國立交通大學

資訊科學與工程研究所

博士論文

無線網際網路之快速換手的
安全與頻寬保留機制

Security Mechanisms and Resource Reservation Schemes for Fast Handoff in Wireless Internet

研究生：王瑞堂

指導教授：曾建超 教授

中華民國 九十七 年 四月

Security Mechanisms and Resource Reservation Schemes for Fast Handoff in Wireless Internet

Student: Jui-Tang Wang

Advisor: Dr. Chien-Chao Tseng

A Dissertation Submitted to
Department of Computer Science
College of Computer Science
National Chiao Tung University
In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Computer Science

Hsinchu, Taiwan, Republic of China

January, 2008

Abstract in Chinese

無線網際網路之快速換手的安全與頻寬保留機制

學生：王瑞堂　　　　　　　　　　指導教授：曾建超 博士

國立交通大學 資訊科學與工程研究所 博士班

摘　　要

在本論文中，我們提出兩套群體認證與金鑰分配機制（group-based authentication and key agreement scheme）、一種快速安全認證機制（fast authentication scheme）和兩種行動頻寬保留機制（mobile resource reservation scheme）。使用者在漫遊網路中針對不間斷的即時服務（real-time services）可以利用這些機制快速執行互相認證、建立安全連線與完成頻寬保留進而完成使用者的服務品質保證（Quality of Services）。

群體認證與金鑰分配機制是基於群體概念所設計出來的，使用群體概念可以減少認證訊息的傳送，也可以降低服務網路儲存空間的浪費。所以群體認證與金鑰分配機制可以加快每一個使用者認證時間，同時也避免因為認證時間過長，而造成使用者即時服務中斷。我們分別考量使用者群組與同時考量使用者群組及服務網路群組提出了兩種群體認證分配機制—以群體金鑰為基底的群體認證與金鑰分配機制與以群體簽章為基底的群體認證與金鑰分配機制。在以群體金鑰為基底的群體認證與金鑰分配機制中，一旦使用者群組裡面其中一個使用者完成認證與金鑰分配之後，相對應的服務網路會取得所謂的群體認證資料。服務網路得到群體認證資料意味著得到使用者群組的家網路的充分授權，此時，服務網路與使用者群組裡面的每一位使用者完成互相認證與金鑰分配流程而不再需要使用者家網路的加入。與以群體金鑰為基底的群體認證與金鑰分配機制最大的不同是，以群體簽章為基底的群體認證與金鑰分配機制的群體概念不僅採用使用者群組，也同時採用服務網路群組。也就是說，一旦使用者群組裡面其中一個使用者與服務

網路群組中其中一個服務網路完成互相認證與金鑰分配流程，使用者群組裡面的每一個使用者可以與服務網路群組裡面的每一個服務網路快速完成彼此互相認證與金鑰分配，而不再需要使用者家網路的參與。

針對即時服務的頻寬保證，在無線區域網路與無線網狀網路下，我們進一步提出快速認證機制，進而縮短使用者在漫遊時認證所需要的時間。在不失去安全性的情況下，使用者利用快速認證機制可以在漫遊時快速與網路端互相認證，並且同時建立安全連線，而不需要再一次執行認證與金鑰分配流程。在多躍步網路中，我們提出整合安全領域機制。整合安全領域機制以不影響 IEEE 802.11i RSN 之安全性為前提，消除安全機制於 multi-hop 網路拓樸所產生之繞送效能耗損，並以無線網狀網路(WLAN Mesh Networks)為例，降低 inter-MAP 換手延遲，使無線網狀網路可提供即時性服務更良好的 QoS 支援。

為了支援使用者漫遊，提供不間斷且保有連線品質的即時服務，我們利用行動智慧代理人（Mobile Intelligent Agent）提出了行動使用者頻寬保留機制。使用此機制可以很快的在使用者的相鄰網路建立具有品質保證的備份連線，一旦使用者漫遊到相鄰網路，可以立即使用具有品質保證的備份連線，減少換手的等待時間。這個機制不僅降低使用者連線的斷線率，同時也增加了頻寬的利用率。針對整網行動(network mobility)方面，我們也提出訊息整合機制。此機制不但自動幫助使用者維護連線品質，同時也整合在整網行動之網路的控制訊息，進而提高整網行動網路的頻寬利用率。

透過使用 G-AKA, ISD, IARSVP 或 MBA 四個機制，無線網路可提供即時性服務更良好的服務品質支援。當使用者在網路中漫遊時，透過 G-AKA 和 ISD，使用者可以與網路快速認證，建立安全連線。透過 IARSVP 和 MBA，使用者可以獲得即時服務所需要的頻寬，進而降低即時服務中斷率。我們可以在適當的時候採用 G-AKA, ISD, IARSVP 或 MBA 的機制，提供更好品質的即時性服務。

關鍵詞：整合式安全網域、快速認證、集中式無線區域網路架構、無線區域網狀網路、隨機行走模型、資源保留機制、認證機制。

Abstract

Security Mechanisms and Resource Reservation Schemes for Fast Handoff in Wireless Internet

Student: Jui-Tang Wang Advisor: Dr. Chien-Chao Tseng

Department of Computer Science
National Chiao Tung University

ABSTRACT

In this thesis, we propose several security and bandwidth mechanisms to support fast handover in wireless networks. In wireless network, a mobile node (MN) or a network moves as a whole, henceforth referred to as network mobility or NEMO [20] for short, may move from one location to another. When an MN or a NEMO enter a new location, it may need to perform authentication and key agreement (AKA), re-authentication, and resource reservations. These three processes are normally time consuming and may affect the Quality of Service (QoS) of real-time applications, such as Voice over IP (VoIP). This thesis aims to propose new mechanisms to reduce or eliminate the latency caused by the above three processes.

In order to resolve the time-consuming AKA process, we propose two Group-based AKA (G-AKA) schemes, that is, Group Key-based AKA (GK-AKA) scheme and Group Signature-based AKA (GS-AKA) scheme, to shorten authentication process. Experimental results show that G-AKA schemes not only can reduce authentication latency but also the number of signal messages between a network visited by an MN and the MN's home network. In addition, G-AKA schemes can retain the same security level as the other AKA protocols do.

For the re-authentication process, we present an integrated security domain (ISD) mechanism for multi-hop network, such as wireless LAN Mesh Networks, to reduce the re-authentication delays. The ISD mechanism integrates the security domains of an IEEE 802.11i WLAN and an IEEE 802.11s Mesh Network so that it not only can reduce the number of authentications but also eliminate the overhead caused by the link layer security protocols.

As for the resource reservations process, we propose a mobile Intelligent Agent-based Resource reSerVation apProach (IARSVP) that can support QoS aware packet transmissions for mobile IP (MIP) networks. Mobile Intelligent Agents (MIAs) are characterized by their ability to move across wide-area networks, operate autonomously on foreign hosts, and perform tasks on behalf of the originating hosts. With MIAs, IARSVP can allocate resources in advance for neighbor locations an MN may visit next. MIAs carries the mobility security association, QoS requirement and administration specification, and associated executable codes of an MN. Therefore, they can perform location updates on behalf of the MN, and adjust autonomously in accordance with the network topology and resource usage when locating the forwarding points (FP) for the MN. As a consequence, IARSVP can avoid redundant resource reservations made in common routes, support route optimization and regional registration naturally, and discover alternative routes dynamically.

Furthermore, in order to resolve the mobility unawareness and excessive signals problems for all nodes inside a NEMO, we present a Mobile Bandwidth-Aggregation (MBA) reservation scheme to support QoS guaranteed services for NEMOs. In MBA, the mobile router (MR) of a NEMO is the proxy that aggregates and reserves the bandwidth required for all nodes inside the NEMO. Mathematical analysis and simulation results show that the proposed MBA scheme can significantly reduce the signal overhead for bandwidth reservations and maintenance. Furthermore we also conduct simulation to evaluate the performance of MBA in terms of blocking probabilities and bandwidth utilizations under three different reservation policies.

With the aforementioned G-AKA and ISD schemes, an MN not only can speedup AKA and re-authentication procedures but also reduce the number of signal messages exchanged between a visiting network and the home network. Furthermore, with IARSVP and MBA, an MN or a NEMO can make resource reservations more effectively. Hopefully, these schemes can help to provide better QoS for real-time services when applicable.

Key words: Security, Authentication and Key Agreement, UMTS AKA, 802.1X, Fast Authentication Mechanism, RSVP, MRSVP, HMRSVP.

