trust density (0.9). When the effective TAH scope ETS is increased in an attempt to make the load fairly distributed among the POAs (shown in Figure 5.11), the MH with the THOA enabled may experience an increase of handover delay in the medium and high trust

density scenarios (demonstrated in Figure 5.10). However, in the low trust density scenarios, Figure 5.10 shows that the ETS has much limited impact on handover delay. The widening of the ETS in the THOA would result in a better load balance pattern among the access networks as demonstrated in Figure 5.11. But, the LBF tends to go into a stable state. Then, further increase of the ETS would deteriorate the performance. This phenomenon manifests itself on the performance of both handover delay and LBF. The simulation results give guidance on how the THOA can be tuned to meet the performance requirements in various handover scenarios.

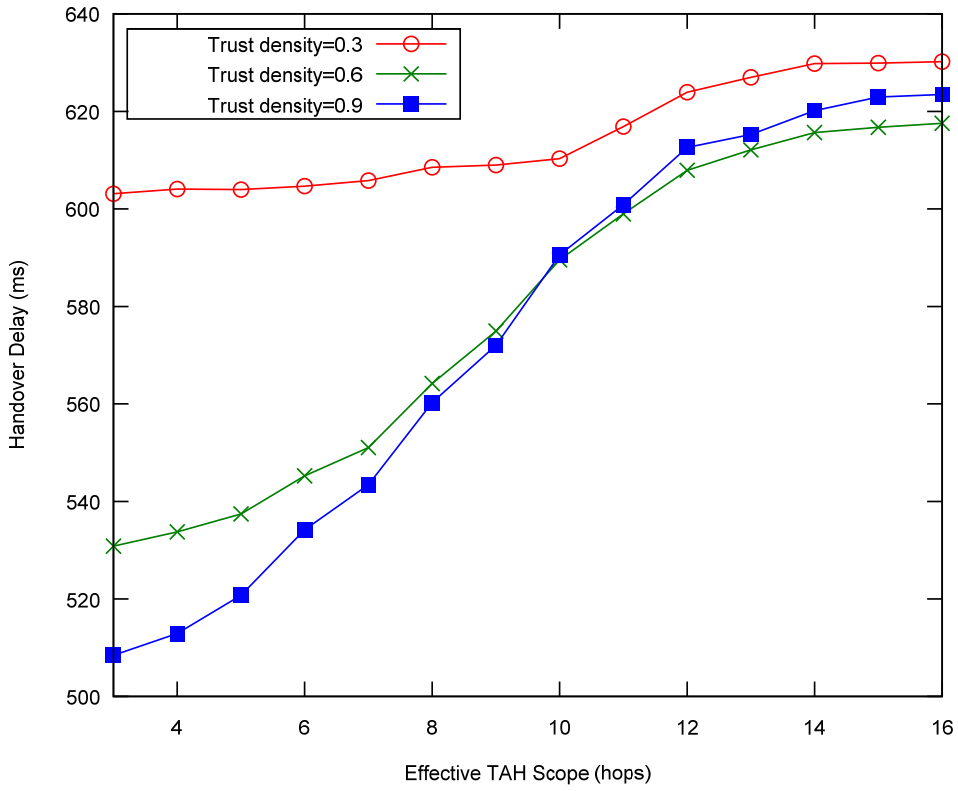


Figure 5.10 The effect of THOA ETS on handover delay

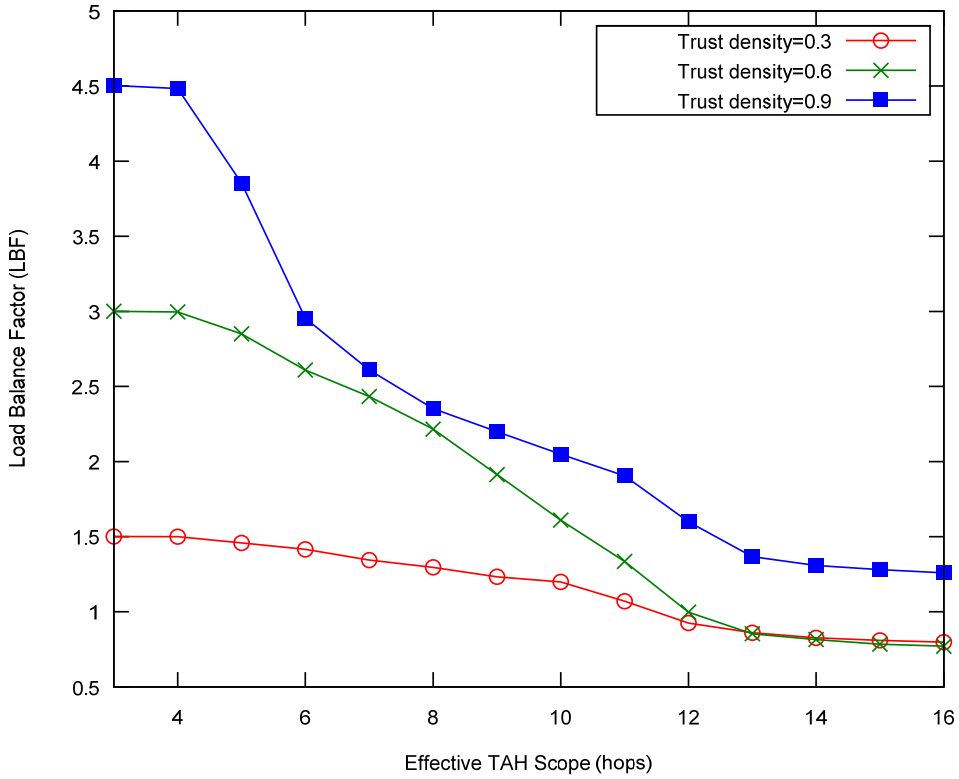


Figure 5.11 The effect of THOA ETS on load balance factor

5.6.3 Quality of Service

Besides the aforementioned objectives, another consideration in doing a handover is to guarantee the quality of service QoS. Three parameters: bandwidth, network latency and packet loss are employed to evaluate the QoS observed during the MH's attachment to a new POA.

The simulation results for all three parameters are illustrated as (a-c) of Figure 5.12. The three figures show that trust density can have either negative or positive effect on the QoS of a user connection. Generally, the QoS obtained by the MH got much improved in the high trust density cases (with a trust density>0.7) irrespective of what parameters the handover decision algorithms were taking.

In regards to the average bandwidth, applying the THOA may lead to a slight reduction of the bandwidth available to the MH compared with the MHOA. With an additional parameter ETS, the THOA may apply strict rules in selecting a POA due to trust considerations. In an extreme case of the THOA (ETS=4, which means much stricter

selection criteria), the bandwidth obtained can be greatly affected as shown in Figure 5.12. An increase of trust density in networks may not even improve available bandwidth. However, if the THOA is tuned with a high ETS (ETS=15), the bandwidth observed in using the THOA can be comparable to what is available with the MHOA. In high trust density scenarios, the performance of the THOA is very close to that of the MHOA in regards to the average bandwidth.

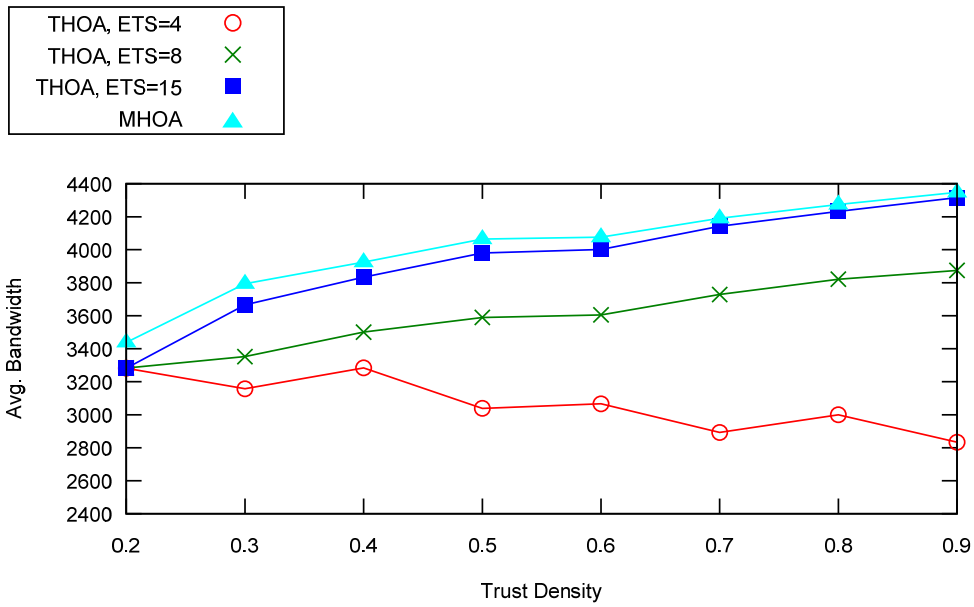
The simulation results for the average network latency also demonstrate that the average latency incurred in implementing the THOA may be comparable to that of applying the MHOA, although QoS does not take priority in the THOA. High latency with the THOA can be avoided in most trust density scenarios by adjusting the ETS of the THOA according to Figure 5.12.

The average packet loss gives similar comparison results of the other two parameters. Generally, the performance of the THOA is less desired than that of the MHOA. However, this degradation in using the THOA can be well compensated by tuning the THOA algorithm. The performance difference between the THOA and the MHOA is less a concern when trust density of networks is quite high, which many networks may be available for network selection.

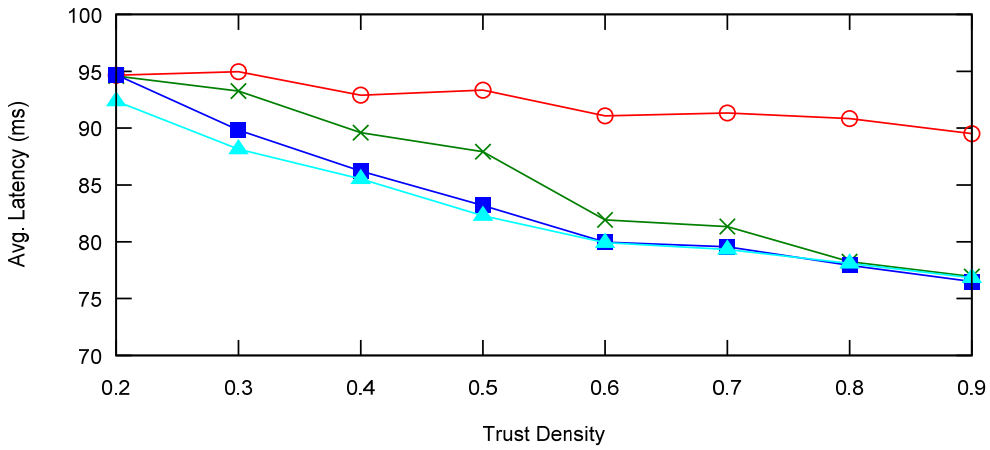
The impact of handover decision algorithms on QoS can be more obvious when the POAs are heavily loaded and a large number of mobile users compete for local resources.

Compared with the THOA with a medium range of ETS, the MHOA appears to provide better QoS. Applying the MHOA on mobile terminals means 6.3% more bandwidth, 14.3% less network latency and 28.6% less packet. However, if the THOA is tuned with a large value of ETS that indicates a more catch-for-all network selection policy, the QoS obtained from using the THOA is comparable to what is provided by the MHOA. When the THOA ETS is set to 15, the QoS gain of the MHOA over the THOA is reduced to 2% for average bandwidth, 0.8% for average network latency, and 6.3% for average packet loss. This is demonstrated as two partially overlapped lines (MHOA vs. THOA with ETS=15) in Figure 5.12. The minor difference can be explained by the fact that the THOA may sometimes select the POA having the slightly less QoS, e.g. a lower

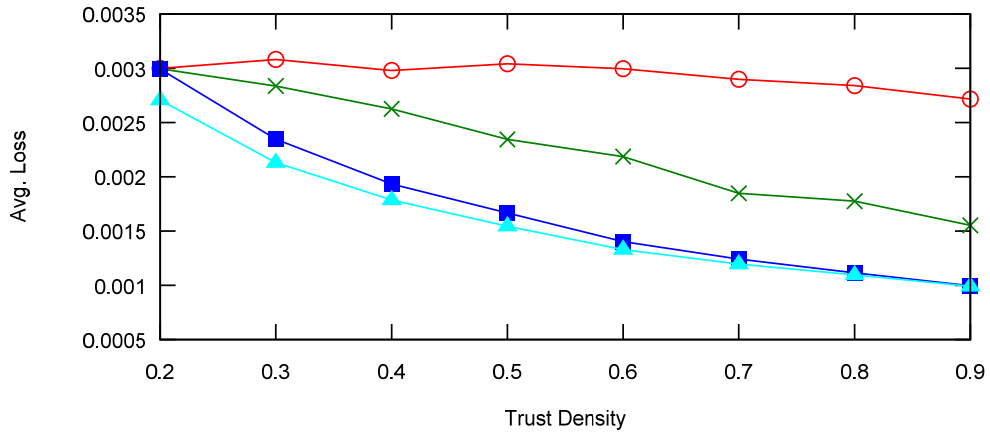
bandwidth POA, to guarantee a more reliable attachment. However, the MH is rewarded a much prompt handover process and stronger adaptability to various trust density scenarios as stated in Sec. 5.6.1. In real systems, a bit sacrifice in QoS (about 3% on average) is often neglectable and can be easily compensated at upper application layers.



(a) Average bandwidth



(b) Average network latency



(c) Average packet loss

Figure 5.12 Comparison of the THOA and the MHOA on QoS

The simulation results shown in Sec. 5.6 demonstrate that the THOA can provide much more reliable handover than the MHOA in a multi-operator environment. The THOA can provide flexibility in balancing QoS and handover delay and thus improve performance output in handover. The proposed algorithm is especially suitable for the low trust density scenarios, where reliable handover is rarely provided using current handover approaches such as MHOA.

5.7 Conclusion

Current known handover approaches such as the MHOA, which were designed for dealing with the network heterogeneities can not deal with the coexistence of multiple network operators. Other solutions such as network-controlled and mobile-controlled operator lists specified by IETF have been proposed to address this deficiency, but only as an add-on to network selection [72]. Unfortunately, they are unable to effectively support the changes of network trust relationship between networks. Current handover solutions have to rely on other mechanisms that appear as an add-on to support AAA in a handover rather than forming an integral part of their handover algorithms.

The trust-assisted handover algorithm is presented to address this problem, since it is anticipated that it would become common as more and more independent operators are involved in interworking. The proposed THOA algorithm provides an efficient and flexible approach to deal with complex network trust relationship during a mobile's handover decision making. At the mobile terminal, the network trust information about neighbours that is sent by the serving network is fed into the mobile's handover decision algorithm. Either the mobile user or its home network operator can set handover policies to adjust the THOA ETS, and thus influence handover AAA delay. The simulation results showed that the THOA can provide a more reliable handover, and result in a 35% reduction in handover delay compared with the MHOA. By tuning the THOA ETS, it was shown that traffic can be evenly distributed among the POAs. It was shown that the implementation of the THOA would not compromise QoS, which is a very important consideration in deploying real wireless networks.

Chapter 6

PROXY BASED AUTHENTICATION LOCALISATION SCHEME FOR HANDOVER

6.1 Introduction

In Next Generation Networks, a number of heterogeneous and distributed wireless networks are expected to converge and have a common all-IP network architecture [6] These technologies could range from Wireless Local Area Network (WLAN) such as IEEE 802.11, to the third generation (3G) cellular networks such as Universal Mobile Telecommunications System (UMTS). These heterogeneous wireless networks are expected to co-exist as overlay networks and enable ubiquitous data services and provide very high data rates in strategic locations [4]. The interworking of different wireless systems and the demand for ubiquitous services present enormous challenges to network security [13, 52].

The trust model for wireless network is identified by three mutual trust relations between network entities [52]: two explicit trust relations between a Mobile Host (MH) and its home network, and between the home network and the Access Network (AN), in addition to one implicit trust relation between the MH and the AN as shown in Figure 6.1. The implicit trust relation between the MH and the AN is often dynamically established during the handover attachment to the AN. The MH and the AN must verify each other during the handover to make sure that both hold a trust relation to MH's home network. Such identity verification in a handover, better known as mutual authentication, often needs the involvement of the home network.

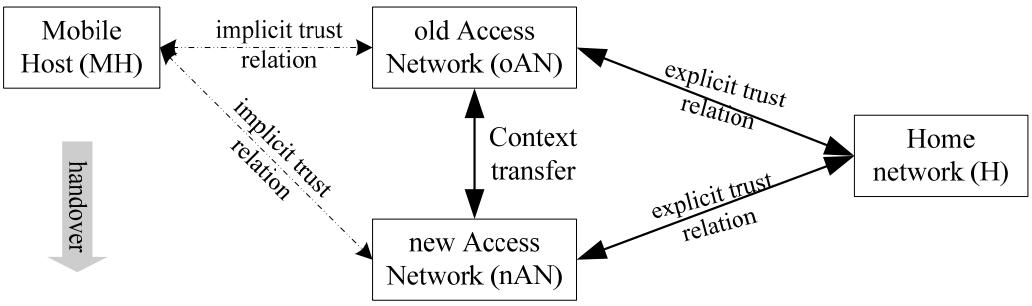


Figure 6.1 A trust model for security analysis of handover

The security of 3G UMTS has been developed to keep maximum compatibility with the current GSM security architecture. Mutual authentication is achieved by showing knowledge of a security key *K* shared between a mobile user and the Authentication Centre (AuC) in its home network. Using the Authentication and Key Agreement (AKA) protocol [93], the AuC generates and transfers a set of security credentials, known as Authentication Vector (AV) of a mobile user to a visited network. With the AV, the visited network performs mutual authentication with the mobile user as described in Sec. 3.2.2.

In IEEE 802.11, a new standard IEEE 802.11i [42] has been developed to strength its security. IEEE 802.11i enhances key management and encryption algorithms by incorporating IEEE 802.1X [47], a port-based network control mechanism. IEEE 802.1X employs the challenge-response Extensible Authentication Protocol (EAP) [18] to provide a variety of authentication mechanisms. The security mechanisms implemented within different wireless networks are limited to their particular architectures.

To integrate WLAN with 3GPP network, 3rd Generation Partnership Project (3GPP) has defined two new mechanisms in *TS 33.234* [75]: Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA) and Extensible Authentication Protocol Method for GSM Subscriber Identity (EAP-SIM). The full authentication methods defined in these protocols need at lease three round trips with home AAA server and AuC. Thus, a lightweight process, the EAP-AKA fast re-authentication method has been proposed. It re-uses keys generated from the previous authentication process to save processing time. However, with the fast re-authentication, at lease, two round trips are still needed to authenticate and authorise the mobile user

and generate session keys [74]. The delay introduced by the authentication procedure adds to the handover latency and consequently affects the ongoing communications.

Multimedia communications such as VoIP is very sensitive to handover delay. For this reason, the AAA related round trips between a mobile user and its home network in a handover should be reduced as many as possible. The EAP pre-authentication method specified by IETF in [94] has been proposed for such a purpose. In the EAP pre-authentication, the authentication for a target authenticator is performed while a mobile's session is still in progress via the serving network. The goal of the pre-authentication is to avoid AAA signalling for EAP when or soon after the mobile moves [94]. Accordingly, two approaches are possible: 1) pre-authenticate a mobile user directly to a target network; 2) pre-authenticate the mobile user indirectly via its serving network. The direct pre-authentication approach heavily relies on the simultaneous use of multiple interfaces on a mobile device [95], and as the support from the mobile's AAA server so as to allow registration of multiple IP addresses. This solution increases the mobile's power consumption and the complexity of deployment. The second approach makes use of the secure channel between the serving network and the candidate networks to transfer pre-authentication messages. The Inter-Access Point Protocol (IAPP), also known as IEEE 802.11F [96] has been specified to allow the transfer of security context information between two 802.11 APs within the same Distribution System (DS). The IAPP is commonly referred to as a layer 2 (L2) context transfer protocol. The Context Transfer Protocol (CTP) specified in RFC 4067 [97] supports context transfers over various L2 access technologies at layer 3 (L3). However, these L2 and L3 solutions rely on a trust relationship being established between the old AN (oAN) and the new AN (nAN) as shown in Figure 6.1 before a context transfer can be made. This greatly limits their applicability in a multi-operator environment, where the adjacent networks may belong to different network operators.