Acknowledgement

每篇博士論文的背後，一定有一位偉大的指導教授。即便我半路出家，大學與碩士班都不是專供無線網路，曾建超教授仍然有教無類的指導著我在這個領域茁壯，在學習的過程中，也讓我體認到--良好的研究態度是完成研究最重要的基礎，老師，我知道我的英文讓您多所擔心，您提醒的文法和發音，仍是我未來努力的重點。

同時要感謝曹孝櫟教授提供精闢的見解與建議，使論文的完整性得以提升。文中針對行動節點的 handoff 的分析，我們又加強的行動節點是不是經常換手或者是很少換手等等內容皆得自於曹老師的啟發。

另外還有口試委員陳耀宗教授、楊竹星教授、鄭瑞光教授、竇其仁教授老師、紀光輝教授於百忙之中抽空審閱我的論文並擔任口試委員，謝謝老師們所提供的寶貴建議讓我的論文更為完善，我對於學術領域渴求的知識，由於各位的幫助，讓我得到灌溉。

當然我不會沒有人性的忘了實驗室兄弟們對我提供的所有生活上、學術上、心靈上的幫助及建議，你們多年來陪著我歷經寫論文痛苦的瓶頸，與甜美果實的綻放；陪著我一起紓解壓力，陪著我一起分享快樂，你們從不吝給我一雙手，幫助我站上高處，看到遠方，尤其是乃心、鈺玟、人順、合翰（按照女士優先原則及筆畫順序排列）對我提供的幫助，不論在課業上或是生活上，你們都是我的營養。還有多位系辦助理，我忙碌異常的學生生活中偶有迷糊，謝謝你們給我匪淺的幫助。

最後我要感謝一直支持我的父母親與家人，這篇論文是這麼多年來他們殷殷切切對我唯一的盼望，我在此雙手呈現給他們，他們是我的骨，我的血，我一切的原點。

Contents

List of Figures

E S A
1896

List of Tables

Chapter 1

Introduction

1.1 Motivation

With the advance of wireless technologies and portable devices, it is now possible to support Voice over IP (VoIP) services or other real-time applications on wireless handsets, such as cellular phones or Personal Digital Assistors (PDAs). In order to support real-time applications in wireless networks, Authentication and Key Agreement (AKA) and Resource Reservation has become two important issues. Through AKA procedures, subscribers and network service providers can perform mutual authentication and generate keys for secure communication. With Resource Reservations, network service providers could reserve sufficient network resources for subscribers to guarantee the quality of real-time applications.

In wireless networks, a mobile node (MN) may move from one location to another, henceforth referred to as handover. During handover, an MN needs to perform an AKA procedure and make resources reservation for the new location. However, these two procedures are normally time consuming and may cause service interruption to the real-time applications. In order to avoid such service interruption during handover, several researchers have proposed schemes to reduce the latencies of AKA procedures and make effective resource reservations. However, previous solutions are still not satisfactory as described below.

1. Excessive Signal Overhead for Time-consuming AKA processes

   Current AKA protocols [1, 4] operate mostly on per-station/network basis; that is the AKA procedure treat each MN or network as an individual entity and performs authentication and key agreement between an MN and a visited network. Therefore, a network needs to perform an AKA procedure for each visiting MN, and similarly an MN needs to conduct an AKA procedure for each network it visits. This per-station/network-based AKA protocol results in multiple AKA procedures, each of which requires multiple signal messages exchanged between an MN's home and the visited network.

   In this thesis, we propose two Group-based AKA (G-AKA) schemes that can utilize the group relations among a set of MNs or among a set of networks to localize authentication or re-authentication procedures in the network visited by an MN. As a consequence, G-AKA can reduce the excessive signal messages exchanged between the MN's home network and the visited network. Furthermore, G-AKA can also reduce the authentication delay if an MN visits a network where another MN of the same group has visited previously. In addition, G-AKA can reduce the handover delay if an MN moves among the networks of the same group.

2. Long re-authentication latency in multi-hop networks

   When an MN moves from one AP to another, it needs to perform a re-authentication. The re-authentication process involves a full 802.1X authentication that is normally very time consuming and may take 750-1200 milliseconds [2]. Therefore, several researchers have proposed mechanisms [27, 7] that can either reduce or eliminate the latencies of 802.1X authentications [49]. However, each of the previous proposals has its own drawback, and still can not resolve the long authentication problem. For examples, the proposal of sharing the Pairwise Master Key (PMK) among APs will result in serious security flaws, the pre-authentication mechanism requires precisely prediction of target APs, and the introduction of new key hierarchies is not compatible with the conventional devices [49]. Furthermore, the multi-hop networks, such as wireless LAN Mesh Networks, aggravate the latencies of re-authentications further because of the hop-by-hop transmission and propagation delays.

In this thesis, we propose an approach to integrating the security domains of an 802.11i [22, 48, 49] RSN network and a multi-hop network. The approach not only can reduce the number of authentications but also eliminate the overhead caused by link layer security protocols.

3. Lack of common route discovery for mobile nodes

   Several researchers have proposed solutions (ex. MRSVP [42] and HMRSVP [45]) to QoS-guaranteed seamless handover for real-time connections. Neither MRSVP nor HMRSVP can adjust themselves in locating the Forwarding Points (FPs) dynamically in accordance with the topology or the current resource usages of the networks. In this thesis, we propose an Intelligent Agent-based ReSource reserVation aPproach (IARSVP) that can avoid redundant resource reservations made in common routes, support route optimization and regional registration automatically, and discover alternative routes dynamically.

4. Lack of effective resource reservation scheme for network mobility

   We further propose a resource reservation mechanism for a NEtwork that is MObile (NEMO) [20]. A NEMO [20] usually consists of at least one Mobile Router (MR) that attaches to the infrastructure via one or more wireless interfaces, and manages all external communication for all nodes inside a NEMO. Because a NEMO moves as a whole, MRSVP and HMRSVP posses two problems in supporting QoS for NEMOs, that is, mobility unawareness and excessive signal overhead. In order to resolve these two problems, we propose a Mobile Bandwidth-Aggregation (MBA) reservation scheme to support QoS guaranteed services for NEMOs. In MBA, an MR of a NEMO serves as the proxy of all nodes inside the NEMO, and aggregates and reserves the bandwidth required for them.

1.2 Objectives

In this thesis, we present several solutions to authentications, re-authentication and resource reservations for fast handover in wireless networks. In order to resolve the time-consuming AKA process, we propose two Group-based AKA (G-AKA) schemes, that is, Group Key-based AKA (GK-AKA) scheme and Group Signature-based AKA (GS-AKA) scheme, to shorten authentication process. Experimental results show that G-AKA schemes not only can reduce authentication latency but also the number of signal messages between a network visited by an MN and the MN's home network. In addition, G-AKA schemes can retain the same security level as the other AKA protocols do.

For the re-authentication process, we propose an integrated security domain (ISD) mechanism for multi-hop networks, such as WLAN Mesh networks, to reduce the long re-authentication delays. The ISD mechanism integrates the security domains of an IEEE 802.11i WLAN and an IEEE 802.11s Mesh Network so that it not only can reduce the number of authentications but also eliminate the overhead caused by the link layer security protocols.

As for the resource reservations process, we propose a mobile Intelligent Agent-based Resource ReSerVation protocol (IARSVP) that can support QoS aware packet transmissions for mobile IP (MIP) networks. Mobile Intelligent Agents (MIAs) are characterized by their ability to move across wide-area networks, operate autonomously on foreign hosts, and perform tasks on behalf of the originating hosts. With MIAs, IARSVP can allocate resources in advance for neighbor locations an MN may visit next. MIAs carries the mobility security association, QoS requirement and administration specification, and associated executable codes of an MN. Therefore, they can perform location updates on behalf of the MN, and adjust autonomously in accordance with the network topology and resource usage when locating the forwarding points (FP) for the MN. As a consequence, IARSVP can avoid redundant resource reservations made in common routes, support route optimization and regional registration naturally, and discover alternative routes dynamically.

Furthermore, in order to resolve the mobility unawareness and excessive signals problems for all nodes inside a NEMO, we present a Mobile Bandwidth-Aggregation (MBA) reservation scheme to support QoS guaranteed services for NEMOs. In MBA, the mobile router (MR) of a NEMO is the proxy that aggregates and reserves the bandwidth required for all nodes inside the NEMO. Mathematical analysis and simulation results show that the proposed MBA scheme can significantly reduce the signal overhead for bandwidth reservations and maintenance. Furthermore we also conduct simulation to evaluate the performance of MBA in terms of blocking probabilities and bandwidth utilizations under three different reservation policies.

With the aforementioned G-AKA and ISD schemes, an MN not only can speedup AKA and re-authentication procedures but also reduce the number of signal messages exchanged between a visiting network and the home network. Furthermore, with IARSVP and MBA, an MN or a NEMO can make resource reservations more effectively. Hopefully, these schemes can help to provide better QoS for real-time services when applicable.

1.3 Synopsis

The remainder of this thesis is organized as follows. Chapter 2 introduces related works. Chapter 3 presents two group-based AKA schemes, including group key-based AKA scheme and group signature-based AKA scheme. Chapter 4 introduces the ISD approach to reduce re-authentication overheads for multi-hop networks. Chapter 5 describes IARSVP and MBA resource reservation schemes, respectively, for mobile IP networks and network mobility. Finally, chapter 6 presents the conclusions and future works of this research.

Chapter 2

Related Works

In this section, will describe some related works on authentication and key agreement process, fast authentication mechanisms and mobile resource reservation mechanisms.

2.1 Related work on authentication and key agreement

Here we outline the essence of AKA mechanisms in widespread use by mobile telecommunication networks. For a better exposition we take UMTS AKA [1] as a representative exemplifying their design tenet.

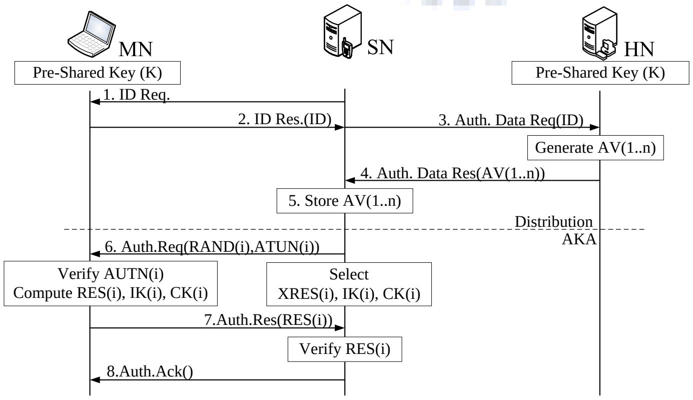


Figure 2.1: UMTS AKA message flow

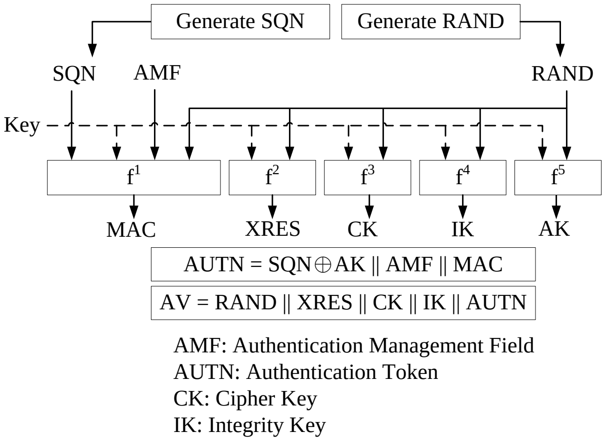

Generate SQN
Generate RAND
SQN
AMF
RAND
Key
f1
f2
f3
f4
f5
MAC
XRES
CK
IK
AK
AUTN = SQN⊕AK || AMF || MAC
AV = RAND || XRES || CK || IK || AUTN
AMF: Authentication Management Field
AUTN: Authentication Token
CK: Cipher Key
IK: Integrity Key

Figure 2.2: UMTS AKA operations on the HN side

UMTS AKA can broadly be divided into two stages: authentication data distribution, and user authentication and key agreement (represented above and below, respectively, the dashed line shown in Fig. 2.1).

The former enables the Home Network (HN) of an Mobile Node (MN) to distribute authentication data to the Serving Network (SN) the MN visits currently. The latter is to establish a new pairwise cipher key and integrity key between the MN and the SN. Overall, UMTS AKA consists of several message exchanges in following lines. In what follows a parenthetical term in any message represents some specific information or parameter to be conveyed in the payload.

1. ID Request: Upon detecting an access request by an MN, the SN initiates an authentication procedure by asking the MN for its identity.

2. ID Response: The MN sends its identify (ID) to the SN.

3. Authentication Data Request: The SN sends this message to acquire n Authentication Vectors AV$(1 \ldots n)$ from the HN. The operation of generating each attribute in AV is depicted in Fig. 2.2.

   In addition to Authentication Management Field (AMF) and Sequence Number (SQN), a Pre-Shared key for the MN and a random number RAND are parameters taken to generate AV(i) comprising the Message Authentication

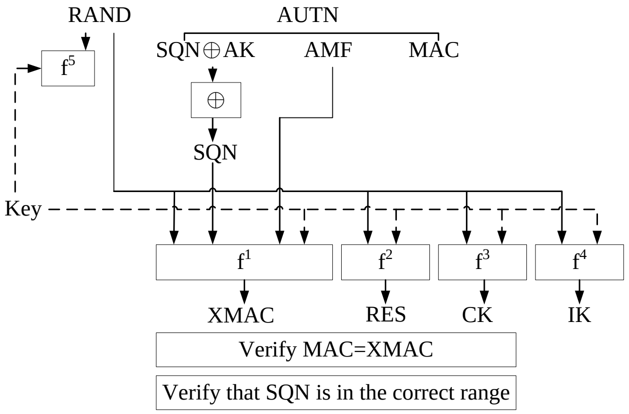


Figure 2.3: UMTS AKA operations on the MN side

Code (MAC), eXpected Response (XRES), Cipher Key (CK), Integrity Key (IK), Anonymity Key (AK), and Network Authentication Token (AUTN).

4. Authentication Data Response: The HN sends back the generated Authentication Vectors (AVs) for the corresponding MN so that the SN is authorized to authenticate the requesting MN.

5. User Authentication Request: Upon receipt of a message containing authentication vectors, the SN sends RAND(i) and AUTN(i) of the i-th selected vector to the MN, enabling the MN to verify the correctness of SQN and compute the proportional response RES(i).

6. User Authentication Response: The MN validates the correctness of SQN by computing XMAC and comparing it with the MAC carried in AUTN(i). If matched, the MN computes and sends the proportional response RES(i) back to the SN in a response message.

7. Authentication Result: Once RES(i) is received and verified correctly, the SN chooses the corresponding CK/IK as the session key to protect data communication. In the meantime, the MN computes its CK/IK accordingly. Hence both the MN and SN reach a common session key, which terminates the UMTS AKA protocol.

Apart from UMTS AKA, other protocols such as UMTS X-AKA [26] have also been devised to reduce signaling traffic in some extent. UMTS X-AKA takes a step further by authorizing an SN to authenticate an MN and establish a master key MK locally with the MN, after the HN has authenticated the MN. Instead of sending authentication vectors to the SN (message 4 in Fig. 2.1), the HN in UMTS X-AKA sends the authentication data of an MN to the SN and authorizes the SN to authenticate the MN henceforward. Later when the SN needs to re-authenticate the MN, the SN can accomplish authentication locally, without the participation of the HN.

However, UMTS X-AKA and other conventional AKA protocols operate mostly on per-station/network basis. When multiple MNs of a group visit an SN, current protocols require the SN to initiate multiple authentication processes for different MNs with the same HN, causing nontrivial overhead during handoff.

2.2 Related worked on fast authentication

The security mechanism of 802.11 WLAN network and 802.11s Wi-Fi mesh network consist of 802.1X authentication, 4-way handshake and encryption protocols, i.e. TKIP and CCMP. The security mechanism is designed to establish a Robust Security Network Association (RSNA) between an MN and an MAP for securing the wireless connection. Due to the delay of the ratification to 802.11i, Wi-Fi Protected Access (WPA), a subset of the 802.11i standard, is adopted by the Wi-Fi Alliance as a transitional solution to WEP insecurities. WPA2 is the full implementation of the 802.11i standard and provides a robust security protocol for WLAN.

The issue of 802.1X authentication has been addressed. Some methods have been proposed to mitigate the overhead of 802.1X authentication and improve the handoff performance. For a better exposition we take pre-authentication mechanism proposed by 802.11i as a representative exemplifying their design tenet.

With the pre-authentication mechanism proposed by 802.11i, an MN is allowed to perform 802.1X authentication with a new AP before associating with it. The

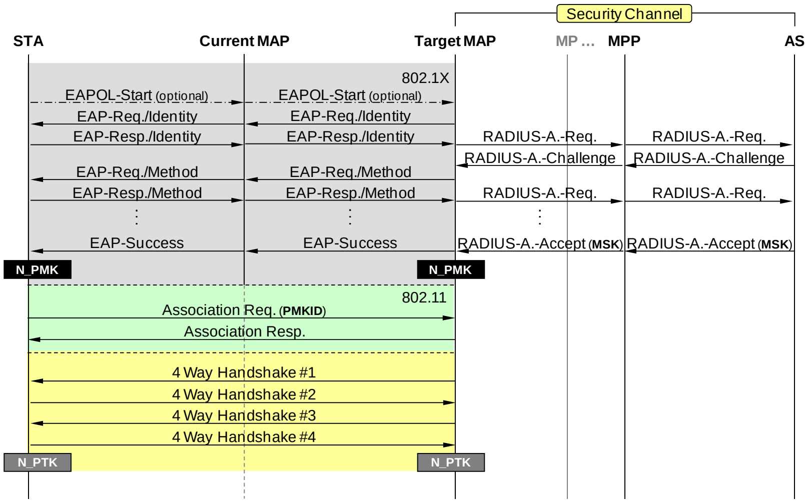


Figure 2.4: 802.11i pre-authentication

pre-authentication separates the commit phase from the authentication phase and permits them to be performed independently.

Because an MN can only associate with an AP at a time, pre-authentication messages transmitted between the MN and the target AP are forwarded via the current AP. Fig. 2.4 shows the procedures of the 802.11i pre-authentication.

1. 802.11i pre-authentication start with an EAPOL-Start message sent from the MN to the target MAP. Except all EAPOL messages are forwarded via the current MAP, the procedures are identical to the 802.1X authentication.

2. After the pre-authentication complete, a new PMK (N_PMK) is derived. The MN and the target MAP cache this N_PMK for further usage.

3. While associating with the target MAP, the MN inserts the PMKID into the Association Request frame to indicate that the N_PMK is cached. If the PMKID is valid, the target MAP will skip 802.1X authentication and initialize 4-way handshake directly.

The benefit of pre-authentication is that the 802.1X authentication is independent of the handoff procedures. An MN is able to authenticate with multiple candidate APs. However, pre-authentication establishes authentication state and key management state on both the MN and candidate APs. To avoid the storage cost and the overhead of pre-authentication burden AP and AS, MN should only pre-authenticate with the AP that it is most likely to handoff to. Therefore, precisely target AP prediction is necessary for MN to perform pre-authentication efficiently.

On the negative side, pre-authentication is expensive in terms of computational power and latency for MN. The 802.1X authentication latency actually is not reduced by the pre-authentication. To guarantee the QoS for real-time applications, an MN needs to perform pre-authentication early enough before the current connection is dropped. Thus, well designed and overlapping coverage areas are essential to perform pre-authentication successfully. In addition, pre-authentication introduces opportunities for DoS attacks. Malicious MNs could burden the AS by pre-authenticating with a large number of APs.

2.3 Related worked on mobile resource reservation

In this section, we will describe the essence of mobile resource reservation mechanisms in wireless Internet. For a better exposition we take Mobile RSVP [42] and Hierarchical Mobile RSVP [45] as representative exemplifying their design tenet.

2.3.1 Mobile RSVP

Mobile ReSource reserVation Protocol (MRSVP) [42] is a mobile resource reservation protocol proposed by Talukdar et al to support multimedia delay-sensitive streaming applications in wireless integrated services Internet [8]. The resource reservations made by MRSVP can be classified as either active or passive. An active reservation is on the path to the MN's current location and has an actual data flow, whereas a passive reservation is on the path to a neighbor location of the MN and does not have an actual data flow.