This chapter mainly addresses security for handover between non trust-associated domains. Here, domain is referred as an administrative domain which has a single Access, Authorisation, and Accounting (AAA) entity for authenticating and authorising its mobile subscribers for accessing network resources. A proxy-based authentication localisation scheme is proposed. It includes two specified phases: *fast authentication*

ticket generation phase and *fast authentication phase*. A third-party entity called AAA proxy is introduced to act as a hub for bridging trust relationships between networks. The trust relationship between the AAA proxy and each network is based on a pre-shared key. When a mobile user hands over to a network, the mutual authentication required for handover is localised at the associated AAA proxy rather than resorting to the mobile's home network. Using appropriate encryption and Mobile-Controlled Handover (MCHO), a mobile user can exert full control over the keying materials for fast authentication (e.g. fast authentication ticket) to be disclosed to only the target network. This effectively avoids the security threats such as Denial of Service (DoS) and masquerading attacks that are specified in other proposals for fast handover [94, 97]. The proposed scheme is to be implemented on the mobile terminal and its home AAA server, without any changes made to access routers that may have a large base of installation. The scheme can be deployed in a cost-effective manner.

The rest of this chapter is organised as follows. The current fast authentication solutions for handover are investigated in Sec. 6.2. In Sec. 6.3, a new fast authentication scheme that addresses authentication for handover between non trust-associated domains is presented. Sec. 6.4 analyses the security of the proposed scheme. The practical implementation of the proposed scheme is discussed in Sec. 6.5. Finally, this chapter is finished with conclusions in Sec. 6.6.

6.2 Related Work

In the direct pre-authentication approach [94], the long authentication delay in a handover is avoided by making use of the secondary network interface on a mobile handset, which can enable simultaneous handover to next point of attachment. The performance improvement is provided by the enhanced capability of the mobile handset (using multiple network interfaces) rather than the optimised authentication mechanism. Therefore, this approach is inapplicable to all handover cases. The current fast authentication solutions [69-71, 96-100] usually focus on the indirect pre-authentication approach as described in Sec. 6.1. A classification of fast authentication approaches can be found in Figure 6.2.

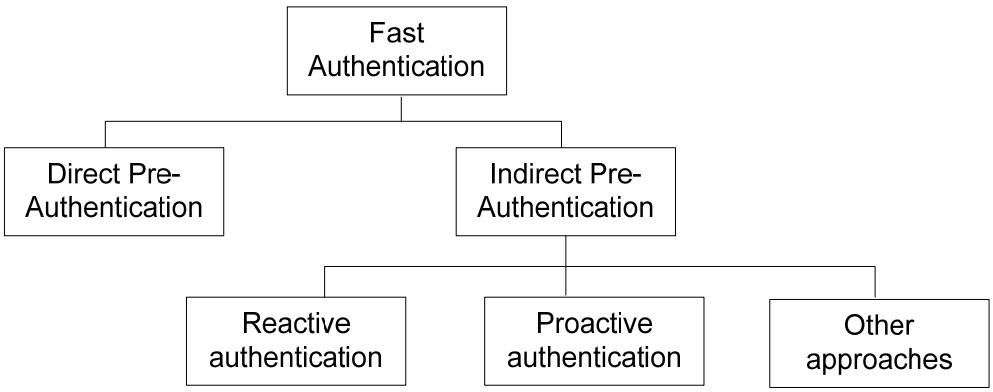


Figure 6.2 Classification of fast authentication approaches

The IAPP protocol [96] defines a standard method to transfer context information between two 802.11 APs when an intra-domain handover takes place. It relies on secure communications between two APs at layer 2, which have to be under the control of the same Extended Service Set (ESS) [96]. To support handover between different ESSs, *Bargh et al.* proposed in [98] that IAPP messages can be encapsulated using the CTP protocol [97]. The IAPP-CTP combined approach [98] requires secure communications between Access Routers (AR) at L3. *Sethom et al.* proposed a distributed architecture with a Location Server (LS) for key derivation [70]. The proposed architecture supports pre-authentication to be performed through the serving AR after a mobile user has determined its next point of attachment. In all these solutions, handover authentication is usually triggered in a reactive manner as shown in Figure 6.2, which means context transfer for fast authentication is made after initiating a handover.

The indirect pre-authentication can be performed in a proactive manner as shown in Figure 6.2. In the proactive authentication approach, fast authentication is achieved by pre-distributing the keying materials to the target network prior to conducting a handover to it. In [69], *Mishra et al.* defined a new data structure – *Neighbour Graph*, which can dynamically identify and maintain the mobility topology of a network. The home AAA server learns the association pattern of a mobile user by observing its neighbour graph, and thus determines the candidate set of APs. Before the mobile user moves to next AP, the home AAA server generates Pairwise Master Keys (PMK), and pre-distributes the keys to the candidate APs. When any one of the candidate APs receives an authentication request for handover from the mobile user, the identity

verification can be performed locally using the pre-distributed credentials. This neighbour graph based proactive approach can be applied to handover within the same administrative domain [69]. Instead of pushing the pre-authentication credentials straight to APs, *Hong et al.* presented a hierarchical key management scheme in [71]. In the hierarchical key management scheme, a Local Master Key (LMK) is generated by the home AAA server, and pre-distributed to a local authentication server for managing pre-authentication for intra-domain handover. However, it is still required that the home AAA server plays a role in providing the necessary LMK to the new local authentication server during an inter-domain handover.

Apart from the reactive and proactive authentication solutions, other approaches for fast authentication have been studied in the literature. In an attempt to localise authentication in the roaming across WLANs, *Long et al.* utilised the public key certificate structure to establish trust relationships between each pair of operators [99]. The public key certificate based authentication needs every network to store *(n-1)* public certificates of its own, and *(n-1)* public keys of other networks as discussed in [99]. Consequently, it may not be scalable when a large number of networks get involved. The Seamless Authentication Protocol (SAP) [100] is another operator-shared-key based scheme for facilitating fast authentication. It supports the sharing of a SAP master key among different AAA servers. Temporary security keys are derived from this master key for local identity verification at AP level. The SAP's scalability can be partially improved by utilising the group-based key update [100]. However, the SAP approach implicitly requires a trust relation between two domains in an inter-domain handover, which may not always be the case.

6.3 Proxy-Based Authentication Localisation

As discussed in Sec. 6.2, current fast authentication solutions for handover are based on the same assumption that there has to be a secure channel between two points of attachment involved in a handover for transferring security context. They are applicable for handovers within the same domain or two independent domains of the same operator. The handover between two separated domains belonging to different operators can be supported when a trust relationship exists between them.

To date, there has been no fast authentication solution specifically designed for handover taking place between two network domains without a trust relation (non trust associated). Current authentication specifications such as 3GPP AKA [93] and EAP-AKA [74] rely on the home AAA server for identity verification during a handover. This inevitably results in long signalling delay in a handover because several round trips between a mobile user and its home AAA server are required for exchanging AAA requests/responses.

This section presents a Proxy-Based Authentication Localisation (PBAL) scheme for handover between non trust associated network domains. The PBAL scheme provides a secure and controllable means of relaying the authentication authority of an AAA server to other AAA proxies, which can then process the authentication requests from its subscribers locally.

6.3.1 A Trust Association Model for the PBAL

A new entity called Fast AAA Proxy (FAP) is introduced to localise authentication in the proposed PBAL scheme. The FAP processes the AAA request from a mobile user and performs the identity verification on behalf of its home AAA server in a handover. As discussed early, to meet the trust relation requirements, the FAP needs to establish trust relationships with both its represented AAA server (HAAA) and the attached Access Network (AN), the latter of which holds a trust relationship with the HAAA for serving its mobile subscribers. The FAP thus acts as a third party proxy that bridges the trust relationship between the HAAA server and the AN.

From the perspective of a mobile user, the FAP acts as a local authentication authority on behalf of its HAAA. The FAP shares a pairwise key (K_{hp}) with the MH's HAAA. The key K_{hp} is used to establish a one-on-one trust association between the HAAA and each FAP. The HAAA may establish trust associations with a number of FAPs, each of which is associated with a group of ANs as shown in Figure 6.3.

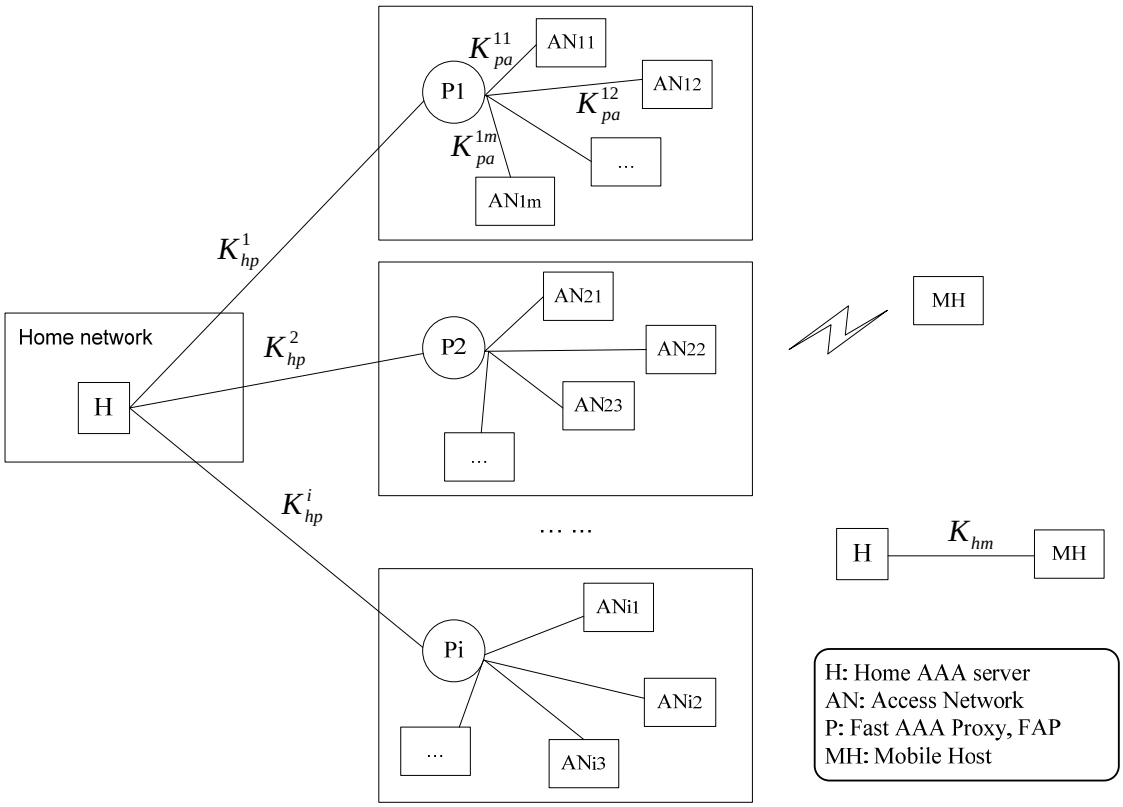


Figure 6.3 An overview of trust associations for fast authentication

The trust association between a FAP and each associated AN is enabled using a pairwise key K_{pa}. The pairwise key K_{pa} is mutually agreed upon by the owner of the FAP and the operator of the AN. By doing do, an AN can confirm that a FAP is legitimate for authenticating a certain group of mobile users. As an independent authority for taking authentication requests, a FAP may be associated with many access networks for fast authentication as shown in Figure 6.3. Meanwhile, an access network can establish trust associations with different FAPs through shared keys at the same time, since the mobile users with different realm portions of Network Access Identifiers (NAI) [64] may require different AAA routing paths. The case of an AN being associated with multiple FAPs has not been shown in Figure 6.3 for the sake of simplicity.

The explicit trust association between a mobile user MH and its HAAA can be established through another key K_{hm}. Table 6.1 lists all the related keys for the PBAL trust association model for fast authentication. The three pre-established keys K_{hp}, K_{pa},

K_{hm} are used to build the two types of explicit trust associations required in the handover attachment. Although their updates can be agreed upon by the involved parties, these keys are mostly regarded as "permanent" in contrast to the two derived temporary keys: Local Authentication Key (LAK) and Pairwise Master Key (PMK). The LAK is a session-related key generated by the HAAA. It is employed by the FAP to perform mutual authentication with the MH and establish the necessary security context associated with the AN. The LAK is used to derive the PMK, which is specified in IEEE 802.11i [42] for wireless link protection.

Table 6.1 A list of the PBAL trust association model related keys

Abbr.	Description	Temporary key
K_{hp}	Pairwise key shared between HAAA and FAP	N
K_{pa}	Pairwise key shared between FAP and AN	N
K_{hm}	Pre-Shared Key (PSK) between HAAA and MH	N
LAK	Local authentication key shared among HAAA, FAP and MH	Y
PMK	Pairwise master key shared among FAP, AN and MH	Y

6.3.2 Fast Authentication Ticket Generation Method

After completing the attachment to a network, the MH can request a Fast Authentication Ticket (FAT) for every nearby access network from its HAAA. The nearby access networks can be either pre-determined using the pre-stored network location information, or determined using the network trust information retrieval scheme presented in Chapter 4. The location based approach requires that a network has accurate location information of its neighbouring networks. Since all the location data have to be pre-loaded, this approach is static and can not cope with changes to networks. With the method proposed in Chapter 4, the serving Access Network (sAN) that has obtained its neighbour network trust pattern can provide the MH's HAAA with an identifier list of its neighbouring access networks.

Note that the determination of the access networks in vicinity is outside the scope of this discussion. It is assumed here that the sAN has the knowledge of its surrounding access

networks, and can provide the neighbour network identifier list to the HAAA along with the MH's request for the FAT.

Following the successful attachment to the serving network sAN, the MH sends a *FAT Request* with an encrypted mobile nonce ($E_{K_{hm}}(N_m)$) to its HAAA to request the FATs via the sAN. The sAN provides an identifier list of its Neighbour Access Networks (NAN) $NAN\_ID_s[AN_1, AN_2, ..., AN_t]$ that includes the identifier of every nearby AN (AN_t). The sAN forwards the FAT request along with the identifier list NAN_ID_s to the MH's HAAA.

Upon receiving the FAT request, the HAAA generates a server nonce (N_s), which will be encrypted in the generated FAT along with other security credentials as shown in Equation 6.1. The server nonce N_s will later be provided to the FAP, and used as a challenge to verify the MH's identity. Since N_s is generated for a specific AN, the HAAA needs to find the associated FAP according to the AN's identifier. The HAAA issues a FAT for every AN included in the neighbour access network identifier list NAN_ID_s:

$$FAT = E_{K_{hp}}(ID_m, N_s, PID_p, SQN, LAK, MAC) \qquad \text{(Equation 6.1)}$$

Each FAT encloses the information about the MH's identity (ID_m), the server nonce (N_s) issued by the HAAA, the Pseudonym Identity (PID) of the FAP (PID_p), the sequence number (SQN), the Local Authentication Key (LAK) for the FAP and the Message Authentication Code (MAC) to be used for verifying the MH's identify.

For security concerns, the server nonce N_s issued will be varied for different FAPs so that different access networks would have different security contexts. The FAP's PID_p is derived using the secret splitting method described in [101]:

$$PID_p = h(K_{hp} \,||\, N_s) \oplus ID_p \oplus N_s \qquad \text{(Equation 6.2)}$$

where *h* is a public strong one-way hash function, and ⊕ is bitwise XOR operation. And ID_p is the FAP's identity.

A sequence number *SQN* is generated to keep the freshness of the FAT, and is incremented by 1 every time a FAT is issued by the HAAA. To be compatible with the authentication framework IEEE 802.11i [42], a Pre-Shared Key (PSK) between the MH and the HAAA is required to derive a high order key Pairwise Master Key (PMK) for protecting wireless link. Acting as the PSK, the key K_{hm} is employed to derive session keys for protecting wireless link between the MH and the AN.

A session-wide secret, the local authentication key LAK is produced and included in the FAT for performing localised authentication associated with the new session. The LAK will be used by the FAP and the MH to conduct mutual authentication, and derive the required PMK. The HAAA generates a LAK for each FAP associated with the nominated ANs. The FAP receives the LAK through the FAT passed by the MH in a handover, and uses this LAK to establish the security context associated with the new AN. The LAK is derived from the current PMK that serves the sAN as follows:

$$LAK = prf(K_{hm}, PMK \,||\, N_m) \qquad \text{(Equation 6.3)}$$

where *prf* denotes a pseudo-random function that can be constructed using the methods described in [102]. The PMK represents the session key associated with the sAN. As such, the LAK is a secret totally determined by the trust association between the MH and its HAAA.

Like the XRES provided to a Visitor Location Register (VLR) for performing authentication in the 3GPP's AKA [93], an Expected MAC (XMAC) will be generated by the HAAA, and provided to the FAP as the expected response when the FAP begins to verify the MH's identity. The usage of MAC for identity verification is seen in the Pre-Shared Key Extensible Authentication Protocol (EAP-PSK) (RFC 4764 [103]). Here, the XMAC is assumed to be a function of the two nounces: N_m and N_s, and the PSK K_{hm}:

$$XMAC = f(K_{hm}, ID_m \,||\, ID_s \,||\, N_m \,||\, N_s) \qquad \text{(Equation 6.4)}$$

where f is a message authentication function that is only known to the MH and its HAAA. The HAAA encrypts the FAT contents as listed in Equation 6.1 using the secret key K_{hp} shared with the corresponding FAP. After every AN on the neighbour access network list has been produced a FAT, the HAAA builds up an encrypted FAT Vector (FATV):

$$E_{K_{hm}}(FATV[FAT_1, ID_{p1}, FAT_2, ID_{p2}, ..., FAT_t, ID_{pt}])$$

which includes every FAT and the corresponding FAP's identity ID_p. The FATV is protected by the key K_{hm} shared between the MH and its HAAA. Then, the encrypted FATV is included in a *FAT Response* message, and sent back to the MH as shown in Figure 6.4. The encryption using the PSK K_{hm} makes sure that the FATV would not be disclosed to other parties (including the sAN), and only the MH can view it. Although the MH can decrypt the encrypted FATV, it is unable to learn the contents of individual FAP. Each FAT is encrypted with a key (K_{hp}) that is only known to the HAAA and the FAP that it is generated for. Neither the MH nor the sAN can modify the received FAT's contents. Once the FAT response is received, the MH increments its SQN_m counter by 1.

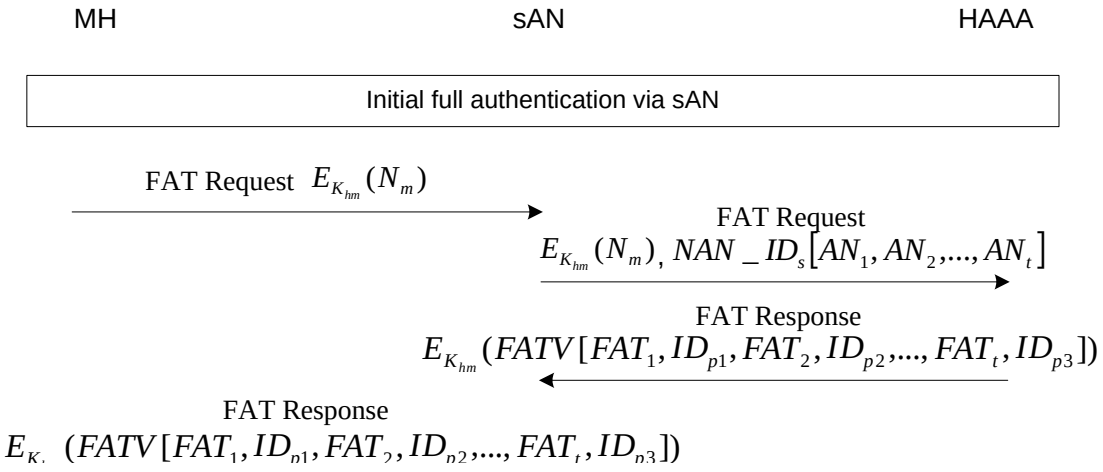


Figure 6.4 Fast authentication ticket generation procedure

The consequence of the fast authentication ticket generation is that a FAT will be produced for every nearby access network before a handover execution. Note that only

the FAT for the selected access network (determined by the handover decision algorithm in the mobile terminal) would be used by the MH for fast authentication.