MRSVP attempts to make passive reservations in advance at the neighbor lo-

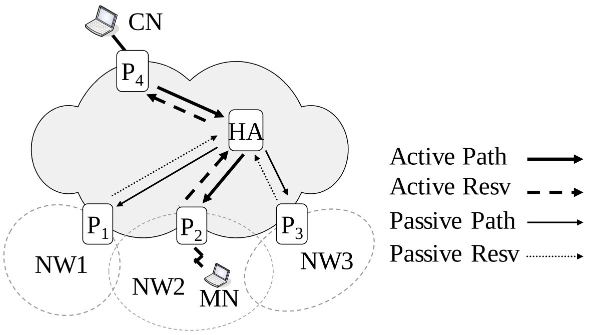


Figure 2.5: MRSVP Active and Passive Reservations

cations where an MN may visit next. If MRSVP reserves the resources for the MN successfully in a neighbor location, the MN can use the pre-reserved resources directly without issuing further resource reservation requests upon entering the location.

MRSVP introduces proxy agents to make resource reservation for the MNs. A proxy agent at an MN's current location is called the *local proxy agent* of the MN; the proxy agents at the MN's neighbor locations are called *remote proxy agents*. The local proxy agent of an MN is responsibility for setting up an active resource reservation whereas the remote proxy agents of an MN are responsibility for making passive reservations. As shown in Fig. 2.5, a receiver MN, its HA and a corresponding sender node CN exchange Path and Resv messages to establish an active reservation and several passive reservations. In this case, we assume Mobile-IP reversed tunneling [34] is in use and HA serves as the anchor node of MN. Therefore, HA aggregates the MN's active and all passive reservations, and makes only a single reservation with CN.

However, MRSVP reserves resources in advance along the path from a CN or the home agent of an MN to each subnet the MN may visit next. These excessive resource reservations may waste too much bandwidth and degrade the network performance.

With passive reservations, an MN can use resources immediately upon entering a new network. However, MRSVP may make unnecessary resource reservations if it reserves bandwidth in advance at all neighbors, and cannot provide mobile resource reservation mechanism for all nodes inside a NEMO [20].

2.3.2 Hierarchical Mobile RSVP

Hierarchical Mobile RSVP (HMRSVP) [45] adopts the hierarchical concept of Mobile-IP regional registration [24] and makes resource reservation in advance for an MN only when the MN resides in the overlapped area of the boundary cells between two regions. A region could be an enterprise or a campus network that consists of a set of routers or subnets. In the base Mobile-IP protocols [37, 29], an MN registers with the MN's HA each time when it changes its point of network attachment. In the cases where the HA is far away, the registration process may become too expensive. In order to support low latency and smooth handoff, Mobile-IP regional registration localizes the registration process within a region when an MN makes an intra-region movement, by organizing Mobility Agents (MAs) in a region hierarchically in accordance with the routing topology of the region. The MA situated on the region gateway is called the Gateway Mobility Agent (GMA) of the region. Due to the hierarchical nature and network-based routing properties of Internet, GMA and regional MAs together can process registrations locally and hide the mobility of a visiting MN from the MN's HA and CNs when the MN moves within the region. Because the reservation setup time within a region is likely to be short, HMRSVP does not make advance resource reservations when MNs move within the region. In stead, HMRSVP makes passive reservations in advance only when an MN may make an inter-region movement when the MN resides in the overlapped area of the boundary cells of two regions.

However, HMRSVP possesses the same bandwidth waste problem for inter-region handovers, which occur when an MN makes a move from one region to another, and still cannot solve the mobility unawareness and excessive signals problems for all nodes inside a NEMO.

Chapter 3

Group Authentication and Key Agreement schemes

As aforementioned, UMTS AKA [1] and other conventional AKA protocols [25, 26] withstand various attacks and operate mostly on per-station/network basic. Through effective, these protocols incur nontrivial communication latency due to potentially prohibitive messages between different network domains.

We note that MNs belonging to the same HN often communicate in a form of group. Such a group is likely to migrate somewhere together. That is, MNs of an HN may visit the same city or country traveling from one place to another, students having a field trip, or mobile routers on a public transportation system. In this text, SNs the concerned group of MNs visit collectively on a route are viewed to form an SN group. Such group-based movement behavior causes many authentication processes within a short period of time, at the expense of significant signaling traffic between the SN and the HN if a traditional protocol like UMTS AKA is used to authentication MNs separately. As a result, signaling overhead between the HN and the SN grows with the involvement of more MNs.

In this chapter, we present two Group-based AKA (G-AKA) scheme to speedup authentication process and improve QoS of real-time connection. The G-AKA scheme consists of Group Key-based AKA (GK-AKA) scheme for MNs and Group

Signature-based AKA (GS-AKA) scheme for both MNs and SNs. In GK-AKA, when the first MN of a group visits, the SN performs a full authentication with the concerned HN and thereby obtains authentication information for the MN and other members. Thus when any other MN of the same group visits, the SN can authenticate locally with them without subsequent involvement of the HN, so as to reduce signaling overhead. In GS-AKA, except the first MN, can perform authentication locally with SNs of an SN group. Therefore, when an MN moves from one SN to another, the MN and the new SN could perform AKA locally without any SN prediction algorithm. The group manager HN also setup group authentication information directly into corresponding SN group in advance if HN knows MN group behavior. Therefore, the G-AKA can reduce signaling message exchange between SN (or SNs) and the HN, decrease authentication delay and speedup the handoff process. Meanwhile, the proposed G-AKA scheme can still guarantee the same secure level as UMTS AKA does.

3.1 Group Key-based AKA Scheme

In this section, we present a group key-based AKA (GK-AKA) scheme addressing the scenario where roaming users with subscribership in a common HN visit a network. In our GK-AKA scheme, every MN provides its identity when visiting an SN. Upon reception, the SN examines whether the MN belongs to an active group of which any member has completed full authentication. If not, the SN acquires authentication data for the MN and its associated group from the concerned HN. This leads MNs of the same group to share the same authentication data including group temporary authentication key (GTK) and other necessary information. To realize, our scheme comprises three procedures: group information setup, authentication data distribution, and mutual authentication and key agreement, as shall be described in following three subsections.

Table 3.1: The index Table in GK-AKA

Group	Group ID	Member ID	Initial Value	Other Information
G1	ID$_{\mathrm{G1}}$	ID$_{\mathrm{M1\text{-}1}}$	IV$_{\mathrm{M1\text{-}1}}$	...
		ID$_{\mathrm{M1\text{-}2}}$	IV$_{\mathrm{M1\text{-}2}}$	...
		$\ldots$	$\ldots$	$\ldots$
		ID$_{\mathrm{M1\text{-}}n}$	IV$_{\mathrm{M1\text{-}}n}$	...
G2	ID$_{\mathrm{G2}}$	ID$_{\mathrm{M2\text{-}1}}$	IV$_{\mathrm{M2\text{-}1}}$	...
		...	...	...

3.1.1 Group Information Setup

In our architecture the HN sends to an SN authentication data for a group of MNs, as opposed to sending authentication data for respective MNs. Initially, the HN configures group information of MNs, including an index table and GAK. As shown in Table 3.1, the index table contains fields of group identity, member identity, initial value (IV_i) for each member i, and other information. For the convenience of illustration, we denote the group in the first entry as $G1$, group in the second entry as $G2$, and so forth. We let the initial value IV_i be large and unique. Besides, IV_i behaves as a sequence number for synchronization between the MN and the SN.

The HN also assigns each MN an individual key for communication confidentiality, and each group a common group key, namely GAK, for authentication purpose. The generation and distribution of GAKs along with MNs joining or leaving a group can be managed by the Authentication Center (AuC) within the home network [47, 46, 43]. Furthermore, the HN, SNs and MNs contain MAC algorithms for authenticating messages. The inputs for MAC algorithms consist of a secret key and some related information, and outputs generated by MAC algorithms are irreversible. Without loss of generality, we denote MAC algorithms as f^0, f^1, f^2, and f^3, respectively, for the HN to authenticate an MN, for an MN to authenticate an SN, for an SN to authenticate an MN, and for key generation.

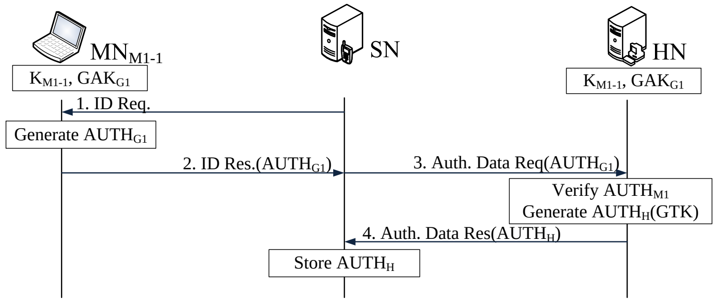


Figure 3.1: Authentication data distribution

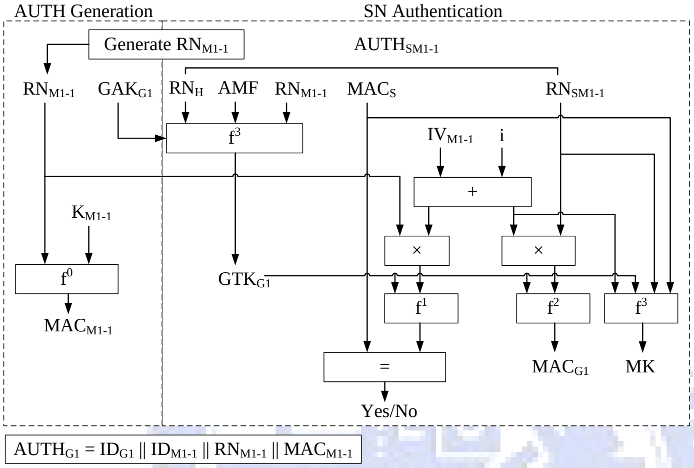


Figure 3.2: An MN generating $\mathrm{AUTH_{G1}}$, verifying $\mathrm{MAC_S}$ and generating $\mathrm{MAC_{G1}}$.