6.3.3 Fast Authentication Method

To handover to a new Access Network (nAN), the MH sends out a *Fast Authentication Request* including the corresponding FAT and its SQN_m to the nAN for requesting local authentication. The MH can influence the AAA routing by modifying the realm portion of its identity [64]. The new identity realm portion can be determined according to the identity ID_p of the FAP associated with the nAN. As a result, the fast authentication request can be redirected to the specified FAP instead of the HAAA. If the modification of the realm portion is not supported by the AAA protocol, the MH may provide the nAN with the FAP's identity ID_p so that the nAN can route the authentication request accordingly. The FAP decrypts the received FAT $E_{K_{hp}}(ID_m, N_s, PID_p, SQN, LAK, MAC)$ using the key K_{hp} shared with the MH's HAAA and retrieves its contents. However, further operations will proceed only if the FAT's *SQN* is equal to the SQN_m provided by the MH. Next, the FAP uses Equation 6.5 to derive its identity from the pseudo one PID_p provided by the HAAA:

$$ID_p = PID_p \oplus h(K_{hp} \,||\, N_s) \oplus N_s \qquad \text{(Equation 6.5)}$$

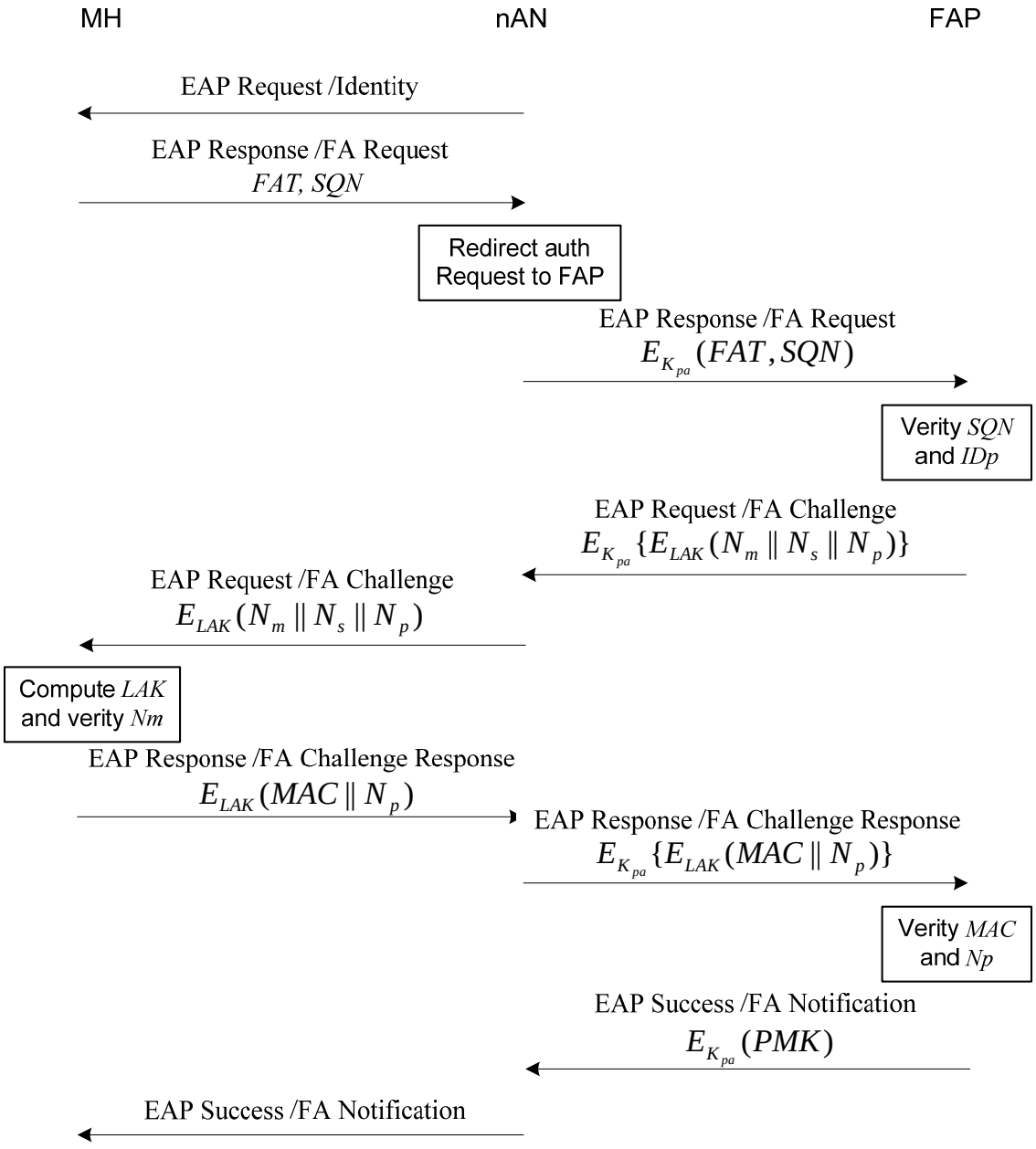


Figure 6.5 Fast authentication procedure

As shown in Figure 6.5, the FAP verifies whether the computed identity ID_p is equal to its real identity such that the identity of the MH's HAAA can be verified. If it passes verification, the FAP issues a proxy nonce N_p, which will contribute to the generation of a new PMK. With all the credentials from the FAT, the FAP encrypts the nounces N_m, N_s and N_p that go as a challenge using the LAK, and returns the

$E_{LAK}(N_m \| N_s \| N_p)$ in a *Fast Authentication Response* to the MH. The MH can compute the LAK using Equation 6.3 based on its knowledge of the required credentials. After receiving the fast authentication response from the FAP, the MH is able to decrypt $E_{LAK}(N_m \| N_s \| N_p)$ with the LAK computed by itself. The N_m obtained should be the same as the original one issued previously for requesting the FAT. By doing so, the MH verifies the authenticity of the visited network nAN.

After verifying the nAN's identity, the MH gets itself authenticated to the nAN. The MH uses the challenge N_s provided by its HAAA, K_{hm}, ID_m, ID_s, and N_m to compute a MAC value using Equation 6.4. The MAC along with N_p are encrypted in $E_{LAK}(MAC \| N_p)$ and delivered to the FAP as a challenge response. With decryption, the FAP verifies the MH's identity by comparing the received MAC with the XMAC provided by the HAAA. If they are equal, the following PMK is generated, and sent to the nAN to build the necessary security context.

$$PMK = prf(LAK, N_m \| N_s \| N_p) \qquad \text{(Equation 6.6)}$$

A fast authentication response is returned to the MH to notify the authentication result. Then, the MH can build the corresponding PMK for communications with the nAN.

6.3.4 Session Key Renewal

In the PBAL, a session key renewal method is proposed so that a mobile user may renew its session key (PMK) that is used to protect the wireless link according to IEEE 802.11i [42]. Change of session key would reduce the risk that the mobile user uses a compromised session key to communicate with an access network.

In the PBAL session key renewal method, the session key renewal can be initiated by either the sAN or the MH. Figure 6.6 shows how the session key PMK being shared between the MH and the sAN can be renewed as requested by the MH. The MH generates a new mobile nonce N_m', and sends a *Key Renew Request* including this new nonce N_m' along with the original server nonce N_s (previously issued by the HAAA) to the FAP to initiate the session key renewal process. N_m' and N_s are encrypted with

the local authentication key LAK shared between the MH and the FAP. The AN acts just as a pass-through, and can not view or modify the two passed nonces. After receiving the key renew request, the AN forwards it with the encrypted $E_{LAK}(N_m' \| N_s)$ to the FAP. $E_{LAK}(N_m' \| N_s)$ is decrypted at the FAP later using the cached LAK. The FAP verifies whether N_s provided by the MH is the same as what has been included in the original FAT. If the MH could pass this request origination check, the FAP generates a new proxy nonce N_p', and computes the new PMK (*PMK'*) as follows:

$$PMK' = prf(LAK, PMK \| N_m' \| N_p') \qquad \text{(Equation 6.7)}$$

The new *PMK'* along with the encrypted proxy nonce $E_{LAK}(N_p')$ will be delivered to the AN through the FAP-sAN connection (protected by K_{pa} as shown in Figure 6.6). The AN keeps the *PMK'* for renewing the related session key, and forwards $E_{LAK}(N_p')$ to the MH. With the decryption, the MH obtains the new proxy nonce N_p' from $E_{LAK}(N_p')$, and can derive the new pairwise master key *PMK'* according to Equation 6.7.

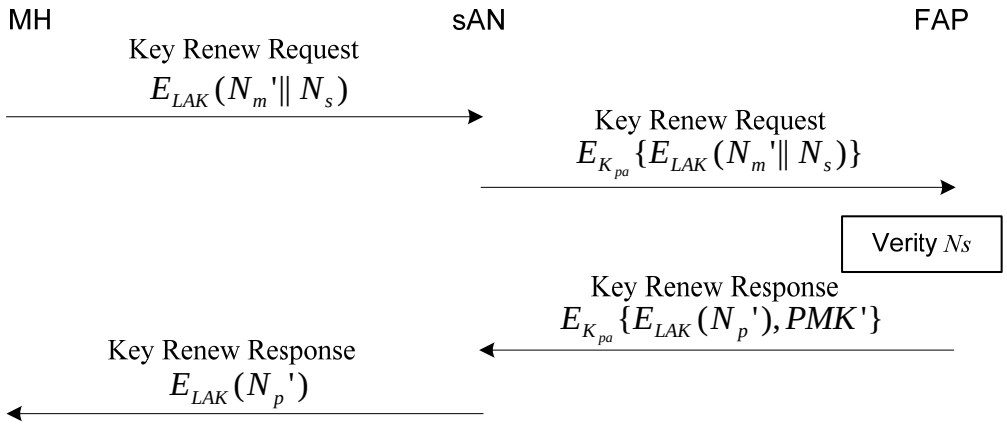


Figure 6.6 Session key renewal procedure initiated by MH

The sAN can start the session key renewal procedure. This is started by sending a *Key Renew Request* to the FAP. Then, the FAP generates the new proxy nonce N_p', and sends this nonce N_p' along with the received original mobile nonce N_m to the MH for verification by the MH. The N_p' and N_m are encrypted using the LAK so that they can

not be overheard by the sAN. The sAN forwards these session key renew credentials to the MH. If N_m that is included in the key renew response is valid, the MH generates another mobile nonce N_m', and establishes the new session key *PMK'* according to Equation 6.7. The newly issued nonce N_m' will be delivered to the FAP as shown in Figure 6.7. Based on the two new nonces N_m' and N_p', the FAP can derive the corresponding *PMK'*. Later on, the derived *PMK'* is provided to the AN for updating the session key.

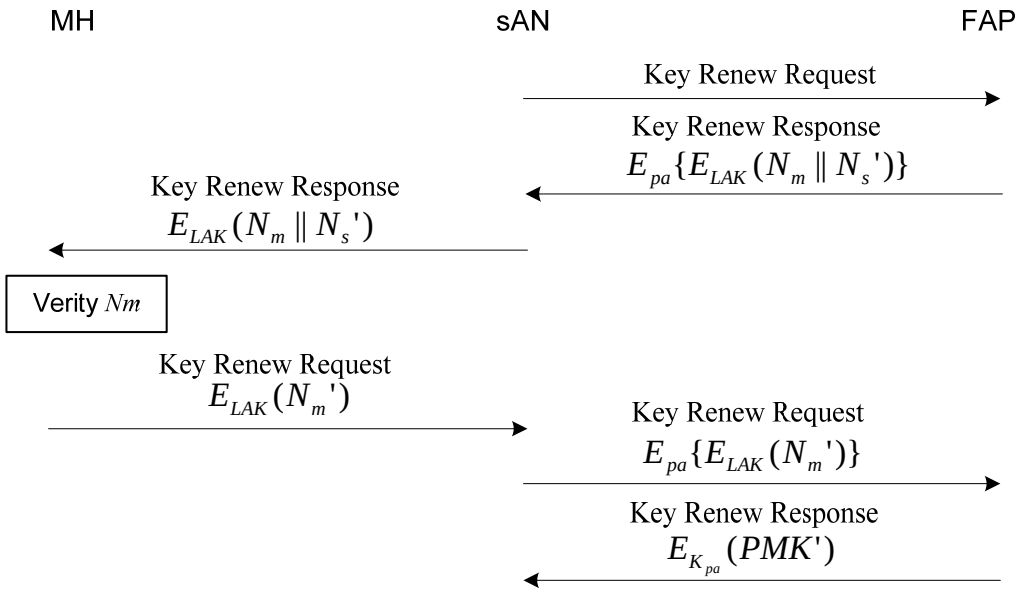


Figure 6.7 Session key renewal procedure initiated by AN

In this thesis, it is assumed that the communications between a MH and an AN are always protected by other link layer security mechanisms. For example, IEEE 802.11i [42] defines how the PMK can be utilised to ensure a secure association with an access point.

6.4 Security Analysis for the PBAL

In the proposed PBAL scheme, the authentication authority is temporarily relayed to a third-party entity FAP when a mobile subscriber roams outside the territory of its home network. The FAP plays a similar role as the Visitor Location Register (VLR) of 3GPP in verifying a mobile user's identity using the home-supplied authentication vectors, as

specified in 3GPP Authentication and Key Agreement (AKA) protocol [93]. However, instead of explicitly requesting authentication vectors from the home network, the FAP obtains security credentials through the fast authentication ticket FAT presented by the mobile user. The release of the FAT by the party who makes handover decision can ensure that sensitive security information can only be disclosed to the selected network. In this section, the security of the PBAL is analysed. First, it shows how mutual authentication in a handover is provided. Then, the study analyses the security of the proposed scheme against some known attacks such as replay attack.

6.4.1 Mutual Authentication

The PBAL provides mutual authentication between a mobile user and FAP. The MH sends a fast authentication request enclosing the related FAT for local authentication at the FAP upon attaching to the selected network. With decryption, the FAP extracts N_m from $FAT(ID_m, N_s, PID_p, SQN, LAK, MAC)$, and encloses it in $E_{LAK}(N_m \| N_s \| N_p)$ sent to the MH. Since N_m was originally issued by the MH and is only known to its home AAA server, only the party that has a trust relationship (K_{hp}) with the home AAA server can demonstrate it. Thus, by checking the correctness of N_m, the MH can verify whether the FAP is a representative of its home AAA server and has been authorised for local authentication.

To get itself authenticated by the FAP, the MH computes the MAC, which is a function of K_{hm}, N_m, N_s. N_m and N_s are two random numbers, which are chosen by the MH and its home AAA server respectively. The N_s encrypted in $E_{LAK}(N_m \| N_s \| N_p)$ was previously sent by the FAP as a challenge. Before being able to retrieve N_s from $E_{LAK}(N_m \| N_s \| N_p)$, the MH must correctly compute the local authentication key LAK that is a function of the preshared key K_{hm} and the pairwise master key PMK associated with the serving AN. With decryption, the MH gets N_s, and computes the MAC_m using its message authentication function *f* according to Equation 6.4. By comparing MAC_m with MAC_h that is provided by the home AAA server, the FAP verifies whether the MH is a legitimate subscriber of the home AAA server.

6.4.2 Security against Replay Attack

Replay attack involves the capture of over the air information and the subsequent retransmission to trick the receiver into unauthorised operations. During the fast authentication procedure of the PBAL, fast authentication requests/responses are exchanged between the MH and the AN via unprotected wireless link before the PMK is established (shown in Figure 6.5). These messages are at risk of being eavesdropped by an adversary. The adversary may mock up a fast authentication request with the same FAT and send it later to gain the right of using the AN. However, such an attempt can be effectively prevented in the PBAL.

Assume that two messages: “FA Request (with FAT)” and “FA Challenge Response” have been eavesdropped by an adversary, when a MH was attempting to attach to the new AN (nAN). The adversary can impersonate the MH, and try to establish a connection with the nAN. To impersonate the MH, the adversary sends the copied FA Request (with the identical FAT) to the nAN, and waits for the challenge from the FAP for further verification. According to the fast authentication procedure of Figure 6.5, the FAP returns a FA challenge $E_{LAK}(N_m \| N_s \| N_p')$ to the adversary. N_p' is a random number generated by the FAP every time it receives a FA request. Obviously, the newly generated N_p' is different from the N_p of the previous session that has been captured by the adversary. The adversary will fail the identity verification at the FAP with the expired N_p. Therefore, using the old credential $E_{LAK}(MAC \| N_p)$ as the FA challenge response can not pass the verification on N_p' to be conducted at the FAP.

In another case, an adversary may impersonate the nAN, and fool the MH to attach to the adversary. Assume that the adversary entices the MH to attach to it soon after eavesdropping the FA challenge $E_{LAK}(N_m \| N_s \| N_p)$ previously sent by the nAN. Such a replay attack can be easily avoided in the PBAL. The MH required both the correct *LAK* and N_m for performing the FAP’s identity verification. The *LAK* is a function of the session key PMK, while N_m is a random number that is valid for a specific FAT. The MH produces a new N_m' for retrieving the FAT after every successful handover. When the MH receives the concocted fast authentication challenge

$E_{LAK}(N_m \| N_s \| N_p)$ from the adversary, the verification on the out-of-date N_m at the MH using the new *LAK'* and N_m' will fail.

6.4.3 Impact of Network Corruption

The traffic on the radio link is protected by the keys derived from the PMK in use, which is shared between the MH and the AN. The corruption of an AN may result in an intended disclosure of PMK. This would make it possible for an adversary to eavesdrop on the radio link, and track the sensitive information to be delivered. Thus, the access security can be affected.

In the PBAL, the nonce N_m generated for requesting the FAT is encrypted using the key K_{hm} shared between the MH and its HAAA as explained in Table 6.1. The FAT vector FATV returned by the HAAA is protected with the key K_{hp} known only to the HAAA and the FAP. Although the traffic on the radio link may be overheard, the disclosure of $E_{K_{hm}}(N_m)$ and the encrypted contents of the FAT will not cause a security threat. No parties other than the ones holding the keys between the HAAA and the FAP can correctly decrypt the enclosed security information used for fast authentication. The MH is unaware of the contents of the received FAT. Therefore, the risk of using a corrupted PMK by an adversary is limited to performing the attacks aforementioned in Sec. 6.4.2.

Although the corruption of PMK does not affect the PBAL fast authentication, it is still considered as a major security hole. In the PBAL, the session key renewal can be periodically initiated by either a mobile user or the FAP as described in Sec. 6.3.4. Access network plays a role of a pass-through by forwarding the key renewal request/response between the mobile user and the FAP. This minimises the impact of a corrupted AN on wireless communications. Moreover, the identity of the key renewal request originator is always verified before proceeding to produce a new PMK as illustrated in Figure 6.6 and Figure 6.7. Such origination verification can make sure that the corruption of the serving PMK will not result in the compromise of a renewed PMK to be used in the future.

6.5 Practical Implementations

The proposed PBAL scheme can be implemented on networks without making any changes to access routers. This is achieved by setting up a third-party fast AAA proxy for localising AAA requests. From the perspective of a mobile user, it interacts with the FAP through an access network during a handover, instead of directly communicating with its home AAA server for authentication. In the PBAL, the access network acts as a pass-through [18] for AAA messages and does not have to understand the authentication method applied. The mobile user is in a position of redirecting its AAA request to the corresponding FAP in a handover. Such a redirection operation is supported by the EAP framework [18], on which various authentication methods can be applied. For example, the EAP-AKA defines that the supplicant (e.g. a mobile user) may modify the realm portion to influence the AAA routing [74]. Keeping the PBAL transparent to access networks is a clear advantage of implementing the PBAL for fast authentication. This provides a large number of network operators with a choice of upgrading to a fast authentication scheme without reflashing old access equipments.

The trust association between the home AAA server and a fast AAA proxy can be set up using a pairwise key K_{hp} as part of a roaming agreement between the two operators. Elements of such a trust association may include cryptographic keys, negotiated cipher suites and other parameters. AAA protocols such as Remote Authentication Dial In User Service (RADIUS, RFC 2865 [49]) and Diameter (RFC 4072 [65]) can be used to negotiate the maximum key lifetime between the home AAA server and a fast AAA proxy. The same protocol can be used to manage the trust association K_{pa} between the fast AAA proxy and each access network. The pre-shared keys for setting up trust associations can be statically configured or dynamically updated in a secure manner. From the perspective of a network operator, its trust association established with a third-party fast AAA proxy has twofold usage: 1) transfers its authentication authorities to a third-party entity for facilitating fast handover of its own subscribers; 2) obtains localised authentication services when processing AAA requests from other roaming mobile users.

In the PBAL, a fast AAA proxy can be integrated on an ordinary AAA broker in the roaming broker infrastructure [53, 87]. This avoids building new separated entities for

deploying the PBAL scheme. With the integration, the PBAL can reuse the cryptographic keys that are used to encrypt and authenticate data exchanged between an AAA server and an AAA proxy [53]. This approach sees a smooth migration into a fast authentication scheme from the current AAA infrastructure. Alternatively, independent AAA proxies can be deployed for providing fast authentication. Considering the dedicated nature of these AAA proxies, a less number of AAA proxies may be required to achieve the same capacity.