3.1.2 Authentication Data Distribution

We now consider how the HN distributes authentication data for some MNs of the same group migrating to an SN. Let $MN_{\text{M1-1}}$ be the first MN initiating authentication in the roaming group $G1$. Furthermore, in what follows a parenthetical term in any message represents some specific information to be conveyed in the payload. The distribution procedure is shown in Fig. 3.1, where challenge-response messages can be embodied by CHAP [40], in following lines.

1. ID Request: The SN attempts to identify $\mathrm{MN_{M1\text{-}1}}$.

2. ID Response ($\mathrm{AUTH_{G1}}$): Upon receiving the ID Request message, $\mathrm{MN_{M1\text{-}1}}$ generates $\mathrm{AUTH_{G1}} = (\mathrm{ID_{G1}} \| \mathrm{ID_{M1\text{-}1}} \| \mathrm{RN_{M1\text{-}1}} \| \mathrm{MAC_{M1\text{-}1}})$, where $\mathrm{ID_{G1}}$ denotes the group identity, $\mathrm{ID_{M1\text{-}1}}$ is the station's identity, $\mathrm{RN_{M1\text{-}1}}$ represents a ran-

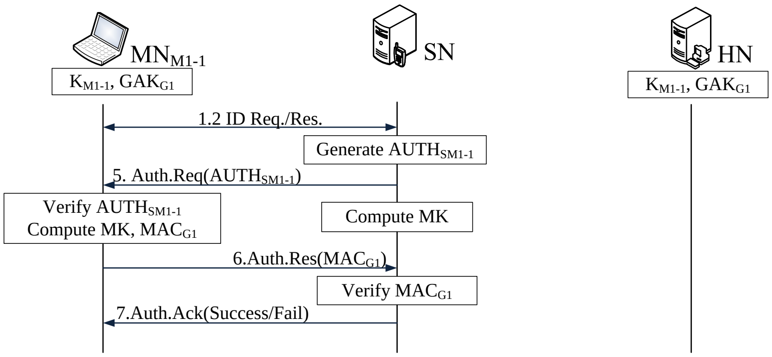


Figure 3.3: Mutual authentication and key agreement

dom number, and $\text{MAC}_{\text{M1-1}} = f^0(\text{K}_{\text{M1-1}}, \text{RN}_{\text{M1-1}})$ for the HN to authenticate $\text{MN}_{\text{M1-1}}$ (see also Fig. 3.2.) Here $\text{K}_{\text{M1-1}}$ is the pre-shared secret key with the HN.

3. Authentication Data Request (AUTH_{G1}): Since $\text{MN}_{\text{M1-1}}$ is new, the SN without knowledge of the MN relays the foregoing message from $\text{MN}_{\text{M1-1}}$ to the HN. The HN shall authenticate the roaming group ($G1$) which $\text{MN}_{\text{M1-1}}$ belongs to.

4. Authentication Data Response (AUTH_{H}): As shown in Fig. 3.4, the HN verifies the received $\text{MAC}_{\text{M1-1}}$ in AUTH_{G1} using $\text{K}_{\text{M1-1}}$ (the pre-shared key with $\text{MN}_{\text{M1-1}}$.) If $\text{MN}_{\text{M1-1}}$ is found authentic, the HN retrieves the corresponding group authentication key GAK_{G1} to generate a Group Transient Key GTK_{G1} $= f^3(\text{RN}_{\text{M1-1}} \| \text{RN}_{\text{H}} \| \text{AMF} \| \text{GAK}_{\text{G1}})$.

Group authentication data sent to the SN contains $\text{AUTH}_{\text{H}} = (\text{RN}_{\text{H}} \| \text{AMF} \| \text{RN}_{\text{M1-1}} \| \text{GTK}_{\text{G1}})$, where RN_{H} is a newly selected random number by the HN, AMF denotes contents of the Authentication Management Field, and $\text{RN}_{\text{M1-1}}$ is the random number chosen a priori by $\text{MN}_{\text{M1-1}}$. The group information for $G1$ is piggy-backed in this message. The SN will keep the received AUTH_{H} in local storage for future use.

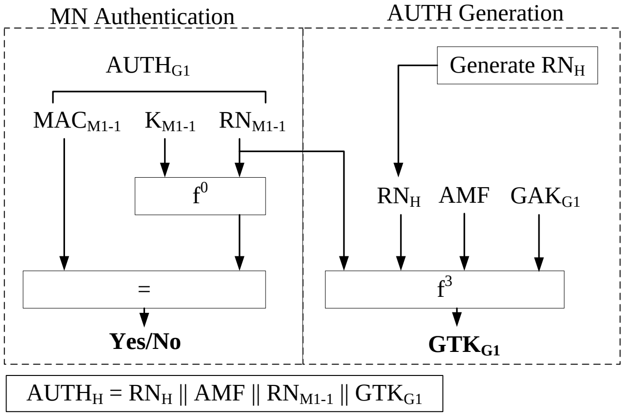


Figure 3.4: The HN verifying AUTH_{G1} and generating AUTH_{H}.

3.1.3 Mutual Authentication and Key Agreement

Upon receiving group authentication data, the SN starts on straightway a procedure of mutual authentication and key agreement with $\text{MN}_{\text{M1-1}}$. This procedure is to achieve mutual authentication and to establish the pairwise session key for message encryption between an MN and an SN. As a result of computing different responses of challenge messages with different arguments, both the MN and the SN can identity each other by verifying the correctness of responses. If both sides are successfully authenticated, the session key is generated to protect the traffic between the MN and the SN. This procedure is depicted as Fig. 3.3, starting with Message 5.

5. Authentication Request ($\text{AUTH}_{\text{SM1-1}}$): After acquiring AUTH_{H} for group $G1$, the SN initiates the i-th run of mutual authentication with $\text{MN}_{\text{M1-1}}$ by generating $\text{AUTH}_{\text{SM1-1}} = (\text{AMF}\|\text{RN}_{\text{H}}\|\text{RN}_{\text{M1-1}}\|\text{MAC}_{\text{S}}\|\text{RN}_{\text{SM1-1}})$, where first three parameters are meant for $\text{MN}_{\text{M1-1}}$ to generate GTK_{G1}, $\text{MAC}_{\text{S}} = f^{1}(\text{GTK}_{\text{G1}}\|\text{RN}_{\text{M1-1}}\|\text{IV}_{\text{M1-1}}+i)$, and $\text{RN}_{\text{SM1-1}}$ is a nonce chosen by the SN to challenge $\text{MN}_{\text{M1-1}}$ later. While waiting for the response from $\text{MN}_{\text{M1-1}}$, the SN computes the Master Key $\text{MK} = f^{3}(\text{GTK}_{\text{ G1}}\|\text{IV}_{\text{M1-1}} + i\|\text{RN}_{\text{M1-1}}\|\text{RN}_{\text{SM1-1}})$ for subsequent sessions with $\text{MN}_{\text{M1-1}}$ in advance. (See Fig. 3.5.)

6. Authentication Response (MAC_{G1}): In response to Authentication Request ($\text{AUTH}_{\text{SM1-1}}$), $\text{MN}_{\text{M1-1}}$ computes GTK_{G1} using first three arguments in $\text{AUTH}_{\text{SM1-1}}$

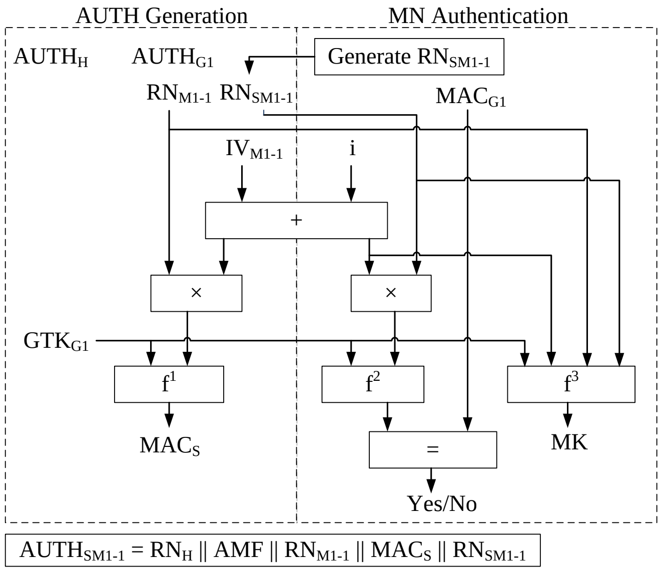


Figure 3.5: An SN generating $\text{AUTH}_{\text{SM1-1}}$ and MK and verifying MAC_{G1}.

and GAK_{G1} stored in each MN of the same group. $\text{MN}_{\text{M1-1}}$ then authenticates the SN by computing and comparing the corresponding result with MAC_{S}. After successfully authenticating the SN, $\text{MN}_{\text{M1-1}}$ calculates the Master Key MK with respect to the SN and generates a message back to the SN containing $\text{MAC}_{\text{G1}} = f^2(\text{GTK}_{\text{G1}} \| \text{RN}_{\text{SM1-1}} \| \text{IV}_{\text{M1-1}} + i)$. Such operations are diagrammed in Fig. 3.2.

7. Authentication Result (Success/Failure): Upon receiving an Authentication Response message carrying MAC_{G1}, the SN checks whether $\text{MN}_{\text{M1-1}}$ has produced the correct response using operations as in Fig. 3.5. Then a message with a status code indicating either success or failure for mutual authentication is sent to $\text{MN}_{\text{M1-1}}$, whence our key agreement procedure is completed.

After full authentication, both $\text{MN}_{\text{M1-1}}$ and its SN share a common MK that can be employed as the material for subsequent key derivations.

When a second group member, say $\text{MN}_{\text{M1-2}}$, arrives and requests for authentication, the SN simply initiates mutual authentication and key agreement with $\text{MN}_{\text{M1-2}}$ using the existing GTK_{G1}. In other words, messages 3 and 4 in the prescribed authentication data distribution procedure can be bypassed, leaving out signaling traffic between SN and HN. In this regard, however, the SN needs to generate a new random number $\mathrm{RN}_{\mathrm{SM1\text{-}2}}$ to create a new challenge message for $\mathrm{MN}_{\mathrm{M1\text{-}2}}$ (message 5 in Fig. 3.3.) Using distinct arguments from those for $\mathrm{MN}_{\mathrm{M1\text{-}1}}$, such as $\mathrm{RN}_{\mathrm{SM1\text{-}2}}$, $\mathrm{RN}_{\mathrm{M1\text{-}2}}$ and $\mathrm{IV}_{\mathrm{M1\text{-}2}}$, our scheme not only assures the freshness of challenge-response messages but also guarantees the uniqueness of MKs for different MNs.

As a remark, the proposed G-AKA scheme reduces messages sent from the HN to the same destination SN repeatedly by providing group authentication data and GTK. The latter is used in place of GAK to prevent the original GAK from being divulged to eavesdroppers. Observe that a GTK allows of periodic updates whenever new random numbers are provided.

3.2 Group Signature-based AKA Scheme

Furthermore, in this section we present a Group Signature-based AKA (GS-AKA) scheme addressing the scenario where roaming users with subscribership in a common HN visit a group of networks. When MNs of a group visit an SN, only the first MN is required to perform full AKA with the HN; the SN is allowed to carry out AKA with (transparent to) other MNs locally. Furthermore, when an MN migrates from an SN to another within the same SN group, the MN need not perform a full AKA procedure; instead, the MN and the new SN can perform AKA mutually to speed up the handoff process.

GS-AKA adopts the concepts of group signatures to reduce repeated full authentications in support of fast handoffs. Group signature was first introduced by Chaum and Heyst [16], in which each group consists of a number of *members*, *verifiers*, and a unique *manager*. Each member of a group has a unique group sign key (*gsk*) to generate a group signature, a verifier has a common group public key (*gpk*) to verify the signature generated by a group member, and the manager has a unique group master secret key (*gmsk*) to identify the member who has generated the signature [6, 13, 12, 14, 16, 17, 5, 41]. In GS-AKA, the HN acts as the group manager, whereas SNs are verifiers of MNs and *vice versa*.

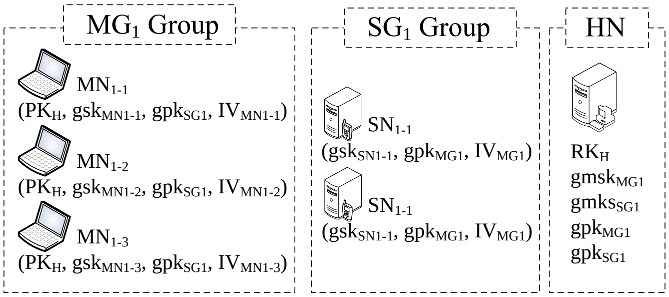

MG1 Group
MN1-1
(PKH, gskMN1-1, gpkSG1, IVMN1-1)
MN1-2
(PKH, gskMN1-2, gpkSG1, IVMN1-2)
MN1-3
(PKH, gskMN1-3, gpkSG1, IVMN1-3)
SG1 Group
SN1-1
(gskSN1-1, gpkMG1, IVMG1)
SN1-1
(gskSN1-1, gpkMG1, IVMG1)
HN
RKH
gmskMG1
gmksSG1
gpkMG1
gpkSG1

Figure 3.6: An example for Group Information Setup

To realize, GS-AKA comprises three procedures: 1) group information setup, 2) authentication data distribution, and 3) mutual authentication and key agreement, as shall be described in following three subsections.

3.2.1 Group Information Setup

Both MN and SN groups could be assigned in a predetermined fashion by network operators. A network operator may group MNs when mobile users start their subscriptions to some services, but group SNs in accordance with the administrative domains. Further, we assume that SNs are trusted and a secure communication channel exists between an SN and the MN's HN.

Fig. 3.6 shows an example for group information setup for an MN group ($\mathrm{MG_1}$) of three MNs, $\mathrm{MN_{1\text{-}1}}$, $\mathrm{MN_{1\text{-}2}}$ and $\mathrm{MN_{1\text{-}3}}$, and an SN group ($\mathrm{SG_1}$) of two SNs, $\mathrm{SN_{1\text{-}1}}$ and $\mathrm{SN_{1\text{-}2}}$. Each MN (or SN) has a unique ID and shares a group ID. For each group, the HN configures one *gmsk*, one *gpk*, and a unique *gsk* for each member of the group. Each member can use its *gsk* to generate a group signature. The verifier can then use *gpk* to verify the signature of a group member, whereas the HN can use *gmsk* to identify the member that generates the signature. Hence an MN can use the HN's public key ($\mathrm{PK_H}$) to verify the signature which the HN generates using its private key ($\mathrm{RK_H}$). Furthermore, the HN assigns each MN a unique initial value (IV) of the authentication sequence. With IVs, a verifier SN can authenticate an MN unambiguously and reject an MN if the MN pretends to be another MN of the same group.

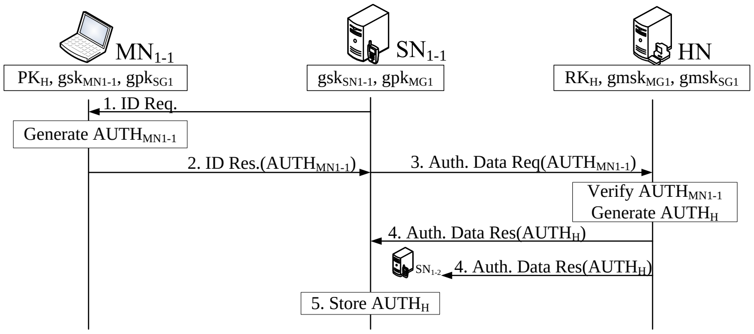

MN$_{1\text{-}1}$
SN$_{1\text{-}1}$
HN
PK$_H$, gsk$_{MN1\text{-}1}$, gpk$_{SG1}$
gsk$_{SN1\text{-}1}$, gpk$_{MG1}$
RK$_H$, gmsk$_{MG1}$, gmsk$_{SG1}$
1. ID Req.
Generate AUTH$_{MN1\text{-}1}$
2. ID Res.(AUTH$_{MN1\text{-}1}$)
3. Auth. Data Req(AUTH$_{MN1\text{-}1}$)
Verify AUTH$_{MN1\text{-}1}$
Generate AUTH$_H$
4. Auth. Data Res(AUTH$_H$)
SN$_{1\text{-}2}$
4. Auth. Data Res(AUTH$_H$)
5. Store AUTH$_H$

Figure 3.7: Authentication Data Distribution

Network operators can configure MNs and SNs beforehand with MN group and SN group information including its IV and security material, respectively, through secure communication channels. However, MNs learn the group public key gpk_{SG_i} of an SN group and SNs receive the group public key (gpk_{MG_i}) and IVs of an MN group dynamically from the HN during the authentication data distribution process.

3.2.2 Authentication Data Distribution

Assume that MN$_{1\text{-}1}$ of MG$_1$ is the first MN that visits SN$_{1\text{-}1}$ of SG$_1$. SN$_{1\text{-}1}$ will request the Authentication Data of MG$_1$ from the HN. However, in addition to SN$_{1\text{-}1}$, the HN also distributes the authentication data of MG$_1$ to other SNs of SG$_1$, as shown in Fig. 3.7.

Such pre-distribution of authentication data enables other SNs to authenticate MN$_{1\text{-}1}$ locally for fast handover. Furthermore, when another MN of MG$_1$ visits an SN of SG$_1$, the MN and the SN can authenticate each other locally without the participation of the HN.

The authentication data distribution process operates in following steps.

1. ID Request: SN$_{1\text{-}1}$ sends an ID Request message to MN$_{1\text{-}1}$ for identification.

2. ID Response (AUTH$_{MN_{1\text{-}1}}$): MN$_{1\text{-}1}$ first generates a random number RN$_{MN_{1\text{-}1}}$ and computes a group signature GSIG$_{MN_{1\text{-}1}}$ $= GSig(gsk_{MN_{1\text{-}1}}, \mathrm{RN}_{MN_{1\text{-}1}})$, where $GSig()$ is a group sign function. MN$_{1\text{-}1}$ then replies with an ID Re-

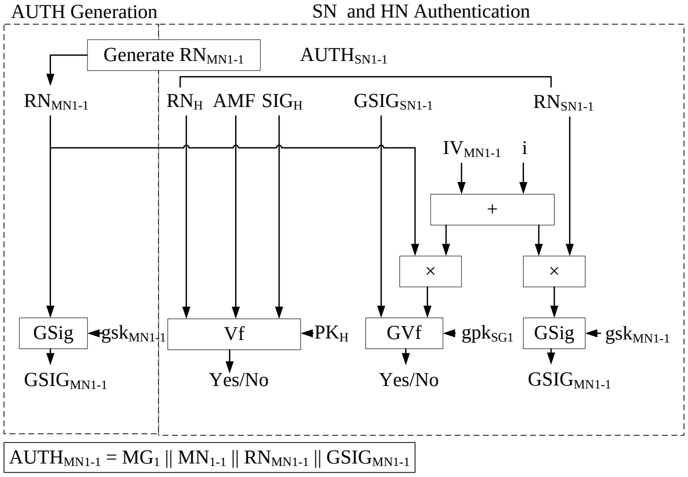


Figure 3.8: An MN generating $GSIG_{MN_{\text{1-1}}}$ and verifying $\text{AUTH}_{SN_{\text{1-1}}}$.