6.6 Conclusion

The past studies on fast authentication have focused on handover either within the same network domain, or between two network domains sharing a trust association. To localise an authentication, their proposed solutions require that a secure channel between two points of attachment in a handover should be available for transferring security context information. However, as the coexistence of multiple network operators is anticipated in the NG heterogeneous wireless networks, more and more handover operations may be performed between network domains without a trust association. For such kind of handover, the authentication request is always delivered to the mobile's home AAA server, because the current fast authentication solutions will not work.

The proxy-based authentication localisation scheme PBAL is presented to address fast authentication in a handover taking place between two network domains without a trust association. The proposed PBAL scheme can provide a roaming mobile user with localised authentication in a handover. The PBAL does not require any communications between two points of attachment (sAN and nAN) before, during and after a handover. By relaying authentication authority to a third-party proxy, the PBAL avoids resorting to a mobile user's home AAA server for identity verification in a handover. As a result, handover signalling delay is greatly reduced in a multi-operator environment. It was proven that the PBAL supports mutual authentication and security protection against some known attacks such as replay attack and network corruption. It provides some additional security features like session key renewal, which makes sure that vulnerable wireless link can be better protected. In the PBAL, both authentication localisation and security enhancement can be conducted in a cost effective manner, since a few round trips are needed. Generally, the PBAL can support fast and localised authentication without compromising security features.

Chapter 7

MULTI-INTERFACE MOBILE MODEL FOR MEDIA INDEPENDENT HANDOVER

7.1 Problem Definition

In the NG heterogeneous wireless networks, mobile users are expected to switch between different wireless networks so as to maintain or optimise their services. To access heterogeneous wireless networks, a mobile user needs to deal with different access technologies. This requires that multiple network interfaces to be equipped on a mobile terminal which are employed for accessing disparate networking media. Types of network interfaces required by a mobile terminal are determined by the integrated network architecture. For example, being equipped with both 3G UMTS and IEEE 802.11 WLAN network interfaces is the prerequisite for a mobile terminal to function in an UMTS-WLAN interworking scenario of Figure 2.1. From the perspective of a mobile user, the multi-interface architecture provides an effective means of dealing heterogenous access technologies.

Multiple heterogeneous wireless interfaces can be integrated in a variety of ways on a mobile terminal. Due to the importance of keeping the independence of every network interface, a multi-interface mobile terminal may have additional requirements on its network interfaces in regards to network protocol, handover processing and system architecture compared with single-interface mobile terminal. During a handover, multiple network interfaces of a mobile terminal have to function in a collaborative manner to enable seamless access to networks. This requires the upward support to upper layer user applications and the downward support on interfacing with disparate

transmission media on a mobile terminal. On a multi-interface terminal, some common features that are required for seamless handover are as follows:

- **Simultaneous communication**: refers to the capability of having multiple network interfaces carry data communications simultaneously.
- **Address management**: host mobility management has been addressed in both Mobile IPv4 [21] and Mobile IPv6 [104]. A mobile user can acquire multiple Care-of-Addresses (CoA) for accessing multiple visited domains. The related IP address assignment on each interface is referred as address management.
- **Traffic redirection**: the ongoing traffic of a mobile terminal is redirected from one network interface to another due to the upcoming handover.
- **Network selection**: is the process of selecting next Point of Attachment (POA) by implementing handover decision algorithms.

A variety of solutions on network selection, multihoming, routing, and transport protocol [105-109] had been proposed for the multi-interface mobile terminal. However, the study on the multi-interface architecture for seamless handover began in recent years. The solution on providing a mobile terminal with multiple network accesses may date back to a single network interface being used for accessing multiple networks [110, 111]. In *MultiNet* [110], the virtualisation of a WLAN adapter was proposed to multiplex a WLAN card across multiple WLAN networks. *SyncScan* [111] provided a low cost replacement by synchronising short listening periods at a mobile terminal with periodic transmission from each AP. These solutions had been proposed for homogeneous wireless networks.

To enable multiple network accesses, multiple network interfaces have to be either simultaneously used or alternatively used. When multiple network interfaces are used simultaneously, a mobile terminal needs several IP addresses at the same time, and thus requires the corresponding processing at upper network layers to serve user applications. In the second approach, a mobile terminal switches its services between multiple network interfaces instead of using all of them at the same time. Alternatively used multiple interfaces present many benefits to a mobile terminal including a reduction of

complexity of upper layer protocols. *F. André et al.* presented a common multi-interface architecture for IPv6 based mobile terminal in [112], but unfortunately with no implementation details. *C.-W. Ng et al.* discussed several issues of using multiple network interfaces, and proposed a tunnel re-establishment technique [113]. To date, the mobile terminal multi-interface architecture has not been a focus in the current research literature. To accommodate multiple heterogeneous interfaces on a mobile terminal, a new mobile terminal architecture has to be developed to integrate various resources and mechanisms provided by heterogeneous network interfaces. The new design has to take into account network protocols already in use and supports seamless handover network selection [84].

7.2 Background

7.2.1 Current Multi-Access Schemes

- **Multi-Interface Simultaneously Used (MISU)**

A mobile terminal may use its multiple network interfaces simultaneously to support soft handover [114] across heterogeneous wireless networks. Soft handover means a mobile user maintains at least one radio connection with a network at any time. During a soft handover, more than one network interface is allowed to carry traffic flow so as to be able to redirect ongoing user sessions seamlessly. Carrying multiple data streams through different network interfaces requires a mobile terminal being configured with multiple IP addresses, referred to as *multihoming* [115]. The multihoming along with other mechanisms such as simultaneous binding [116] provides basic function sets for the MISU based architecture. Additional functions required for the MISU need to be supported at both network and mobile terminal ends. Moreover, some particular processing at network and transport layers is necessary, because multiple data streams may be carried through multiple network domains simultaneously. In the dynamic network interface selection scheme [105], a number of network and transport protocols for host multihoming have been investigated. The network and transport protocols designed for multihoming can deal with access heterogeneities, and make these heterogeneities transparent to user applications. Although the MISU can guarantee reliable data transmission in a handover, it needs particular network/transport protocols

to work with. Considering the huge installed base of TCP/IP on mobile devices, this would need widespread support from networks.

- **Multi-Interface Alternatively Used (MIAU)**

Apart from the MISU, multiple network interfaces can be alternatively used to support hard handover [114] across heterogeneous wireless networks. Hard handover is also known as “break before make”. When multiple network interfaces are alternatively used, only one interface is active at any time and involved in real data transmission. With the MIAU, there is no need to deal with multiple data streams and therefore, the multihoming support is not required on a mobile terminal. Consequently, no special upper layer protocols are needed. From the perspective of user applications, the switching of network interfaces may result in fluctuation of data transmission quality and even a possible break of data stream. Obvious, the MIAU that employs single data stream incurs less traffic burden on a network compared with the MISU. However, the MIAU may result in longer handover delay because no redundant link would be available for continuing data transmission when the old data connection is lost. In the MIAU, only one interface is active at any time and the other interfaces act as back-up. Rather than relying on the multihoming protocols [105], the MIAU can remain compatible with standard upper layer protocols, such as Mobile IP, TCP.

- **Single Interface for Multiple Access (SIMA)**

In some cases, single network interface can be used for accessing multiple networks. This allows a mobile terminal with single network interface to connect to multiple network domains at the same time. To support multiple accesses, a network interface must be made multihomed so as to receive several network prefixes advertised by access routers. Single interface for multi-access is stimulated by various reasons such as soft handover and load balancing [117]. *Chandra et al.* stated in their *MultiNet* proposal [110] that single WLAN card can be multiplexed across multiple networks by applying an adaptive network hopping scheme. Activating WLAN card on different wireless channels at different time slots can thus enable the virtualisation of WLAN into several active virtual adapters. In another scheme known as *SyncScan* [111], the freedom of beacon sending time is explored to synchronise wireless node with the timing of AP

beacons. Basically, beacon periods of a group of access points are scheduled across channels so that a mobile user can synchronise its scanning with broadcasting of network beacons. SyncScan provides a means of continuously tracking signal strength from nearby APs while carrying on communication with the serving AP. However, both MultiNet and SyncScan need a mobile user's synchronisation with networks, and modification to network protocol stack at either mobile terminal or access points.

7.2.2 Media Independent Handover

IEEE 802.21 Media Independent Handover (MIH) standard [118] is a developing effort for enabling handover and interoperability between heterogeneous wireless networks. IEEE 802.21 provides link layer intelligence and other related network information to upper layers to optimise handover between heterogeneous network transmission media [119]. IEEE 802.21 supports cooperative use of mobile terminals and network infrastructure. It is proposed to provide an architecture that enables transparent service continuity when a mobile user switches between heterogeneous network interfaces. Moreover, a set of handover enabling functions within the mobility management protocol stacks of the network elements and the creation therein of a new entity called the MIH function [119].

MIH functional entities can reside at both mobile terminal and network ends, and communicate with each other. The Service Access Points (SAPs) are defined for lower layers and upper layers. Their services are used to facilitate seamless handover between heterogeneous wireless networks. Meanwhile, MIH provides the functionality for the exchange of information between MIH entities. In IEEE 802.21 [118], three types of services are proposed to facilitate information exchange for network discovery and selection in a handover. Media Independent Event Service (MIES) provides events notification in response to changes in state and transmission behaviour of data links. It supports local and remote events if the remote network is of the same media type. Media Independent Command Service (MICS) includes the commands sent from the upper layers to the lower layers as shown in the reference model of Figure 7.1. MICS commands are used to deliver the upper layer decisions to the lower layers, and thus control the behaviour of lower layers. Media Independent Information Service (MIIS)

provides a model and corresponding mechanisms in which MIH entities can gather network information to facilitate handover.

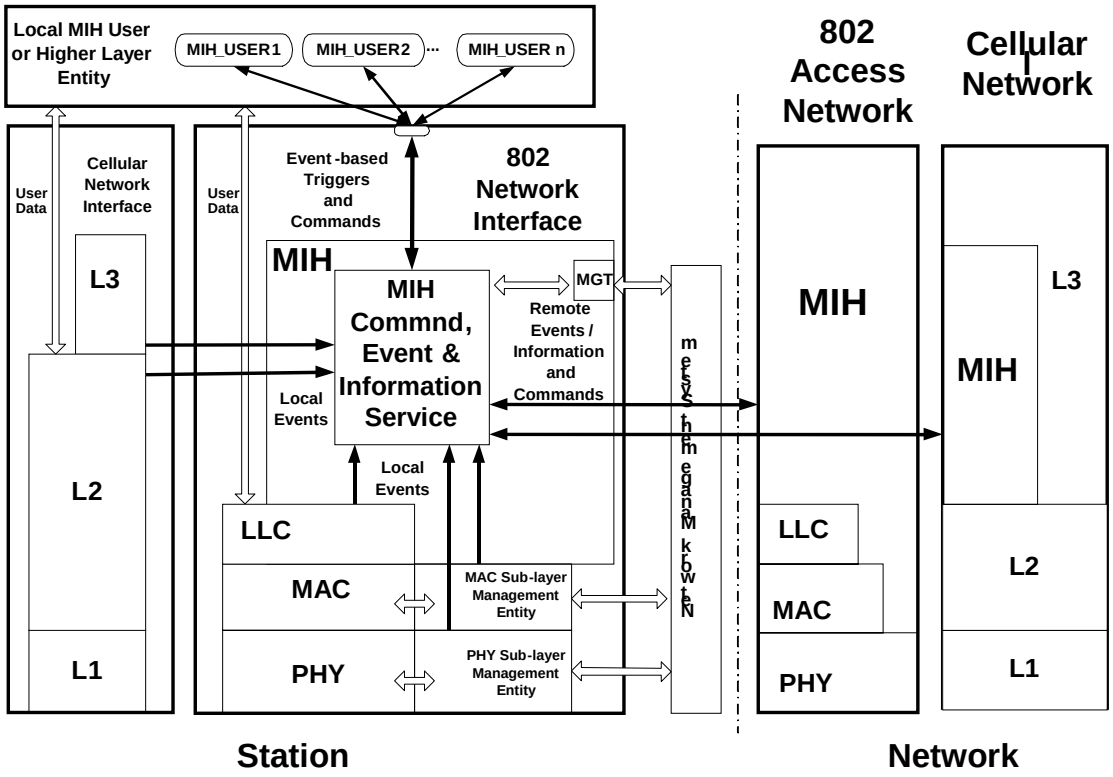


Figure 7.1 General Media Independent Handover reference model [119]

Under IEEE 802.21, a variety of access technologies such as 802.11, 802.16 and 3GPP can be accommodated on a single mobile terminal. Working with handover decision algorithms for heterogeneous wireless networks, IEEE 802.21 is able to provide flexibility in facilitating information retrieving, network discovering and handover triggering of mobile users.

7.3 A Multi-Interface Mobile Terminal Model

In Sec. 7.2.1, two multi-access approaches for accessing heterogeneous wireless networks are described. As discussed, the MISU architecture needs the additional support from upper layers. It causes less service disruption in handover due to being able to enable multiple data streams simultaneously on mobile terminals. In contrast, the MIAU architecture reduces the complexity of upper layers. But, longer handover latency is expected as a result of using multiple network interfaces on a rotate basis.

With the MIAU, data traffic can not be recovered until the switching of network interfaces has been completed in a handover. In this section, a multi-interface mobile terminal model based on the MIAU is presented. The proposed model can retain the MIAU's support for ordinary upper layer protocols, and enable fast handover, which is not achieved in the traditional MIAU solutions.

With the MIAU, the proposed multi-interface model can maximise its compatibility with upper layer protocols such as Mobile IP and TCP. Meanwhile, the support on the latest IEEE 802.21 framework [118] provides extensibility. MIH functional services are utilised for internal communications among the system modules at a mobile terminal. In the proposed model, IEEE 802.21 is utilised in such a way that a mobile terminal can cooperate with both non-MIH networks infrastructure and MIH enabled networks in future. To tackle handover delay problem of the MIAU, the cross-layer design is introduced into the model to shorten the gap between adjacent layers.

The proposed generic multi-interface architecture is shown in Figure 7.2. An intermediate component, named as Handover Management Module (HMM) is proposed between upper layers and lower layers. Acting as a logical layer to isolate the heterogeneities of physical network media from applications, the HMM is responsible for handover related processing. The HMM is composed of four subsystem modules: Policy Manager (PM), Handover Decision Trigger (HDT), Network Selector (NS) and POA Candidate Cache (PCC). Four subsystem modules work together to gathers versatile information from heterogeneous network interfaces and make handover decision. The internal communications can be based on the MIH services or any user-defined services.

The HDT is included in the HMM to receive the cross-layer trigger commands from the NS, and prompt the corresponding processing at Mobile IP [21]. In this thesis, *cross-layer trigger* is referred to as the process of informing Mobile IP of changes to link status at lower layers. Handover policies are managed by the PM, and executed by the NS, in which various handover decision algorithms can be implemented. The communications between the HMM and lower layers are relayed by the PCC. The PCC helps to mitigate the heterogeneity of lower layer network stack, and ensure that lower layer metric information conforms to the common rules for processing at the HMM.

Each network interface can work in either primary or standby mode. At any time, only one network interface is in primary mode at any time and is activated for carrying data and signalling traffic. Other interfaces are in standby mode and are allowed to receive signalling from new POAs by incorporating intelligence at HMM. Thus, a mobile terminal may utilise redundant signalling streams enabled on its standby interfaces for network discovery and other operations.

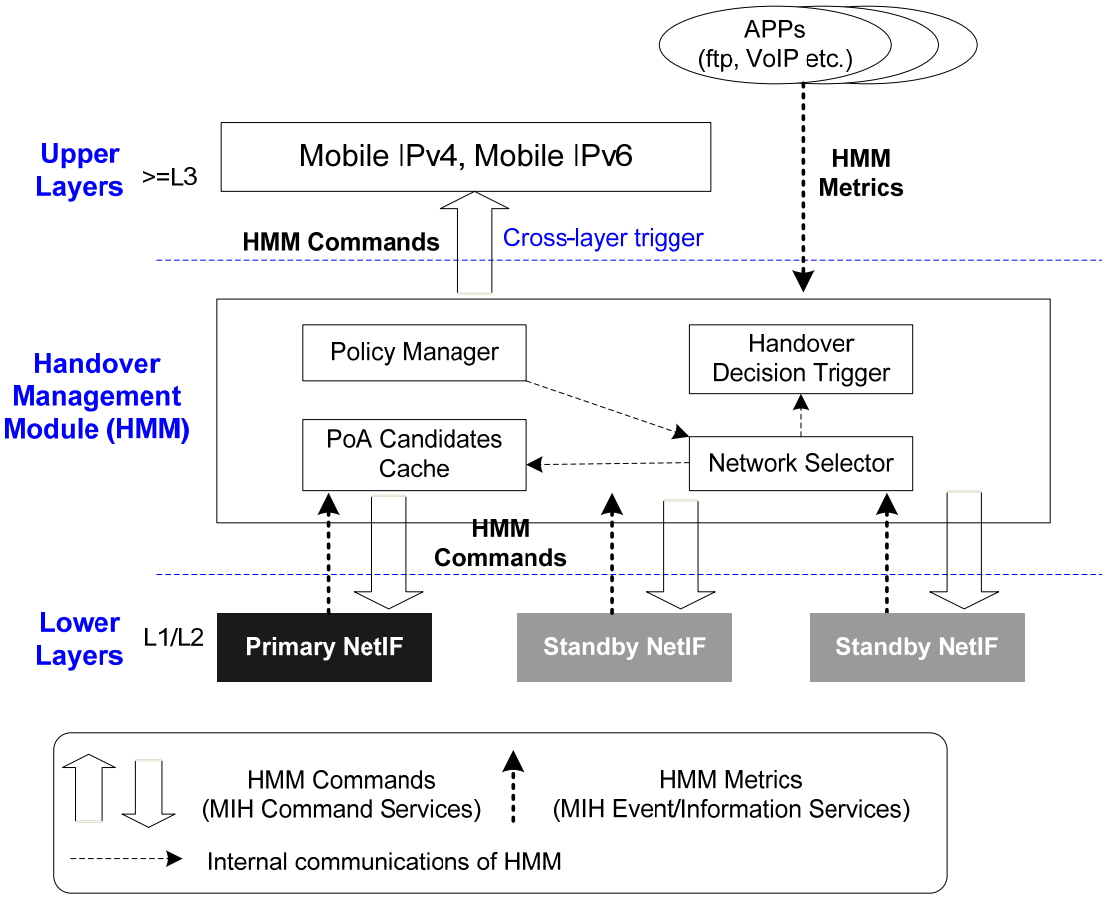


Figure 7.2 A generic multi-interface model for media independent handover

Figure 7.2 illustrates all the subsystem modules, and their possible interactions between each other. Handover metric information such as signal strength, QoS and service type from both lower layers and upper layers is carried on the defined *HMM Metrics*, which are implemented in 802.21 MIH Information service. The HMM Metrics provide network condition information on to the HMM to assist its handover decision. The MIH event services are utilised to notify the HMM of link events. In response, the HMM sends *HMM Commands* to trigger the corresponding operations at either lower layers or

upper layers. The HMM Commands to lower layers are used to instruct the switching of network interfaces. In contrast, the HMM Commands to upper layers indicate the actions that can be taken at upper layers when the predefined events, e.g. loss of radio connection, take place at lower layers. The notifications via the HMM can thus enable cross-layer trigger, which resolves handover delay problem of the MIAU.

7.4 Implementation Issues

According to the proposed generic multi-interface model, a dual-interface mobile terminal has been implemented in Network Simulation 2 (ns-2) [120]. The implementation is intended to find out the implementation issues in real systems. In the dual-interface implementation, a mobile terminal is supposed to be equipped with two IEEE 802.11 WLAN cards. And a Handover Manager (HOMgr) resides on the mobile terminal to execute the HMM functions. The architecture of the mobile terminal is shown in Figure 7.3. ns 2.29 [120] is used as the platform for development. The following modifications have been made to ns 2.29 for the implementation of the proposed mobile architecture:

- Support for multiple wireless channels on a mobile terminal;
- Probe embedded in lower layer objects for handover metrics' gathering;
- Customisation of NO Ad-Hoc Routing Agent (NOAH) [121] for routing capability in the dual-interface mobile terminal;
- Proactive triggering mechanism added to the mobile IP implementation of ns 2.29;

In ns 2.29, only single network interface is originally supported for a mobile terminal [122]. For this reason, a number of new components and functions that enable the controlling and management of multiple interfaces have been developed.