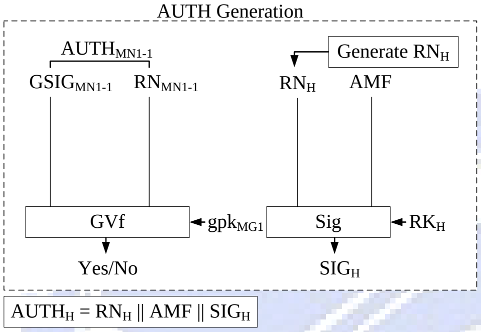


Figure 3.9: $\text{AUTH}_{MN_{\text{1-1}}}$ verification and AUTH_{H} generation in the HN.

sponse ($\text{AUTH}_{MN_{\text{1-1}}}$) message, where $\text{AUTH}_{MN_{\text{1-1}}}$ consists of its Group ID (MG_1), individual ID ($\text{MN}_{\text{1-1}}$), $\text{RN}_{MN_{\text{1-1}}}$, and $\text{GSIG}_{MN_{\text{1-1}}}$. See also Fig. 3.8.

3. Authentication Data Request ($\text{AUTH}_{MN_{\text{1-1}}}$): Since $\text{MN}_{\text{1-1}}$ is new, $\text{SN}_{\text{1-1}}$ without knowledge of $\text{MN}_{\text{1-1}}$ relays the forgoing message from $\text{MN}_{\text{1-1}}$ to the HN. The HN shall authenticate the roaming MN group, say MG_1, to which $\text{MN}_{\text{1-1}}$ belongs.

4. Authentication Data Response (AUTH_{H}): As shown in Fig. 3.9, the HN first checks the freshness of the random number $\text{RN}_{MN_{\text{1-1}}}$ and then verifies the signature $\text{GSIG}_{MN_{\text{1-1}}}$. If $\text{MN}_{\text{1-1}}$ is found authentic, the HN generates a nonce RN_{H} for challenging MN and computes a signature $\text{SIG}_{\text{H}} = Sig(\text{RK}_{\text{H}}, \text{RN}_{\text{H}}\|\text{AMF})$,

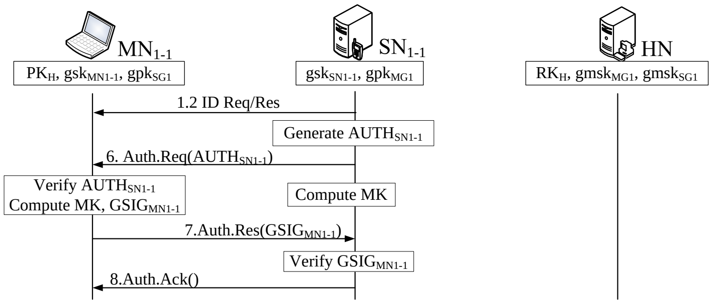


Figure 3.10: Mutual Authentication and Key Agreement

where $\mathrm{RK_H}$ is the private key of the HN, AMF is the Authentication Management Field, and $Sig()$ is single user sign function. After that, the HN sends the group authentication data ($\mathrm{AUTH_H} = \mathrm{RN_H} \parallel \mathrm{AMF} \parallel \mathrm{SIG_H} \parallel gpk_{MG_1} \parallel IV_{MG_1}$) to all members of $\mathrm{SG_1}$, namely $\mathrm{SN_{1\text{-}1}}$ and $\mathrm{SN_{1\text{-}2}}$, for local authentications and fast handoff.

5. Store Authentication Data: $\mathrm{SN_{1\text{-}1}}$ and $\mathrm{SN_{1\text{-}2}}$ store $\mathrm{AUTH_H}$ for later use.

3.2.3 Mutual Authentication and Key Agreement

Upon receiving group authentication data, $\mathrm{SN_{1\text{-}1}}$ starts on a procedure of mutual authentication (or re-authentication) and key agreement with $\mathrm{MN_{1\text{-}1}}$ locally. This procedure can also establish a common Master Key (MK), with which both $\mathrm{SN_{1\text{-}1}}$ and $\mathrm{MN_{1\text{-}1}}$ can generate integrity and cipher keys to secure communication in-between. As a result of computing different responses of challenge messages with different arguments, both the MN and the SN can identify each other by verifying the correctness of response. If both sides are successfully authenticated, the master key is generated to protect the traffic between the MN and the SN.

Fig. 3.10 shows how Mutual Authentication and Key Agreement are achieved in our architecture, in following steps:

6. Group Authentication Request ($\mathrm{AUTH}_{SN_{1\text{-}1}}$): After acquiring $\mathrm{AUTH_H}$ for group $\mathrm{MG_1}$, $\mathrm{SN_{1\text{-}1}}$ initiates the i-th mutual authentication between $\mathrm{SN_{1\text{-}1}}$ and

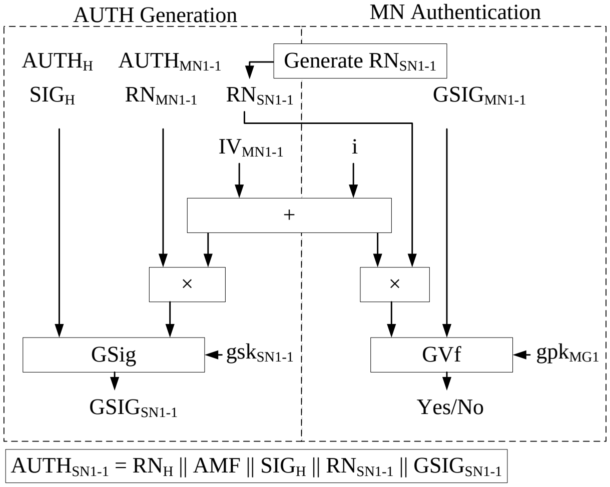


Figure 3.11: AUTH$_{SN_{1\text{-}1}}$ generation and GSIG$_{MN_{1\text{-}1}}$ verification in SNs

MN$_{1\text{-}1}$ by generating AUTH$_{SN_{1\text{-}1}}$ = (SIG$_{\mathrm{H}}\|$RN$_{\mathrm{H}}\|$AMF$\|$GSIG$_{SN_{1\text{-}1}}\|$RN$_{SN_{1\text{-}1}}$), where the first terms are from AUTH$_{\mathrm{H}}$ to authenticate the HN, group signature GSIG$_{SN_{1\text{-}1}} = GSig(gsk_{SN_{1\text{-}1}}, \mathrm{SIG_H} \,\|\, \mathrm{RN}_{MN_{1\text{-}1}} \,\|\, IV_{MN_{1\text{-}1}} + i)$, and RN$_{SN_{1\text{-}1}}$ is a nonce chosen by the SN to challenge MN$_{1\text{-}1}$ later. While waiting for the response from MN$_{1\text{-}1}$, SN$_{1\text{-}1}$ calculates the master key MK by some key generation algorithm, such as Diffie-Hellman [39], for subsequent sessions with MN$_{1\text{-}1}$ in advance. (See Fig. 3.11.)

7. Group Authentication Response (GSIG$_{MN_{1\text{-}1}}$): In response to Group Authentication Request (AUTH$_{SN_{1\text{-}1}}$), MN$_{1\text{-}1}$ authenticates the HN and SN$_{1\text{-}1}$, respectively, by verifying SIG$_{\mathrm{H}}$ and GSIG$_{SN_{1\text{-}1}}$ inside AUTH$_{SN_{1\text{-}1}}$. After successfully authenticating the HN and SN$_{1\text{-}1}$, MN$_{1\text{-}1}$ calculates a master key MK with respect to SN$_{1\text{-}1}$ and generates a message back to SN$_{1\text{-}1}$ containing GSIG$_{MN_{1\text{-}1}} = GSig(gsk_{MN_{1\text{-}1}}, \mathrm{RN}_{SN_{1\text{-}1}} \| IV_{MN_{1\text{-}1}} + i)$. Such operations are diagrammed in Fig. 3.8.

8. Authentication Result (Success/Failure): Upon receipt of a Group Authentication Response message carrying GSIG$_{MN_{1\text{-}1}}$, the SN checks whether MN$_{1\text{-}1}$ has produced the correct response using operations as in Figs. 3.8, 3.9 and 3.11. Then a message with a status code indicating either success or fail-

ure for mutual authentication is sent to $\mathrm{MN}_{1\text{-}1}$, whence our key agreement is completed.

After full authentication, both $\mathrm{MN}_{1\text{-}1}$ and $\mathrm{SN}_{1\text{-}1}$ share a common MK that can be employed as the material for subsequent key derivations.