- Handover Manager (HOMgr) with handover decision and policy engine;
- New functions for accommodating additional network interface on a mobile terminal and interface switching;

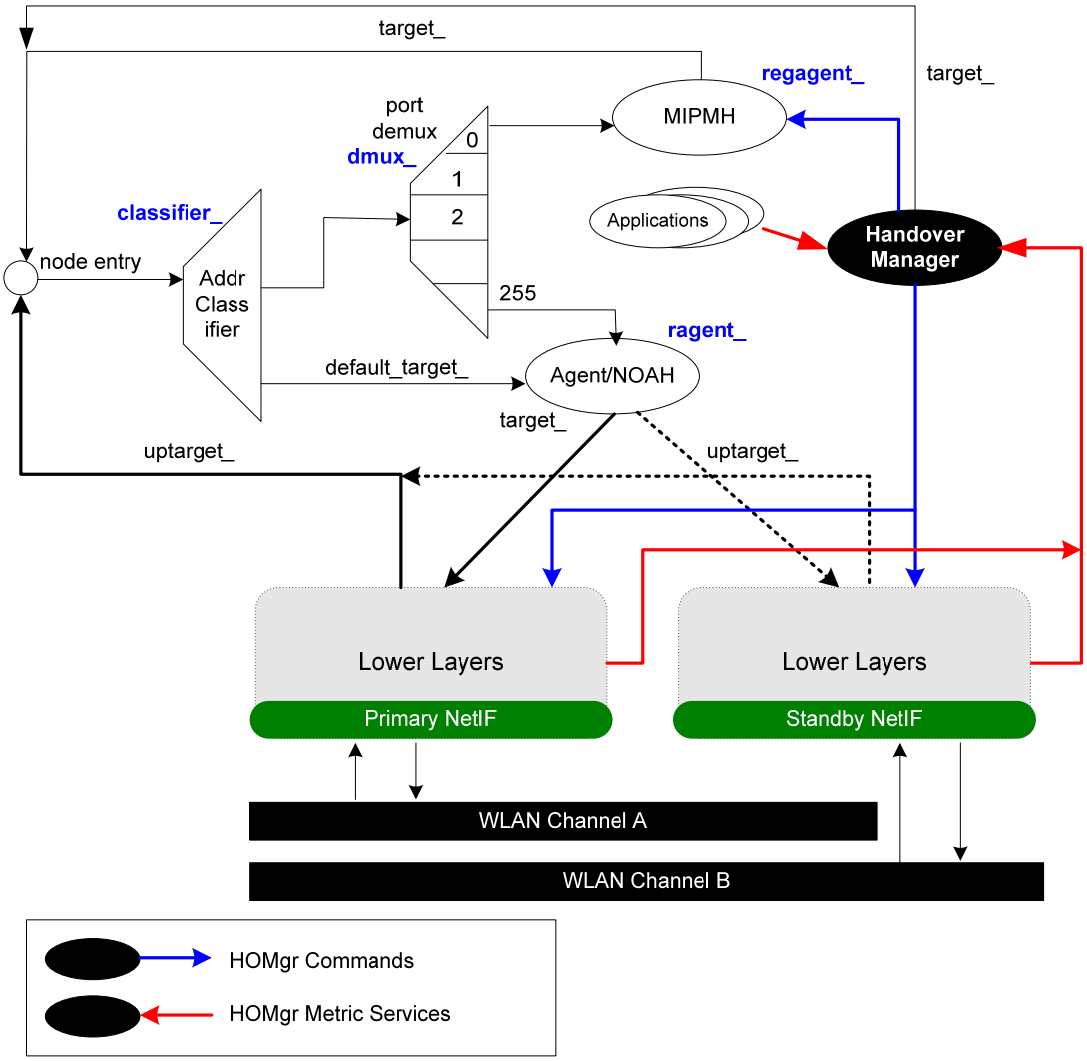


Figure 7.3 Schematic of a dual-interface mobile terminal under the generic multi-interface model

In the dual-interface architecture, the NOAH [13] is applied to direct communications between a mobile terminal and APs. Two 802.11 interfaces are tuned to different wireless channels, and are attached to the mobile terminal without a physical connection between each other. Both of the interfaces can interact with the HOMgr following the design of Sec. 7.3. Mobile IPv4 [21] has been chosen as the network layer protocol for mobility management. Foreign Agent (FA) Care-of-Address (CoA) is utilised on both interfaces when the mobile terminal is connecting to visited networks. The dual-interface mobile terminal works in such a way: at anytime if the standby interface is activated, then the interface in primary mode goes into standby mode. This is followed by the termination of the data traffic on the previously activated interface. The HOMgr

coordinates the operations. The upper layers (TCP/Applications) are not aware of events in lower layers. Periodically, the HOMgr retrieves metric information through the embedded probes at different layers. In this implementation, the HOMgr employs the signal power based handover decision algorithm [82].

It is assumed that the implemented mobile terminal skips channel scanning process in discovering new POAs, which has been studied in [111]. Authentication is not implemented for handover association. However, the multi-interface with a HOMgr provides sufficient extensibility of accommodating other types of network interface such as GPRS and UMTS.

7.5 Performance Analysis

7.5.1 Single-interface vs. Dual-interface

Figure 7.4 illustrates the simulated scenario containing a Mobile Host (MH), Correspondent Node (CN), Home Agent (HA) and Foreign Agent (FA). The HA Access Router (AR) and FA AR are connected to a Wide Area Network (WAN). The CN is attached to the WAN through a fixed link of 5Mbps. The HA Access Point (AP) and FA AP are tuned to different radio channels. They provide partially overlapped radio coverage as shown in the MH's received signal power (Figure 7.5). The MH starts from the HA, and moves towards the FA at a speed of 20m/s, and then goes back to the HA. The coverage of the HA and FA is a circle of radius 450m. TCP/IP is applied between network nodes. Constant Bit Rate (CBR) traffic with a rate of 672Kbps (equivalent to 210-byte packet is sent every 2.25ms) is carried between the CN and the MH. The simulation was run for mobile terminal with both dual-interface and single-interface.

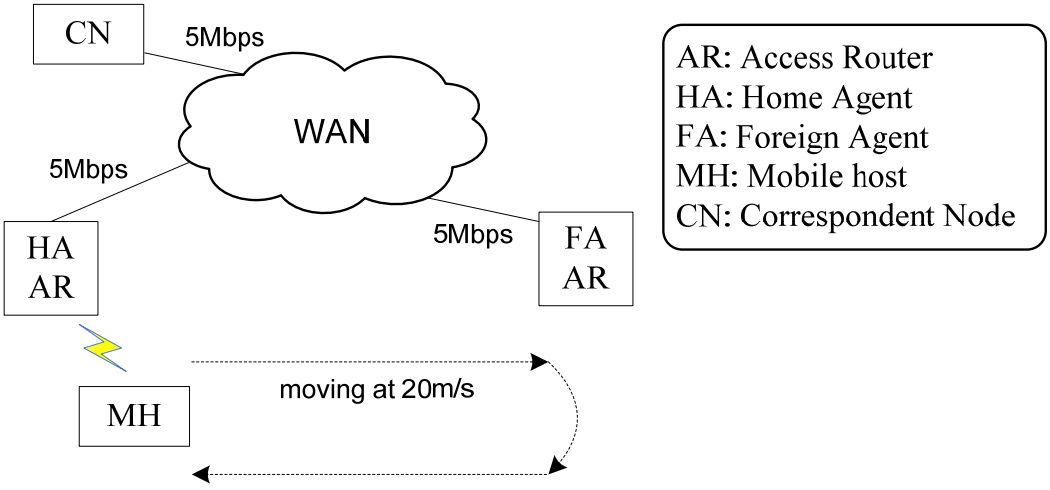


Figure 7.4 Mobility scenario 1

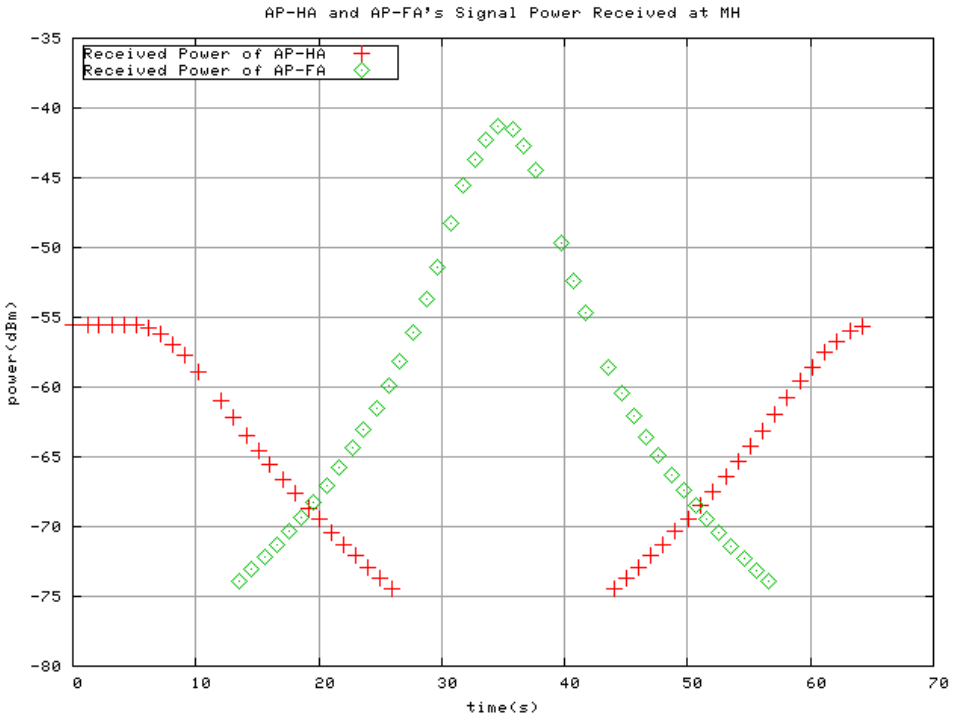


Figure 7.5 Received signal power at the MH

The dual-interface MH is compared with the single-interface MH in regards to their data throughput, end-to-end packet delay and handover delay. The data throughputs of the dual-interface and the single-interface mobile terminals are demonstrated in Figure 7.6 and Figure 7.7 respectively. Both types of mobile terminal show common data transmission characteristics in non radio overlap areas. In the overlap areas, the dual-

interface MH conducted a faster handover than the single-interface MH. For the single-interface MH, the break interval of TCP transmission is 5.14s for 'HA→FA' handover, and 2.64s for 'FA→HA' handover. In contrast, the dual-interface MH took 67ms to recover its data transmission in the 'HA→FA' handover, and 547ms in the 'FA→HA' handover.

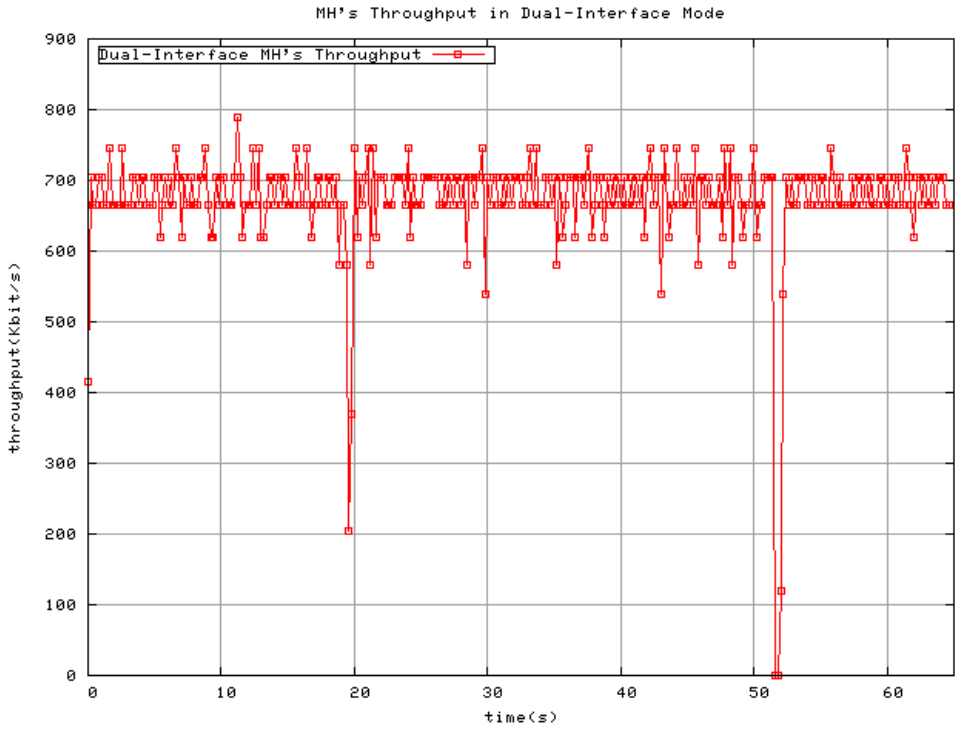


Figure 7.6 The dual-interface MH's throughput

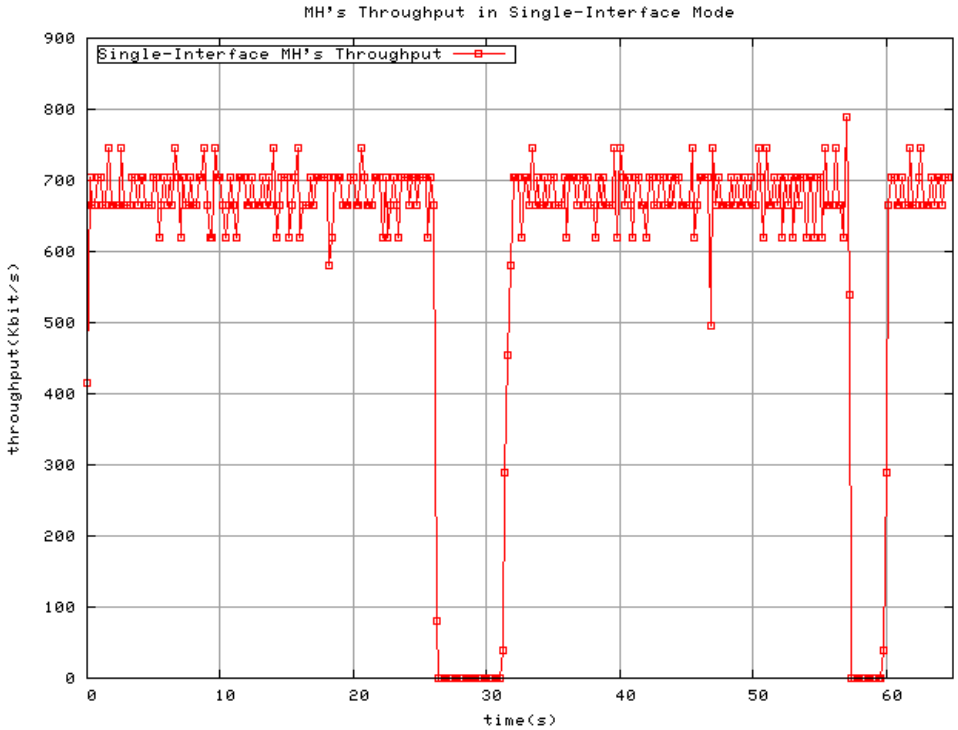


Figure 7.7 The single-interface MH's throughput

The end-to-end packet delay is defined as the packet travelling time from its data source (CN) to its destination (MH). In the simulation, the TCP packet is checked for the parameter. The simulation results of Figure 7.8 and Figure 7.9 show that the end-to-end packet delay during handover is in proportion to transmission break interval. Intuitively, handover causes longer end-to-end packet delay. Although dual-interface architecture can reduce handover delay effectively (as illustrated in Figure 7.6), long end-to-end packet delay is still incurred in the handover of the dual-interface MH. However, the incurred delay can be noticeably reduced to 700ms if a dual-interface architecture is enabled on the MH, in comparison to over 5s delay for the single-interface MH.

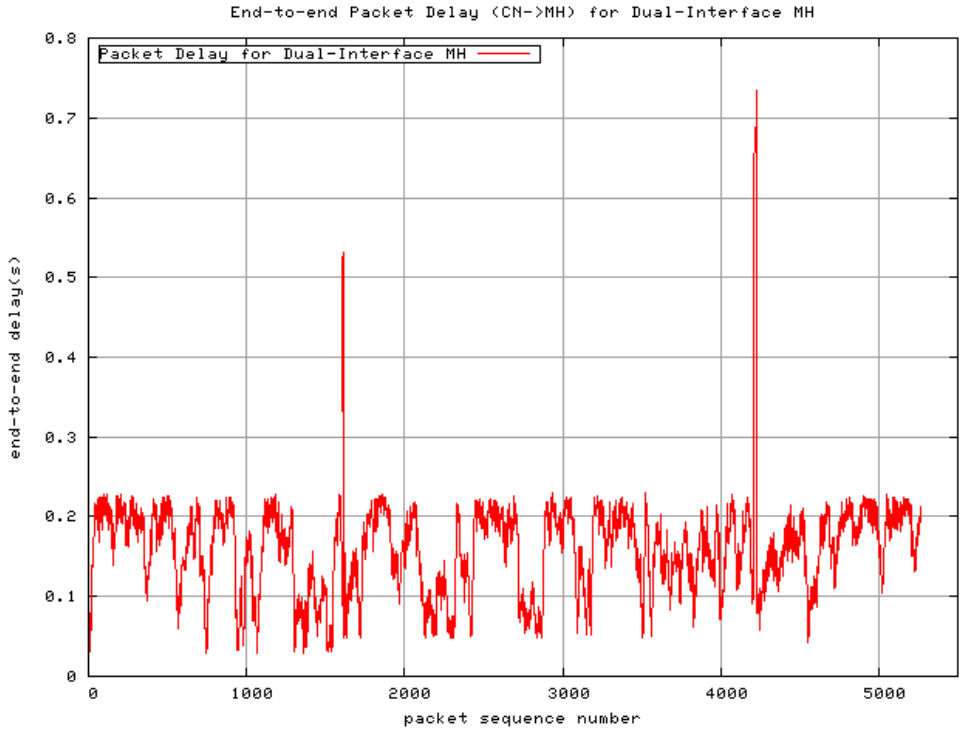


Figure 7.8 End-to-end packet delay from the CN to the dual-interface MH

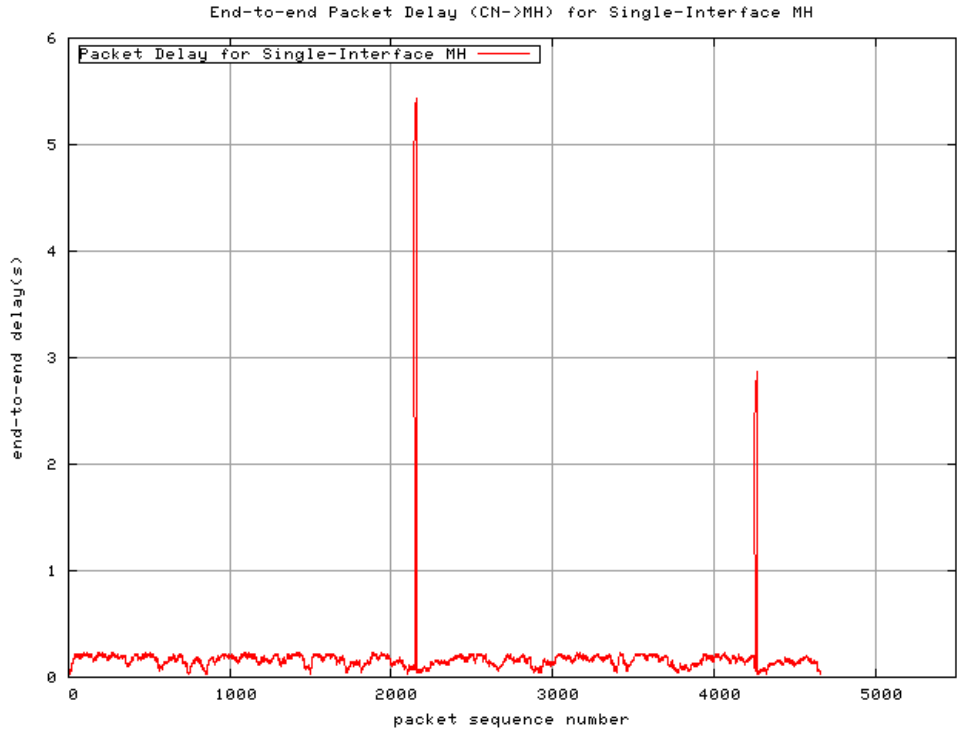


Figure 7.9 End-to-end packet delay from the CN to the single-interface MH

7.5.2 Impact of Access Heterogeneities on Handover

The access heterogeneities are defined as the differences in access technologies that are employed by heterogeneous wireless networks for providing network services. The scenario of Figure 7.10 is used to evaluate the impact of access heterogeneities on handover. The dual-interface MH moves anti-clockwise along the triangle formed by the FA1, FA2 and HA. Data traffic is carried between the CN and the MH. The radio coverage areas of three nodes are illustrated in Figure 7.10. According to the defined trajectory, the MH is expected to hand over in both the radio overlap (of the FA1 and FA2) and non-overlap areas. Different Advertisement Intervals (ADI) and Advertisement Lifetime (ADF) are configured on the FA1 AR and FA2 AR. The MH starts at 5 sec from the FA1, and arrives at the FA2 at 30 sec. Then, it moves towards the HA, and reaches the HA at 70 sec. The MH finally returns to the FA1 at 110 sec, by which the simulation stops.

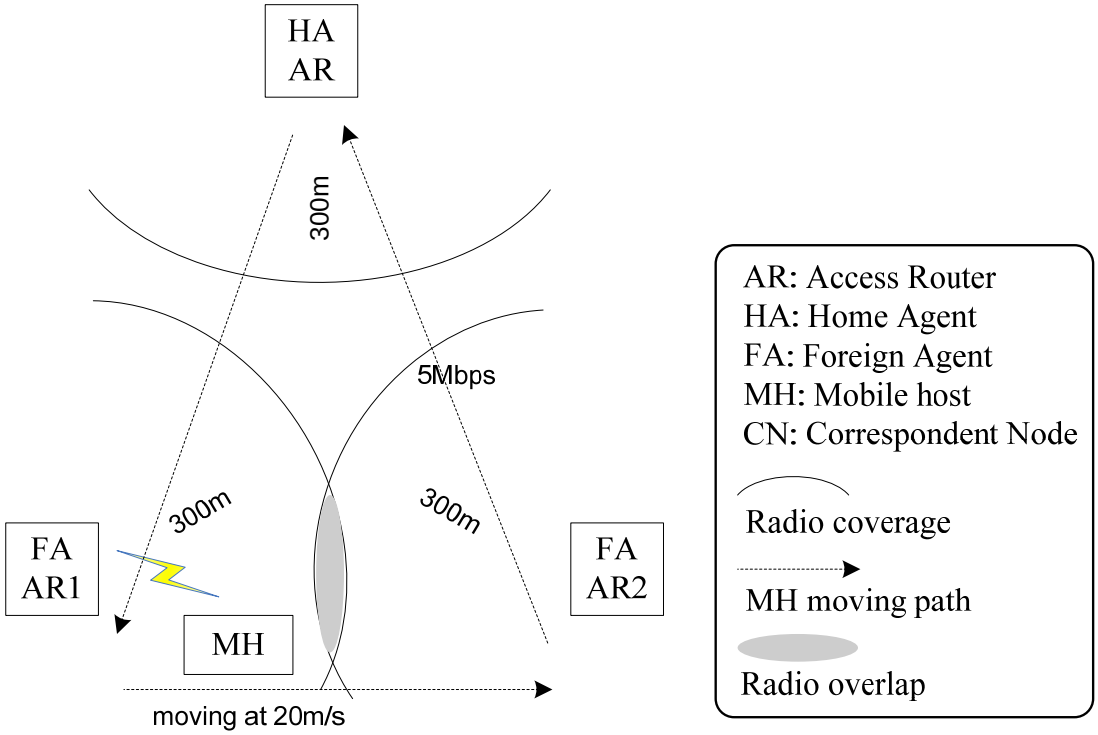


Figure 7.10 Mobility Scenario 2

The agent routers of different network domains may come up with different settings for mobility management, in particular on ADI and ADF. In the scenario of Figure 7.10, the

ADF is set to be 3 times the ADI. The setting means that the MH is allowed to miss three consecutive advertisements before the attachment with the current agent is regarded as being invalid by the MH. The AP that the MH attaches to is referred to as the *old AP*, in contrast to the *new AP* that the MH would switch to in a handover. The ADI of the old AP is fixed to 1s. The ADI of the new AP was varied in the simulation, and thus the ratio of ADI_{old_AP} to ADI_{new_AP} may change accordingly. When the cross-layer trigger of Figure 7.2) is enabled through the HOMgr, the handover delay can keep a constant value at around 40ms as illustrated Figure 7.11. However, when the cross-layer trigger is not applied, the handover delay fluctuates at 3s. The corresponding results can be found in Figure 7.12. In both cases, the handover delay appears not to be influenced by the difference of the ADI between heterogeneous wireless networks. The introduction of the cross-layer trigger mechanism in the proposed multi-interface model can effectively reduce handover delay down to 40ms.

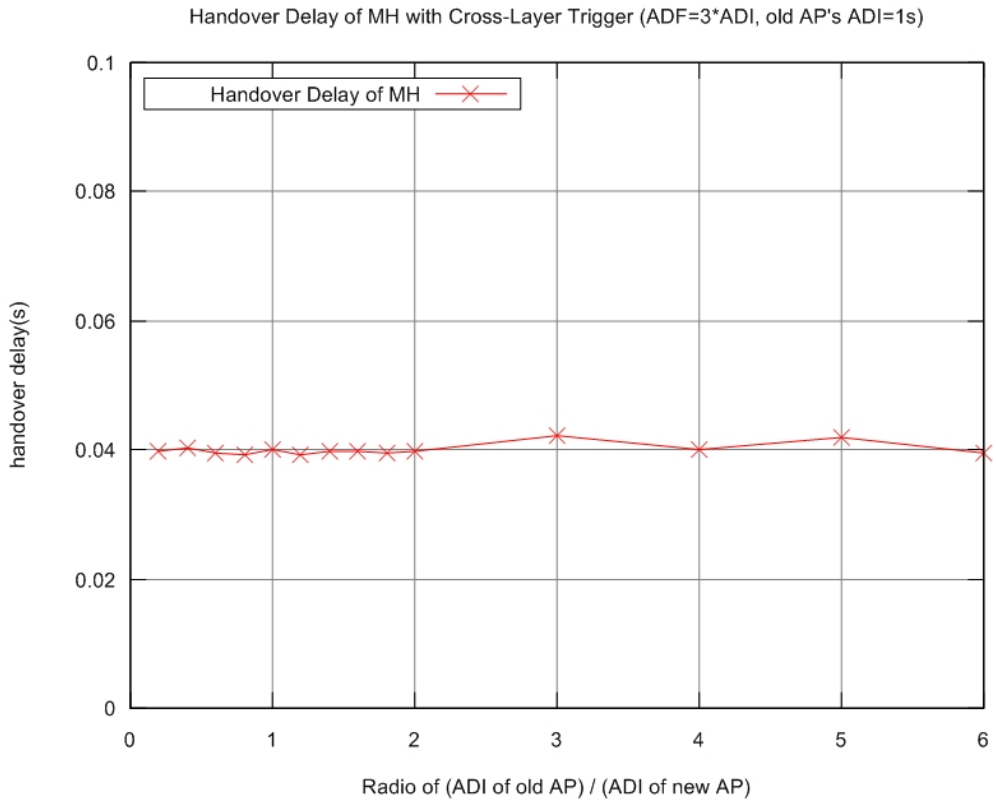


Figure 7.11 Heterogeneity of ADI vs. Handover delay (with cross-layer trigger mechanism)

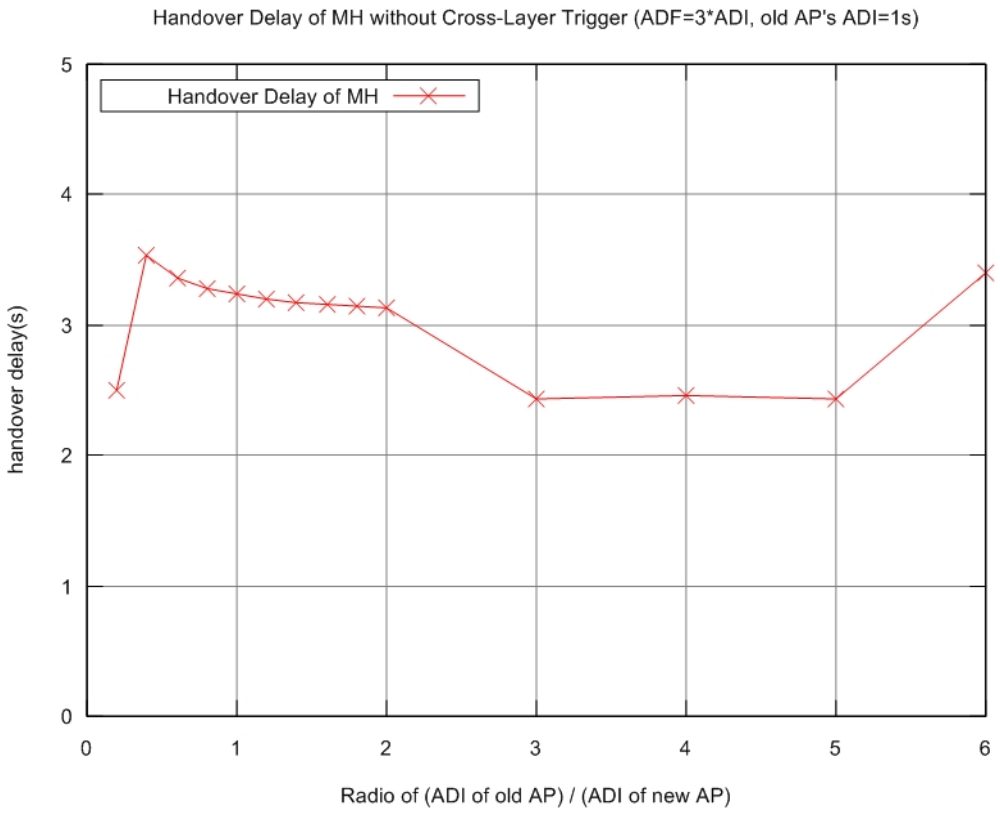


Figure 7.12 Heterogeneity of ADI vs. Handover delay (without cross-layer trigger mechanism)

To evaluate the impact of ADI on handover delay, the ADI of the agent router is varied with ADF = 3*ADI. Figure 7.13 shows that the handover delay is not influenced by the varying ADIs (of both the old and the new APs) in the cross-layer trigger scenario. The MH with the cross-layer trigger enabled can trigger the mobile IP registration with the new AP immediately once the loss of radio connection is detected at link layers. However, if no cross-layer trigger mechanism is applied on the MH, move detection would not occur until the advertisement from the current AP expires. In this case, the ADF of access routers would have an impact on handover delay. Moreover, through the simulation, it is found that the advertisement from the new AP may get lost due to the collision at the MAC layer during handover. Therefore, the handover delay moves up as the ADI rises (as shown in Figure 7.14). The MH may experience the smaller handover delay of 40ms if the cross-layer trigger is enabled.

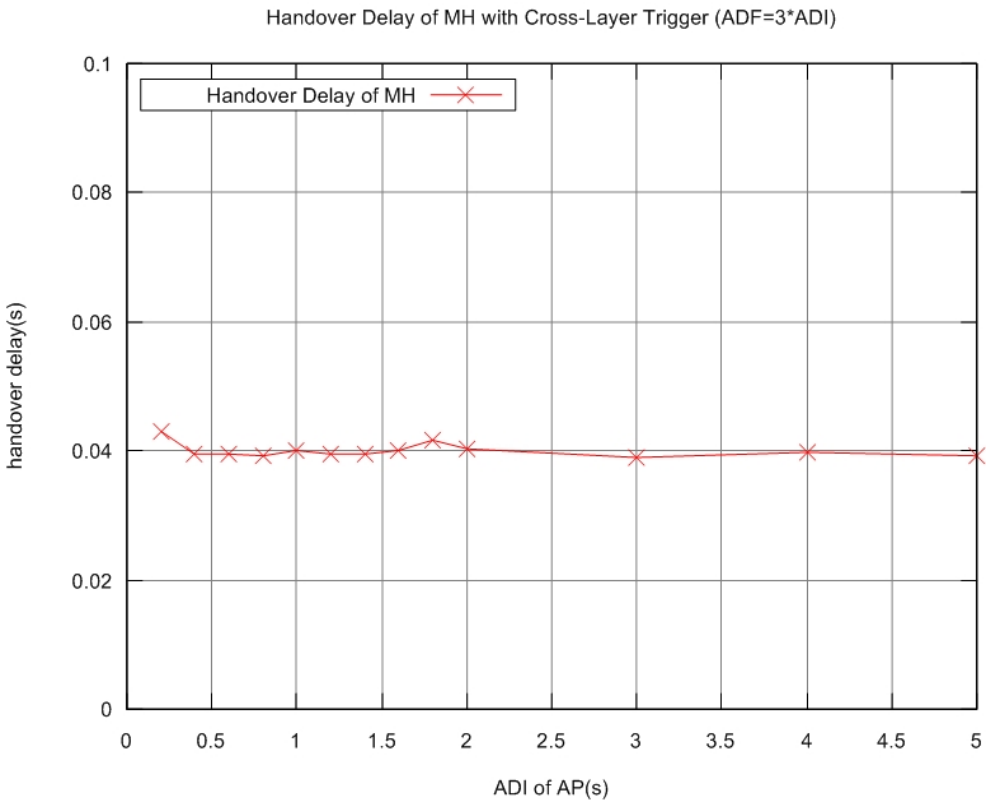


Figure 7.13 ADI vs. Handover delay (with cross-layer trigger)

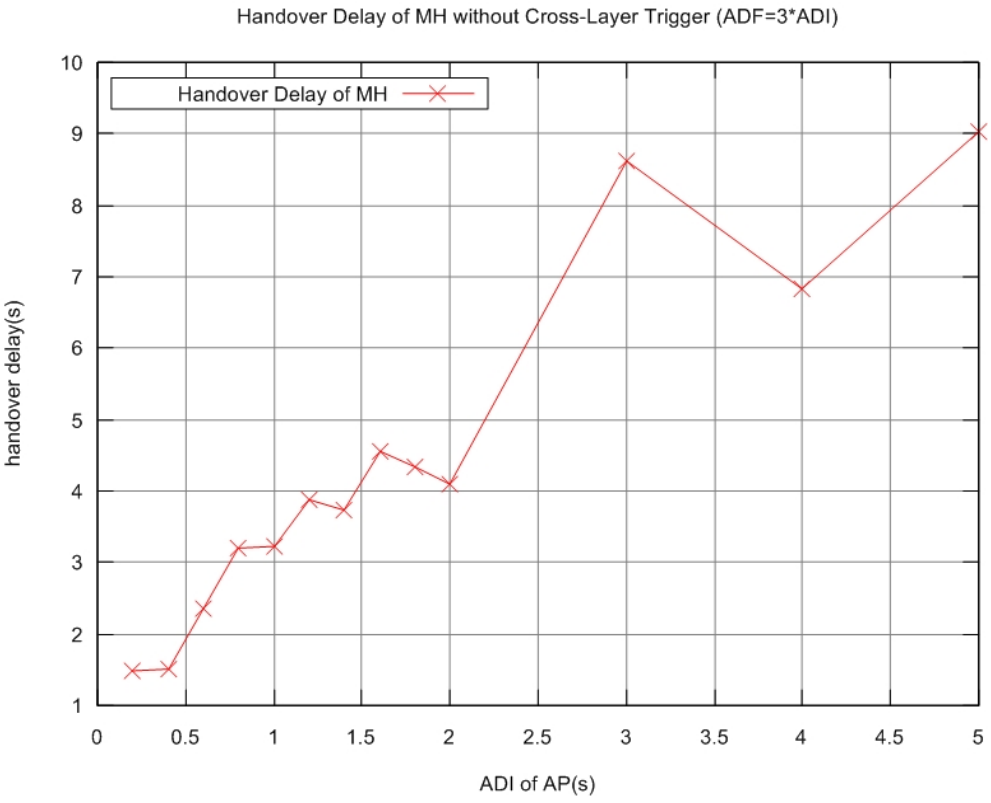


Figure 7.14 ADI vs. Handover delay (without cross-layer trigger)

The experiment moves further to transport layers. The TCP sequence number of the non real time traffic being carried from the CN to the MH is used for performance evaluation. The experimental results for cross-layer trigger and non cross-layer trigger are illustrated in Figure 7.15 and Figure 7.16 respectively. For the MH with the cross-layer trigger, the interruption of packet transmission during the handover at the radio overlap area is not obvious due to the short handover delay (around 40ms). But, the TCP transmission gap in the handover at the non-radio-overlap area is noticeable because of lack of radio coverage (illustrated as handover in the non-overlap area in Figure 7.15). The cross-layer trigger makes the TCP sequence curve fairly smooth. In contrast, the data transmission in non cross-layer trigger case has a longer breaking time, as demonstrated in Figure 7.16. The larger ADI can result in an excessive transmission recovery time. It is found that TCP retransmission mechanism has contributed to a TCP transmission break in handover. It is observed that the transmission break of 50s is incurred in radio overlap areas when ADI = 4s.

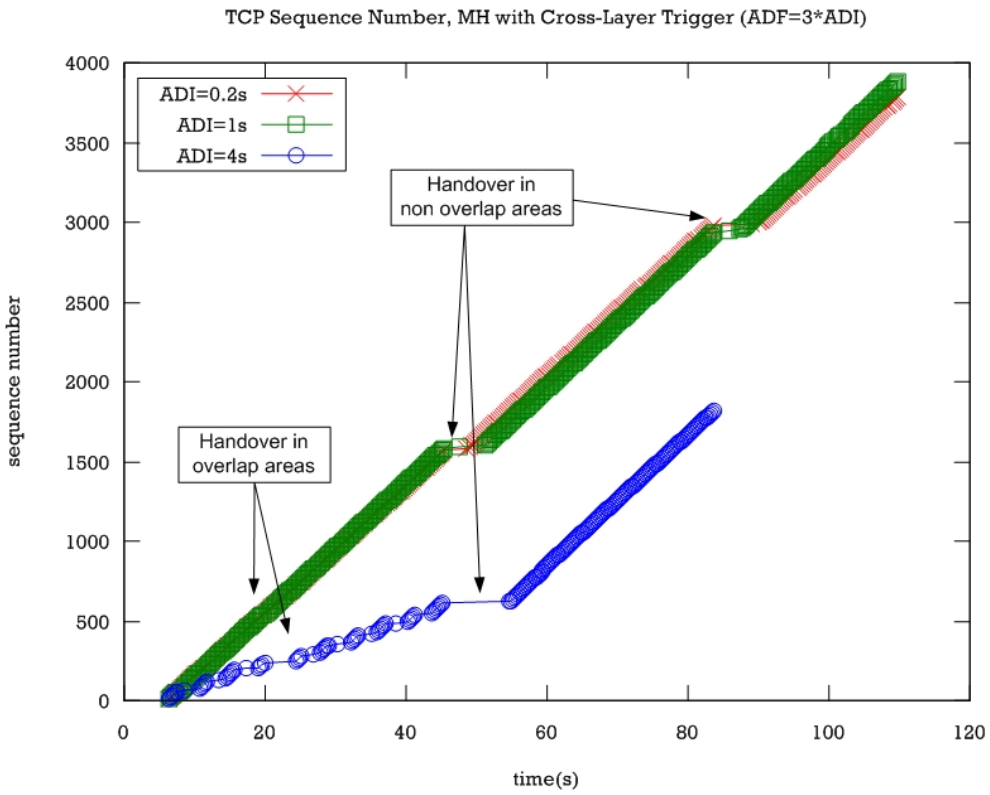


Figure 7.15 TCP sequence number (with cross-layer trigger)

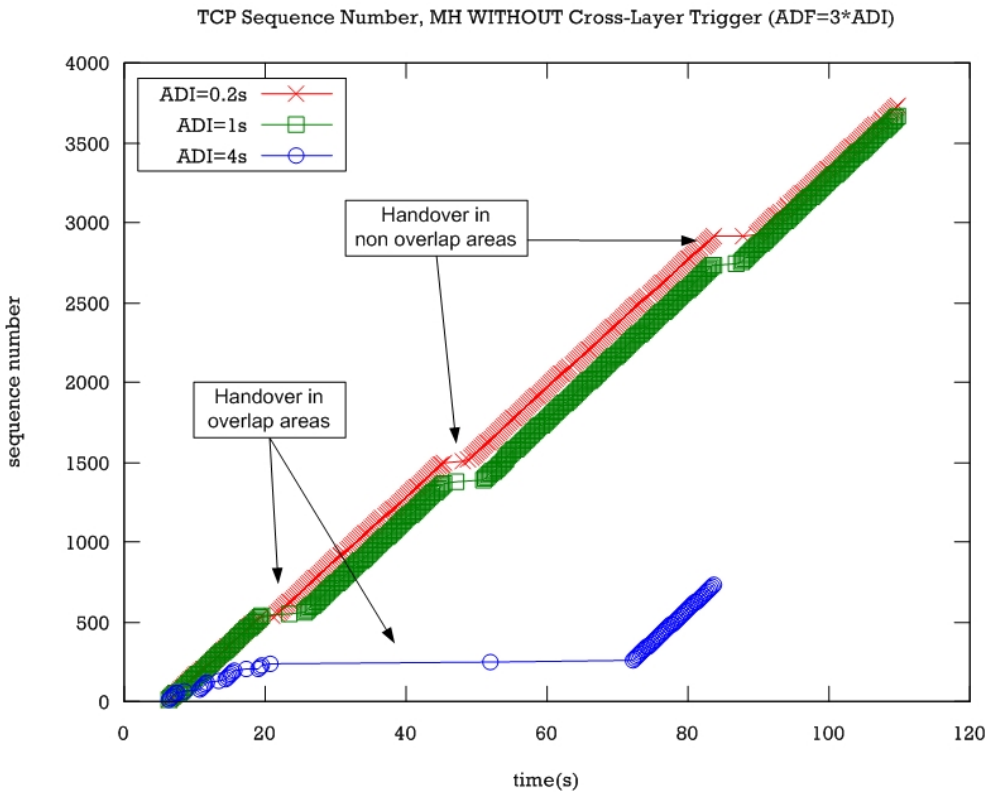


Figure 7.16 TCP sequence number (without cross-layer trigger)

7.6 Conclusion

In this chapter, a generic multi-interface mobile terminal model is introduced. The proposed model can work with IEEE 802.21 for media independent handover, and supports the accommodation of multiple network interfaces on a mobile terminal. The simulation on ns-2 proves that it can support ordinary upper layer protocols such as TCP, UDP and Mobile IPv4. By implementing a cross-layer trigger mechanism, it greatly reduces handover delay, and makes handover delay independent of heterogeneities of access networks. It was shown that a smooth transmission of non real time traffic being carried on TCP can be expected in a handover when the cross-layer trigger mechanism is applied and visited networks have small advertisement interval (<1s).

Chapter 8

CONCLUSIONS AND FUTURE RESEARCH WORK

This chapter summarises the thesis contributions and discuss potential directions for future research in the area.

8.1 Summary of Thesis Contributions

The research work in this thesis focused on the development of new techniques for supporting seamless handover in a multi-operator and multi-technology network environment. The thesis contributions have been made in the four areas.

8.1.1 Dynamic Trust Information Retrieval for Global Roaming

In Chapter 4, a network trust information retrieval scheme for global roaming was presented. The concept "Network Trust Correlation" (NTC) was proposed to represent a versatile and complex network trust relationship between two networks. A quantitative NTC model based on trust association hops was proposed. With this NTC model, the thesis presents a network trust information retrieval and distribution method. The proposed method makes use of a mobile user's mobility and handover process to retrieve and exchange network trust information between two neighbouring networks without any direct communication between the neighbours. Consequently, when a number of mobile users are involved, a network can build up its neighbour network trust pattern. Instead of relying on other mechanisms for getting trust information required for roaming, a mobile user can dynamically obtain the necessary trust information from its serving network.

In regards to the practical implementation of the proposed scheme, two implementation solutions corresponding to two different operation modes of mobile terminal were

presented. In the Active Operation Mode NTC (AOMN) solution, a mobile user carrying active sessions is involved in the NTC process. The AOMN solution relies on actual handover event and can produce the most accurate information about neighbours' trust relationships. In contrast, the Power Save Mode NTC (PSMN) approach takes advantage of the location update and paging process of a mobile user in a dormant state to trigger the NTC process. The PSMN derives neighbours' trust relationships from analysing location update records of mobile users involved in the NTC process.

To evaluate the performance of the proposed scheme, a hexagonal random walk model with the Markov transition probabilities was used to simulate a multi-operator environment, in which a group of mobile users performed handover between networks. A series of simulations were conducted on a number of parameters. The simulation results showed that the proposed trust information retrieval scheme can function in a cost-effective manner, generating 3.7% additional signalling overhead on networks compared with the standard handover signalling cost. Moreover, it was shown that one mobile user per network would be sufficient to establish a complete network trust pattern in most cases when the proposed scheme is widely deployed in networks. From the perspective of real deployment, this makes sure that the implementation of the proposed NTC scheme would not be a burden for operational networks. It was found that the time taken for constructing neighbour trust pattern can be speeded up by increasing the number of mobile users involved.

8.1.2 Trust Assisted Handover Algorithm for Reliable Handover

In Chapter 5, a Trust assisted Handover Algorithm (THOA) was presented to deal with the coexistence of multiple network operators, which has not been addressed in current handover approaches. A THOA cost function that can take into account network trust relationship along with multiple handover metrics was proposed. The proposed THOA cost function works with a trust coefficient model that normalises the NTC data in order to have networks holding different trust values accurately weighed in network selection. To implement the proposed algorithm at the mobile end, the thesis presented an optimised network selection process that is designed to provide build-in trust awareness. The optimised network selection process along with the THOA can guarantee more

reliable handover in a multi-operator environment. In this way, the THOA makes the support on dealing with the multiplicities of network operators an integral part of the mobile's handover algorithm. The THOA provides flexibility for a network operator to define its own handover policies and exert control on handover of its subscribers.

A system analysis framework was proposed for the performance evaluation of the THOA algorithm. The framework covers four parts: Movement Detection, Network Selection, Address Configuration and Network Registration, all of which contribute to handover delay. The proposed analysis framework was applied to a network model that was created for simulating a multi-operator heterogeneous wireless environment. The proposed multi-operator heterogeneous network model can mock up an interworking of UMTS and WLAN networks, and the variations of their network conditions using a Markov chain based transition probability matrix. Furthermore, two new parameters "Trust Density" and "Load Balance Factor" (LBF) were presented for the first time for the THOA related performance analysis. Their mathematical definitions were given.

The simulation results showed that the proposed THOA can effectively reduce unnecessary handover attempts. This made a reduction of up to 35% in handover delay compared with other handover approaches. The results for the LBF proved that the THOA can provide a mechanism of having network load evenly distributed among available networks by adjusting its THOA Effective TAH Scope (ETS). More simulations on handover delay suggested that the THOA can work well with the fine-grained AAA policies implemented by networks. In addition, through the simulations, the thesis elaborated how the THOA can be tuned to meet the performance requirements in various trust density scenarios. The simulations on QoS showed that the probable degradation of QoS from applying the THOA can be compensated by its own mechanism. Maintaining QoS in a handover would make the THOA very attractive to the deployment of real wireless networks.

8.1.3 Proxy Based Authentication Localisation Scheme

In Chapter 6, a Proxy Based Authentication Localisation (PBAL) scheme was proposed to provide fast authentication in a handover taking place between two networks without a trust relation. This is achieved by localising authentication in a handover at a proposed entity called Fast AAA Proxy (FAP). The FAP acts as a third-party entity for processing AAA requests coming from the networks associated with it. A trust association model based on pairwise keys shared between network entities was proposed. The proposed trust association model allows a large number of heterogeneous wireless networks to be interconnected in a loosely coupled manner through a few FAPs.

The thesis presented a fast authentication ticket generation and encryption mechanism. With this mechanism, a mobile's home AAA server proactively produces the security credentials that can be used by the mobile user later for its fast authentication in a handover. The PBAL provides a mutual authentication mechanism, through which both a mobile user and its visited network's associated FAP can perform identity verification mutually in a handover. This proposed authentication mechanism is further protected by a session wide Local Authentication Key (LAK) that is derived from the trust association between the mobile user and its home AAA server. To minimise the risk of using a compromised session key, a session key renewal method was proposed in the PBAL. This session key renewal method allows both the mobile user and its visited network to initialise session key renewal.

The thesis adopted an analytical approach to evaluate the security of the proposed PBAL scheme. The analysis on mutual authentication showed that a mobile user is able to verify whether a FAP is an authorised agent representing its home AAA server. The mobile user authenticates to the FAP by showing its knowledge of the LAK that can be computed using the correct preshared key applied between the mobile user and its home AAA server. The PBAL scheme demonstrated strong security against the replay attacks in the form of an adversary impersonating a mobile user. Moreover, the PBAL provides sufficient protection against the replay attacks, in which an adversary tries to fool a mobile user by impersonating an access network. This is done by using the LAK for the FAP verification, which is transitive and session related. Further security analysis was

conducted on the impact of network corruption at an access network applying the PBAL. It was proved that overhearing radio link and the disclosure of encrypted fast authentication ticket contents would not be a serious problem, because the related risk is limited to receiving the above mentioned replay attacks.

In summary, the proposed PBAL provides an effective and secure mechanism of localising authentication at a third-party proxy in a handover without compromising security features in a multi-operator environment.

8.1.4 Multi-Interface Mobile Model for Media Independent Handover

In Chapter 8, a multi-interface mobile terminal model for media independent handover was proposed. The proposed model was designed for facilitating seamless handover across heterogeneous wireless networks from the perspective of a mobile terminal. It is based on the idea of "Multi-Interface Alternatively Used". In the proposed multi-interface architecture, a Handover Management Module (HMM) was presented which includes four subsystem modules: Policy Manager (PM), Handover Decision Trigger (HDT), Network Selector (NT) and POA Candidate Cache (PCC). The proposed architecture uses a cross-layer design approach, and is compatible with IEEE 802.21 Media Independent Handover (MIH) standard.

Based on the proposed multi-interface architecture, a dual-interface 802.11 prototype had been implemented in *ns-2*. A Handover Manager was developed to simulate the HMM functions for the mobile terminal. The simulation in *ns-2* proved that the proposed model can support common upper layer protocols such as TCP, UDP and Mobile IPv4. The cross-layer trigger mechanism applied greatly reduced handover delay, and made handover delay independent of heterogeneities of access network technologies. A smooth transition of non real time traffic being carried on TCP can be expected in a handover when such a cross-layer trigger mechanism is applied.

8.2 Future Research Work

The future wireless systems are expected to be based on heterogeneous wireless networks that are interconnected for providing ubiquitous high bandwidth services. Some of the future research directions that are related to handover across heterogeneous wireless networks belonging to multiple network operators are listed and briefly discussed below:

- **Decentralised-AAA security framework for the NG heterogeneous wireless networks:** In this thesis work, it is assumed that the current security framework which was originally designed for homogeneous cellular networks will remain for the interworking of heterogeneous wireless networks in the future. The current security framework relies on a centralised AAA server for identity verification in a handover. However, heterogeneous wireless systems may rely on different security mechanisms and their network operators may execute different security policies. When a mobile user authenticates to these heterogeneous wireless networks in a handover, a centralised AAA server based security framework may become the bottleneck for provisioning IP-based multimedia services. Therefore, an alternative approach "a decentralised-AAA security framework" may be an effective add-on for the interworking of heterogeneous wireless systems. The decentralised-AAA security framework moves the primary AAA mechanism of a network to a group of networks to facilitate network collaboration. Future work can investigate this new security approach that is expected to bring many benefits to delay sensitive multimedia services in a handover. The decentralised-AAA security approach should be combined with the centralised-AAA security framework in the NG heterogeneous wireless networks.

- **QoS mechanisms for handover across heterogeneous wireless networks:** The thesis assumed that "always best connected" [34] was the objective of performing a handover. However, maximising QoS in a handover can not always guarantee smooth transfer of data transmission in a handover across heterogeneous wireless networks. The smooth transfer of data transmission is characterised by the QoS provided by the source and target systems, which is expected to be nearly equal.

Therefore, the thesis research can be extended to explore the QoS mechanisms for smooth data transfer of seamless handover. QoS context transfer from the source to target systems has been proposed in a tightly coupled interworking architecture. However, this approach is not applicable to a handover from the source to target systems without a trust relationship. In a multi-operator environment, such kind of handover may occur, and will place new challenges because heterogeneous wireless networks may lack necessary mechanisms for negotiating QoS.

- **Network control over handover across heterogeneous wireless networks:** All the solutions proposed in this thesis are based on the assumption that individual mobile users control their handover processes in heterogeneous wireless networks for the clear advantages [123]. However, it may be necessary for network operators to exert control over their network resources to be allocated to a handover. This requirement will become more prominent when network collaboration is widely expected and there is no central party in control. Hence, it will call for a network controlled handover approach. Network controlled handover that relies on communications between the source and target systems is available in homogeneous wireless systems or tightly coupled heterogeneous wireless systems, but will be a challenging issue for handover across heterogeneous wireless systems belonging to multiple network operators. Future work may investigate the network controlled handover mechanisms for such kind of handover.

ABBREVIATIONS

3G	3rd Generation Wireless Network
3GPP	3rd Generation Partnership Project
4G	4th Generation Wireless Network
AAA	Authentication, Authorisation, and Accounting
ABC	Always Best Connected
ADF	Advertisement Lifetime
ADI	Advertisement Interval
AKA	Authentication Key Agreement Protocol
AMF	Authentication Management Field
AMPS	Advanced Mobile Phone Service
AN	Access Network
AP	Access Point
AR	Access Router
AR	Agent Router
AuC	Authentication Centre
AUTN	Authentication Token
AV	Authentication Vector
AVP	Attribute Value Pair
BS	Base Station
BSS	Basic Service Set
CBR	Constant Bit Rate
CDMA	Code Division Multiple Access
CIPGW	Cellular IP Gateway
CK	Cipher Key
CN	Correspondent Node
CoA	Care-of-Address

CTP	Context Transfer Protocol
DoS	Denial of Service
DS	Distribution System
EAP	Extensible Authentication Protocol
EAP-AKA	Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement
EAPOL	EAP over LAN
EAP-SIM	Extensible Authentication Protocol Method for GSM Subscriber Identity
EAP-TLS	EAP Transport Layer Security
ESS	Extended Service Set
ETS	Effective TAH Scope
FA	Foreign Agent
FAAA	Foreign AAA Server
FAP	Fast AAA Proxy
FAT	Fast Authentication Ticket
FATV	Fast Authentication Ticket Vector
FDMA	Frequency Division Multiple Access
GCoA	Global Care of Address
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
GTK	Group Temporal Key
GTP	GPRS Tunnelling Protocol
HA	Home Agent
HAAA	Home AAA Server
HLR	Home Location Register
HSDPA	High-Speed Downlink Packet Access
LAK	Local Authentication Ticket
LAN	Local Area Network
LCoA	Local Care of Address
LMK	Local Master Key

LS	Location Server
MH	Mobile Host
IAPP	Inter-Access Point Protocol
IdM	Identity Management
IK	Integrity Key
IKE	Internet Key Exchange
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
KCK	(EAPOL) Key Confirmation Key
KEK	(EAPOL) Key Encryption Key
MA	Mobility Agent
MAC	Message Authentication Code
MAHO	Mobile Assisted Handover
MCHO	Mobile Controlled Handover
MCRT	Mean Cell Residence Time
MH	Mobile Host
MHOA	Multicriteria Handover Algorithm
MIC	Message Integrity Code
MIES	Media Independent Command Service
MIIS	Media Independent Information Service
MSC	Mobile Switching Centre
NAI	Network Access Identifier
NAN	Neighbour Access Network
NAS	Network Access Server
NASREQ	Network Access Server Requirements
NCHO	Network Controlled Handover
NG	Next Generation
NOAH	NO Ad-Hoc Routing Agent
NTC	Network Trust Correlation
ns-2	Network Simulation Platform 2
QoS	Quality of Service
PAN	Port Access Entity

PBAL	Proxy-Based Authentication Localisation
PDG	Packet Data Gateway
PID	Pseudonym Identity
PLMN	Public Land Mobile Network
PMK	Pairwise Master Key
POA	Point of Attachment
PRF	Pseudo-Random Function
PS	Packet-Switched
PSK	Pre-Shared Key
PTK	Pairwise Transient Key
RADIUS	Remote Authentication Dial-in User Service
RAI	Routing Area Identity
RAND	Random Number
RNC	Radio Network Controller
MD5	RSA Message Digest Algorithm 5
RFC	Request For Comment
RSN	Robust Security Network
RSNA	Robust Security Network Association
RSS	Received Signal Strength
RTT	Round Trip Time
SA	Subnet Agent
SAP	Seamless Authentication Protocol
SAP	Service Access Point
SCTP	Stream Control Transmission Protocol
SGSN	Serving GPRS Support Node
SLA	Service Level Agreement
SQN	Sequence Number
SRES	Signed Response
SSID	Service Set Identifier
STA	Station (IEEE 802.11)
TAH	Trust Association Hop
TCP	Transmission Control Protocol

THOA	Trust Assisted Handover Algorithm
TK	Temporal Key
TMSI	Temporary Mobile Subscriber Identity
TR	Trust Relationship
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunications System
VLR	Visitor Location Register
VoIP	Voice over IP
WAN	Wide Area Network
WEP	Wireless Encryption Protocol
WLAN	Wireless Local Area Network
XMAC	Expected Message Authentication Code
XRES	Expected Response

REFERENCES

[1] N. D. Tripathi, J. H. Reed, and H. F. VanLandinoham, "Handoff in cellular systems," *Personal Communications, IEEE [see also IEEE Wireless Communications]*, vol. 5, pp. 26, 1998.

[2] IEEE, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications - Amendment 4: Further Higher-Speed Physical Layer Extension in the 2.4 GHz Band," in *IEEE Std. 802.11g*: IEEE, June 2003.

[3] K. Ahmavaara, H. Haverinen, and R. Pichna, "Interworking architecture between 3GPP and WLAN systems," *Communications Magazine, IEEE*, vol. 41, pp. 74, 2003.

[4] A. K. Salkintzis, "Interworking techniques and architectures for WLAN/3G integration toward 4G mobile data networks," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 11, pp. 50, 2004.

[5] M. Buddhikot, G. Chandranmenon, S. Han, Y. W. Lee, S. Miller, and L. Salgarelli, "Integration of 802.11 and third-generation wireless data networks," presented at IEEE INFOCOM, April 2003.

[6] T. Zahariadis, "Trends in the path to 4G," *Communications Engineer*, vol. 1, pp. 12-15, 2003.

[7] 3GPP, "3GPP system to Wireless Local Area Network (WLAN) interworking; System description (Release 6)," in *TS 23.234*: 3rd Generation Partnership Project, 2005.

[8] I. F. Akyildiz, S. Mohanty, and X. Jiang, "A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems," *Communications Magazine, IEEE*, vol. 43, pp. S29, 2005.

[9] Y.-F. R. Chen and C. Petrie, "Ubiquitous Mobile Computing," *IEEE Internet Computing*, vol. 7, pp. 16 - 17, April 2003.

[10] M. Stemm and R. H. Katz, "Vertical handoffs in wireless overlay networks," *Mobile Networks and Applications*, vol. 3, pp. 335 - 350, 1998.

[11] M. Gast, "Federated Network Authentication," O'Reilly Network Wireless DevCenter, 2005.

[12] C. Politis, T. Oda, S. Dixit, A. Schieder, H. Y. Lach, M. I. Smirnov, S. Uskela, and R. Tafazolli, "Cooperative networks for the future wireless world," *Communications Magazine, IEEE*, vol. 42, pp. 70, 2004.

[13] S. Y. Hui and K. H. Yeung, "Challenges in the migration to 4G mobile systems," *Communications Magazine, IEEE*, vol. 41, pp. 54-59, 2003.

[14] C. Kalmanek, J. Murray, C. Rice, B. A. G. B. Gessel, R. A. K. R. Kabre, and A. A. M. A. Moskal, "A network-based architecture for seamless mobility services," *Communications Magazine, IEEE*, vol. 44, pp. 103-109, 2006.

[15] I. F. Akyildiz, X. Jiang, and S. Mohanty, "A survey of mobility management in next-generation all-IP-based wireless systems," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 11, pp. 16, 2004.

[16] V. Marques, V. Marques, R. L. Aguiar, C. Garcia, J. I. A. M. J. I. Moreno, C. A. B. C. Beaujean, E. A. M. E. Melin, and M. A. L. M. Liebsch, "An IP-based QoS architecture for 4G operator scenarios," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 10, pp. 54-62, 2003.

[17] IEEE Std. 802.1X, "IEEE standard for local and metropolitan area networks - Port-based network access control," LAN/MAN Standards Committee, October 2001.

[18] B. Aboba, L. Blunk, J. Vollbrecht, J. Carlson, and H. Levkowetz, "RFC 3748 - Extensible Authentication Protocol (EAP)," IETF Network Working Group, June 2004.

[19] G. M. Koien, "An introduction to access security in UMTS," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 11, pp. 8-18, February 2004.

[20] J.-C. Chen and H.-W. Lin, "A gateway approach to mobility integration of GPRS and wireless LANs," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 12, pp. 86-95, 2005.

[21] C. Perkins, "RFC 3220 - IP Mobility Support for IPv4," IETF Network Working Group, January 2002.

[22] S. Minghui, S. Xuemin, and J. W. Mark, "IEEE 802.11 roaming and authentication in wireless LAN/cellular mobile networks," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 11, pp. 66-75, August 2004.

[23] F. M. Chiussi, D. A. Khotimsky, and S. Krishnan, "Mobility management in third-generation all-IP networks," *Communications Magazine, IEEE*, vol. 40, pp. 124, 2002.

[24] J. H. Schiller, *Mobile communications*, 2 ed. Boston, MA: Addison-Wesley, 2003.

[25] N. Banerjee, W. Wei, and S. K. Das, "Mobility support in wireless Internet," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 10, pp. 54-61, 2003.

[26] A. T. Campbell, J. Gomez, S. Kim, A. G. Valko, C.-Y. Wan, and Z. R. Turanyi, "Design, implementation, and evaluation of cellular IP," *Personal Communications, IEEE [see also IEEE Wireless Communications]*, vol. 7, pp. 42-49, 2000.

[27] R. Ramjee, K. Varadhan, L. Salgarelli, S. R. Thuel, W. Shie-Yuan, and T. La Porta, "HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks," *Networking, IEEE/ACM Transactions on*, vol. 10, pp. 396-410, 2002.

[28] S. Das, A. McAuley, A. Dutta, A. Misra, K. Chakraborty, and S. K. Das, "IDMP: an intradomain mobility management protocol for next-generation wireless networks," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 9, pp. 38, 2002.

[29] IEEE Std. 802.11F, "IEEE trial-use recommended practice for multi-vendor access point interoperability via an inter-access point protocol across distribution systems supporting ieee 802.11 operation," LAN/MAN Standards Committee, June 2003.

[30] Y.-B. Lin and I. Chlamtac, *Wireless and Mobile Network Architectures*. New York: Wiley Computer Publishing, 2001.

[31] A. Calvagna and G. D. Modica, "A cost-based approach to vertical handover policies between WiFi and GPRS," *Wireless Communications and Mobile Computing*, vol. 5, pp. 603 - 617, Aug 2005.

[32] F. Zhu and J. McNair, "Multiservice Vertical Handoff Decision Algorithms," *EURASIP Journal onWireless Communications and Networking*, vol. 2006, pp. 13, May 2006.

[33] R. L. Aguiar, S. Sargento, A. Banchs, C. J. A. B. C. J. Bernardos, M. A. C. M. Calderon, I. A. S. I. Soto, M. A. L. M. Liebsch, T. A. M. T. Melia, and P. A. P. P. Pacyna, "Scalable qos-aware mobility for future mobile operators," *Communications Magazine, IEEE*, vol. 44, pp. 95-102, June 2006.

[34] E. Gustafsson and A. Jonsson, "Always best connected," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 10, pp. 49, 2003.

[35] J. Blau, "Wi-Fi hotspot networks sprout like mushrooms," *Spectrum, IEEE*, vol. 39, pp. 18-20, September, 2002.

[36] J. D. Avila, "Wi-Fi Users, Beware: Hot Spots Are Weak Spots," *The Wall Street Journal*, January 16, 2008.

[37] Wi-Fi Alliance, "Wi-Fi Alliance 2006 Annual Report," http://www.wi-fi.org/, 2006.

[38] M. Audeh, "Metropolitan-scale Wi-Fi mesh networks," *Computer*, vol. 37, pp. 119-121, December 2004.

[39] FON, "What's FON?," http://www.fon.com, 2008.

[40] M. Nakhjiri and M. Nakhjiri, *AAA and Network Security for Mobility Access: Radius, Diameter, EAP, PKI and IP Mobility*: John Wiley & Sons, 2005.

[41] C. Kaufman, "RFC 4306 - Internet Key Exchange (IKEv2) Protocol," IETF Network Working Group, December 2005.

[42] IEEE, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications - Amendment 6: Medium Access Control (MAC) Security Enhancements," in *IEEE Std 802.11i*: IEEE, July 2004.

[43] R. Housley and B. Aboba, "RFC 4962 - Guidance for Authentication, Authorization, and Accounting (AAA) Key Management," IETF Network Working Group, July 2007.

[44] K. Boman, G. Horn, P. Howard, and V. Niemi, "UMTS security," *Electronics & Communication Engineering Journal*, vol. 14, pp. 191, 2002.

[45] 3GPP, "Specification of the MILENAGE Algorithm Set: An example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*," in *TS 35.205*: 3rd Generation Partnership Project, 2007-06.

[46] IEEE, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," in *IEEE Standard 802.11, 1999 Edition*: IEEE, June 2003.

[47] IEEE, "IEEE standard for local and metropolitan area networks - Port-based network access control," in *IEEE Std. 802.1X*: IEEE, October 2001.

[48] J.-C. Chen, M.-C. Jiang, and Y.-w. Liu, "Wireless LAN security and IEEE 802.11i," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 12, pp. 27-36, February 2005.

[49] C. Rigney, S. Willens, A. Rubens, and W. Simpson, "RFC 2865 - Remote Authentication Dial In User Service (RADIUS)," IETF Network Working Group, June 2000.

[50] B. Aboba and D. Simon, "RFC 2716 - PPP EAP TLS Authentication Protocol," IETF Network Working Group, October 1999.

[51] C. Rensing, M. Karsten, and B. Stiller, "AAA: a survey and a policy-based architecture and framework," *Network, IEEE*, vol. 16, pp. 22-27, November 2002.

[52] M. Shin, J. Ma, A. Mishra, and W. A. Arbaugh, "Wireless network security and interworking," *Proceedings of the IEEE*, vol. 94, pp. 455-466, 2006.

[53] L. Salgarelli, M. Buddhikot, J. Garay, S. Patel, and S. Miller, "Efficient authentication and key distribution in wireless IP networks," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 10, pp. 52, 2003.

[54] C. d. Laat, G. Gross, L. Gommans, J. Vollbrecht, and D. Spence, "RFC 2903 - Generic AAA Architecture," IETF Network Working Group, August 2000.

[55] S. Yan Lindsay, Y. Wei, H. Zhu, and K. J. R. A. L. K. J. R. Liu, "Information theoretic framework of trust modeling and evaluation for ad hoc networks," *Selected Areas in Communications, IEEE Journal on*, vol. 24, pp. 305-317, 2006.

[56] C. E. Perkins, "Mobile IP joins forces with AAA," *Personal Communications, IEEE [see also IEEE Wireless Communications]*, vol. 7, pp. 59-61, 2000.

[57] W. Wang, "Authentication for Inter-Domain Roaming in Wireless IP Networks," in *Security And Routing in Wireless Networks*, Y. Xiao, J. Li, and Y. Pan, Eds.: Nova Publishers, 2005, pp. 21-54.

[58] G. M. Koien and T. Haslestad, "Security aspects of 3G-WLAN interworking," *Communications Magazine, IEEE*, vol. 41, pp. 82-88, November 2003.

[59] B. Aboba and P. Calhoun, "RFC 3579 - RADIUS (Remote Authentication Dial In User Service) Support For Extensible Authentication Protocol (EAP)," IETF Network Working Group, September 2003.

[60] B. Aboba and J. Vollbrecht, "RFC 2607 - Proxy Chaining and Policy Implementation in Roaming," IETF Network Working Group, June 1999.

[61] B. Aboba et al., "RFC 2989 - Criteria for Evaluating AAA Protocols for Network Access," IETF Network Working Group, November 2000.

[62] P. Calhoun, J. Loughney, E. Guttman, G. Zorn, and J. Arkko, "RFC 3588 - Diameter Base Protocol," IETF Network Working Group, September 2003.

[63] P. Calhoun, G. Zorn, D. Spence, and D. Mitton, "RFC 4005 - Diameter Network Access Server Application," IETF Network Working Group, August 2005.

[64] B. Aboba and M. Beadles, "RFC 2486 - The Network Access Identifier," IETF Network Working Group, January 1999.

[65] P. Eronen, T. Hiller, and G. Zorn, "RFC 4072 - Diameter Extensible Authentication Protocol (EAP) Application," IETF Network Working Group, August 2005.

[66] V. K. Varma, S. Ramesh, K. D. Wong, M. Barton, G. Hayward, and J. A. Friedhoffer, "Mobility management in integrated UMTS/WLAN networks,"

presented at IEEE International Conference on Communications (ICC), May 2003.

[67] G. Fodor, A. Eriksson, and A. Tuoriniemi, "Providing quality of service in always best connected networks," *Communications Magazine, IEEE*, vol. 41, pp. 154-163, 2003.

[68] M. Bernaschi, F. Cacace, G. Iannello, S. Za, and A. Pescape, "Seamless internetworking of WLANs and cellular networks: architecture and performance issues in a Mobile IPv6 scenario," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 12, pp. 73, 2005.

[69] A. Mishra, S. Min Ho, N. L. Petroni, Jr., T. C. Clancy, and W. A. Arbaugh, "Proactive key distribution using neighbor graphs," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 11, pp. 26-36, 2004.

[70] K. Sethom, H. Afifi, and G. Pujolle, "A distributed and secured architecture to enhance smooth handoffs in wide area wireless IP infrastructures," *SIGMOBILE Mob. Comput. Commun. Rev.* , vol. 10, pp. 46 - 57, July 2006.

[71] K. Hong and S. Jung, "A Hierarchical Key Management Scheme for Authentication of Roaming Mobile Nodes between Domains in Mobile Networks," *IEICE Transactions on Communications*, vol. E89-B, December 2006.

[72] 3GPP, "3GPP system to Wireless Local Area Network (WLAN) interworking; WLAN User Equipment (UE) to network protocols V7.4.0," in *TS 24 234*: 3rd Generation Partnership Project (3GPP), 2006-12.

[73] M. Li, K. Sandrasegaran, and T. Tung, "Trust-Assisted Handover Decision Algorithm in Hybrid Wireless Networks," presented at IEEE Wireless Communications & Networking Conference (WCNC'07), Hong Kong, March, 2007.

[74] J. Arkko and H. Haverinen, "RFC 4187 - Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA)," IETF Network Working Group, January 2006.

[75] 3GPP, "Wireless Local Area Network (WLAN) interworking security V7.3.0," in *TS 33.234*: 3rd Generation Partnership Project, 2006-12.

[76] K. Kyamakya and K. Jobmann, "Location management in cellular networks: classification of the most important paradigms, realistic Simulation framework, and relative performance analysis," *Vehicular Technology, IEEE Transactions on*, vol. 54, pp. 687-708, 2005.

[77] M. Liebsch and X. Perez-Costa, "Utilization of the IEEE802.11 power save mode with IP paging," presented at IEEE International Conference on Communications (ICC), Heidelberg, Germany, 2005.

[78] I. F. Akyildiz, J. S. M. Ho, and Y.-B. Lin, "Movement-based location update and selective paging for PCS networks," *Networking, IEEE/ACM Transactions on*, vol. 4, pp. 629-638, 1996.

[79] Y.-B. Lin, L.-F. Chang, and A. Noerpel, "Modeling hierarchical microcell/macrocell PCS architecture," presented at Communications, 1995. ICC 95 Seattle, Gateway to Globalization, 1995 IEEE International Conference on, 1995.

[80] E. P. C. Kao, *An Introduction to Stochastic Processes*, 1 ed: Duxbury Press, March 1996.

[81] C. W. Lee, L. M. Chen, M. C. Chen, and Y. S. Sun, "A framework of handoffs in wireless overlay networks based on mobile IPv6," *Selected Areas in Communications, IEEE Journal on*, vol. 23, pp. 2118, 2005.

[82] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J. P. Makela, R. Pichna, and J. Vallstron, "Handoff in hybrid mobile data networks," *Personal Communications, IEEE*, vol. 7, pp. 34, 2000.

[83] J. McNair and Z. Fang, "Vertical handoffs in fourth-generation multinetwork environments," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 11, pp. 8, 2004.

[84] Q. Song and A. Jamalipour, "Network selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 12, pp. 42, 2005.

[85] V. Gazis, N. Alonistioti, and L. Merakos, "Toward a generic "always best connected" capability in integrated WLAN/UMTS cellular mobile networks (and beyond)," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 12, pp. 20-29, 2005.

[86] N. D. Tripathi, J. H. Reed, and H. F. VanLandingham, "Adaptive handoff algorithms for cellular overlay systems using fuzzy logic," 1999.

[87] J.-S. Leu, R.-H. Lai, H.-I. Lin, and W.-K. Shih, "Running cellular/PWLAN services: practical considerations for cellular/PWLAN architecture supporting interoperator roaming," *Communications Magazine, IEEE*, vol. 44, pp. 73-84, 2006.

[88] H. J. Wang, R. H. Katz, and J. Giese, "Policy-enabled handoffs across heterogeneous wireless networks," presented at The Second IEEE Workshop on Mobile Computing Systems and Applications (WMCSA), Feb 1999.

[89] M. Ylianttila, J. Mäkelä, and K. Pahlavan, "Analysis of handoff in a location-aware vertical multi-access network " *Computer Networks, Elsevier North-Holland*, vol. 47, pp. 185-201, 2005.

[90] R. Verdone and A. Zanella, "Performance of received power and traffic-driven handover algorithms in urban cellular networks," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 9, pp. 60, 2002.

[91] 3GPP, "Review of Network Selection Principles V7.2.0," in *TR 22.811*: 3rd Generation Partnership Project, 2006-06.

[92] M. Li, K. Sandrasegaran, and T. Tung, "Performance Evaluation of A Multi-Interface Model for Media Independent Handover," presented at The 7th International Symposium on Communications and Information Technologies, Sydney, Australia, October, 2007.

[93] 3GPP, "3G Security; Security architecture V7.1.0," in *TS 33.102*: 3rd Generation Partnership Project, 2006-12.

[94] Y. Ohba, "EAP Pre-authentication Problem Statement," in *draft-ietf-hokey-preauth-ps-00*: IETF Network Working Group, September 2007, pp. IETF Draft.

[95] M. Li, K. Sandrasegaran, and T. Tung, "A Multi-Interface Proposal for IEEE 802.21 Media Independent Handover," presented at The Sixth International Conference on Mobile Business (ICMB), Toronto, Ontario, Canada, July 2007.

[96] IEEE, "IEEE trial-use recommended practice for multi-vendor access point interoperability via an inter-access point protocol across distribution systems supporting ieee 802.11 operation," in *IEEE Std. 802.11F*: IEEE, June 2003.

[97] J. Loughney, M. Nakhjiri, C. Perkins, and R. Koodli, "RFC 4067 - Context Transfer Protocol (CXTP)," IETF Network Working Group, July 2005.

[98] M. S. Bargh, R. J. Hulsebosch, E. H. Eertink, A. Prasad, H. Wang, and P. Schoo, "Fast authentication methods for handovers between IEEE 802.11 wireless LANs," presented at the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots (WMASH), Philadelphia, PA, October 2004.

[99] M. Long, C. H. Wu, and J. D. Irwin, "Localised authentication for inter-network roaming across wireless LANs," *Communications, IEE Proceedings-*, vol. 151, pp. 496-500, October 2004.

[100] S. C.-H. Huang, H. Zhu, and W. Zhang, "SAP: seamless authentication protocol for vertical handoff in heterogeneous wireless networks," presented at the 3rd international conference on Quality of service in heterogeneous wired/wireless networks (QShine), Waterloo, ON, Canada, August 2006.

[101] Y. Jiang, C. Lin, X. S. Shen, and M. Shi, "Mutual Authentication and Key Exchange Protocols for Roaming Services in Wireless Mobile Networks," *Wireless Communications, IEEE Transactions on*, vol. 5, pp. 2569-2577, 2006.

[102] O. Goldreich, S. Goldwasser, and S. Micali, "How to construct random functions," *Journal of the ACM (JACM)*, vol. 33, pp. 792 - 807, October 1986.

[103] F. Bersani and H. Tschofenig, "RFC 4764 - The EAP-PSK Protocol: A Pre-Shared Key Extensible Authentication Protocol (EAP) Method," IETF Network Working Group, January 2007.

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

[105] J. Ylitalo, T. Jokikyyny, T. Kauppinen, A. J. Tuominen, and J. Laine, "Dynamic network interface selection in multihomed mobile hosts," presented at The 36th Annual Hawaii International Conference on System Sciences, Jan 2003.

[106] K. Chebrolu and R. Rao, "Communication using multiple wireless interfaces," presented at IEEE Wireless Communications and Networking Conference (WCNC), Orlando, FL, 2002.

[107] V. S. Kaulgud and S. A. Mondal, "Exploiting multihoming for low latency handoff in heterogeneous networks," presented at Telecommunications, 2005. ConTEL 2005. Proceedings of the 8th International Conference, 2005.

[108] K.-H. Kim, Y. Zhu, and R. Sivakumar, "A Receiver-Centric Transport Protocol for Mobile Hosts with HeterogeneousWireless Interfaces," *Springer Wireless Networks, Special issue: Selected papers from ACM MobiCom 2003*, vol. 11, 2005.

[109] P. Kyasanur and N. H. Vaidya, "Routing and interface assignment in multi-channel multi-interface wireless networks," presented at IEEE Wireless Communications and Networking Conference (WCNC), New Orleans, LA, 2005.

[110] R. Chandra, P. Bahl, and P. Bahl, "MultiNet: connecting to multiple IEEE 802.11 networks using a single wireless card," March 2004.

[111] I. Ramani and S. Savage, "SyncScan: practical fast handoff for 802.11 infrastructure networks," presented at 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), Miami, 2005.

[112] F. André, J.-M. Bonnin, B. Deniau, K. Guillouard, N. Montavont, T. Noel, and L. Suciu, "Optimized Support of Multiple Wireless Interfaces within an IPv6 End-terminal," presented at Smart Objects Conference (SOC'2003), Grenoble, France, May 2003.

[113] C.-W. Ng and T. Ernst, "Multiple access interfaces for mobile nodes and networks," presented at 12th IEEE International Conference on Networks (ICON), Singapore, 2004.

[114] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks*: Prentice Hall, 2002.

[115] N. Montavont, R. Wakikawa, T. Ernst, C. Ng, and K. Kuladinithi, "Analysis of Multihoming in Mobile IPv6," in *draft-ietf-monami6-mipv6-analysis-01*: IETF MONAMI6 Working Group, June 26, 2006.

[116] K. E. Malki and H. Soliman, "Simultaneous Bindings for Mobile IPv6 Fast Handovers," in *draft-elmalki-mobileip-bicasting-v6-06*: IETF Mobile IP Working Group, July 2005, pp. IETF Draft.

[117] T. Ernst, N. Montavont, R. Wakikawa, C. Ng, and K. Kuladinithi, "Motivations and Scenarios for Using Multiple Interfaces and Global Addresses," in *draft-ietf-monami6-multihoming-motivation-scenario-00*: IETF Monami6 Working Group, February 2006, pp. IETF Draft.

[118] IEEE 802.21, "IEEE 802.21 Media Independent Handover," vol. 2006: IEEE 802.21, pp. http://www.ieee802.org/21/.

[119] IEEE 802.21 (joint contribution), "IEEE 802.21 Media Independent Handover Services," in *draft proposal for 802.21*: IEEE 802.21, May, 2005.

[120] ns-2, "The Network Simulator - ns-2," in *ns 2.29*: http://www.isi.edu/nsnam/ns/, 2006.

[121] NOAH, "NO Ad-Hoc Routing Agent," http://icapeople.epfl.ch/widmer/uwb/ns-2/noah/.

[122] The VINT Project, "The ns Manual (formerly ns Notes and Documentation)," K. Fall and K. Varadhan, Eds.: http://www.isi.edu/nsnam/ns/ns-documentation.html, August 14, 2006.

[123] Q.-T. Nguyen-Vuong, N. Agoulmine, and Y. Ghamri-Doudane, "Terminal-Controlled Mobility Management in Heterogeneous Wireless Networks," *Communications Magazine, IEEE*, vol. 45, pp. 122-129, 2007.