When a second group member, say $\mathrm{MN}_{1\text{-}2}$, arrives and requests for authentication, $\mathrm{SN}_{1\text{-}1}$ simply initiates mutual authentication and key agreement with $\mathrm{MN}_{1\text{-}2}$ using the existing $\mathrm{AUTH_H}$. Moreover, when $\mathrm{MN}_{1\text{-}2}$ visits any SN of SG_1, say $\mathrm{SN}_{1\text{-}2}$, $\mathrm{MN}_{1\text{-}2}$ and $\mathrm{SN}_{1\text{-}2}$ can accomplish mutual authentication and key agreement locally as well, without involving the HN. In other words, messages 3 and 4 in the prescribed authentication data distribution procedure can be bypassed, leaving out signaling traffic between SN and HN. In this regard, however, $\mathrm{SN}_{1\text{-}2}$ needs to generate a new nonce $\mathrm{RN}_{SN_{1\text{-}2}}$ to create a new challenge message for $\mathrm{MN}_{1\text{-}2}$ (message 6 in Fig. 3.10.) Using distinct arguments from those for $\mathrm{MN}_{1\text{-}1}$, such as $\mathrm{RN}_{MN_{1\text{-}2}}$, $\mathrm{RN}_{SN_{1\text{-}2}}$ and $\mathrm{IV}_{MN_{1\text{-}2}}$, our authentication protocol assures not only the freshness of challenge-response messages but also the uniqueness of MKs for different MNs within SG_1.

As a remark, the proposed GS-AKA protocol reduces messages sent from the HN to a group of SNs repeatedly by providing group authentication data. Observe that $\mathrm{SIG_H}$ allows of periodic updates whenever new nonce is provided.

3.3 Analysis

This section gives security analysis, performance analysis, and storage analysis of the two proposed G-AKA protocol.

3.3.1 Security analysis

Here we reason that the two proposed G-AKA can achieve the same security level as other contemporary AKA protocols do. First GK-AKA and GS-AKA both enforce mutual authentication. In GK-AKA, an MN authenticates both the SN and its HN

Table 3.2: Comparison of four AKA protocols

	UMTS AKA	UMTS X-AKA	GK-AKA	GS-AKA
Auth. Concept	User	User	Group	Group
Auth. method	Pre-Shared Key	Pre-Shared Key	Pre-Shared Group Auth. Key	Group signature
Cryptography	symmetric Key	symmetric Key	symmetric Key	asymmetric key
Storages in SN	n × AV(s)	n × (TK+AUTH)	g × (GTK+Index)	g × (SIG$_H$+IV)

in step 7 and the SN authenticates the MN using the same GTK generated by its HN in step 8. And in GS-AKA, an MN also authenticates both the SN and its HN in step 7 and the SN authenticates the MN using the key gpk generated by the HN in step 8. Second, the random number RN$_H$ in GK-AKA and GS-AKA assures the freshness of authentication messages and master key. In GK-AKA, the RN$_H$, RN$_{MN1-1}$ and RN$_S$, generated by the HN, the first MN and the corresponding SN respectively, assures freshness of GTK and authentication message 6 and 7, as shown in Fig. 3.2, 3.4 and 3.5. In GS-AKA, the RN$_H$ generated by the HN assures the freshness of SIG$_H$, as shown in Fig. 3.8, 3.9 and 3.11. Furthermore, the proposed GK-AKA and GS-AKA both allow for how many authentications have been performed so far. This guarantees the freshness of each local authentication. Finally each MN has a unique initial value IV so that no MN can impersonate any another MN of the same group.

3.3.2 Performance analysis

We now compare the performance of the proposed GK-AKA protocol and GS-AKA protocol with UMTS AKA and UMTS X-AKA. Table 3.2 presents the comparison of four AKA protocols. First, according to using the group concept, the G-AKA have more efficient to accomplish AKA procedure than UMTS AKA and UMTS X-AKA. Hence, the G-AKA can reduce AKA latency and speedup handoff process. Second, only the cryptography algorithm of GS-AKA takes public key architecture (asymmetric key system) to carry out both MN and SN authentication. Therefore, the GS-AKA needs more computation power and time than another AKA schemes to cryptography functions. However, current computation technology of devices can cover computation problems easily.

Next, the GK-AKA, GS-AKA and UMTS X-AKA authorize an SN to authenticate an MN locally after the HN has authenticated the first MN. In GK-AKA, we take a step by using the GAK concept so that the SN can authenticate other MNs in the same group as the MN. Therefore, GK-AKA can reduce authenticate delay and signaling overhead between the HN and the SN. And GS-AKA takes a step further by using the group signature concept so that any SN of an SN group can authenticate each other with any MN of an MN group locally. Therefore, GS-AKA can reduce both authentication delay and signaling overhead between the HN and SNs. Additionally, due to group signature is asymmetric key system, the computation load of GS-AKA higher than the other three AKA protocols. The synchronization of authentication data of UMTS AKA is between HN and MN, UMTS AKA need more signaling messages exchange between HN and MN to accomplish synchronization. Due to adopt the concept of SN group, GS-AKA can perform pre-authentication for other MNs in the same group as the first MN with SNs of a group without any signaling exchange and SN prediction. That is, GS-AKA provides shorter handoff process. Therefore, according to adopt the MN concept, the GK-AKA is well applicable to any tourist group from the same city or country traveling from one place to another, students having a field. GS-AKA adopts the concept of group

Table 3.3: Number of signals between HN and SN in AKA protocols

AKA Methods	Number of signaling messages					
	1 MN			n MNs		
	1 Req.	m Req.	k SNs	1 Req.	m Req.	k SNs
UMTS AKA	1	m	m	n	$m \times n$	$m \times n$
UMTS X-AKA	1	1	k	n	n	$k \times n$
GK-AKA	1	1	k	g	g	k
GS-AKA	1	1	1	g	g	g

signatures for both SNs and MNs. Therefore, GS-AKA is well applicable to system accommodating a public transportation, such as bus, train, high-speed railway, etc.

Table 3.3 shows the number of authentication messages exchanged between the corresponding SNs and the HN in these AKA protocols. When there is only one MN, all four protocols need one pair of request and response messages for a full authentication. For the next m re-authentication of the same MN, UMTS X-AKA, GK-AKA and GS-AKA can just perform local authentication. As for k SNs of an SN group that the MN will visit, GK-AKA still need k pair of request and response messages for a full authentication but GS-AKA can just perform local authentication for subsequent MNs in the same group as the first MN. Additionally, if n MNs of a group visit an SN, GS-AKA and GK-AKA perform better than UMTS X-AKA because the SN can perform local authentications with the MNs of the same group. Furthermore, if n MNs of a group visit k SNs of a group, GS-AKA perform better than GK-AKA because the SNs can perform local authentications with the MNs of the same group. In general, the number of groups, g, is less than the number of MNs. Fig. 3.12 plots the number of messages between the HN and the SN with

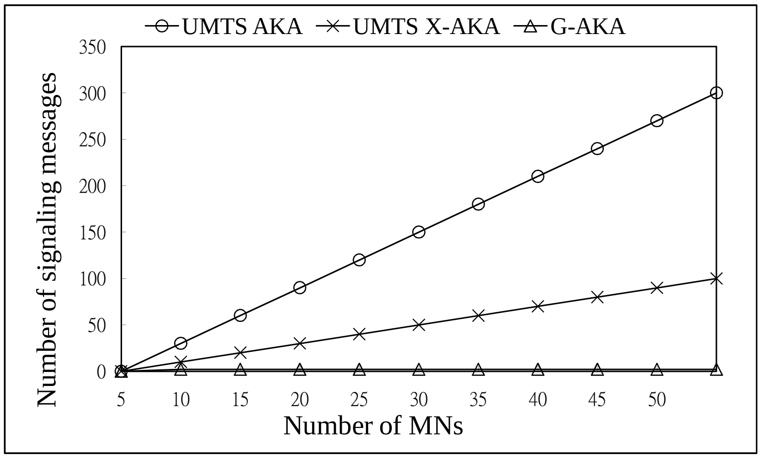


Figure 3.12: Signaling exchange between HN and SN for AKA protocols

respect to the number of MNs of a group that visit the SN.

3.3.3 SN Storage analysis

In the table 3.2, for UMTS AKA, an MN has a set of AVs for mutual authentication kept in each SN so that n MNs occupy an order of $n\times$AVs storage space in each SN. As to UMTS X-AKA, $n\times$(TK+AUTH) space is occupied in each SN because the authentication data (TK+AUTH) are generated for each MN individually. As regards the proposed GK-AKA protocol, personal authentication data are replaced with group authentication data, that is GTK and an index table, hence SN spends only $g\times$(GTK+index table) on storing authentication data for MNs. As for GS-AKA protocol, personal authentication data are replaced with group authentication data, that is SIG$_H$ and initial value (IV). Hence, an SN spends only $g\times$(SIG$_H$+IV) storing group authentication data for MNs. Thus, it can be seen that our GK-AKA and GS-AKA protocol keeps authentication data only for g groups in an SN whereas other approaches require n different authentication data.

3.4 Summary

In this chapter, we present two G-AKA scheme to speedup authentication time and reduce signaling messages between HN and SN. At first, we propose a GK-AKA scheme that adopts the notion of GAK and authorizes an SN to authenticate other MSs of a group locally, after the HN has successfully authenticated the first MN that visits the SN. GK-AKA can reduce signaling overhead significantly between HN and SN when MSs exhibit some group-based moving behaviors. Next, we adopt the concept of group signatures for both SNs and MNs to propose a GS-AKA scheme. Therefore, GS-AKA is well applicable to any system accommodating a public transportation, such as bus, train, high-speed railway, etc. In GS-AKA, all MNs of a group, except the first MN, can perform authentication locally with SNs of an SN group. Furthermore, when an MN moves from one SN to another, the MN and the new SN could perform AKA locally without any SN prediction algorithm. The group manager HN can also sets up group authentication data directly into corresponding SN groups in advance if the HN knows MN group behavior.

Therefore, GK-AKA and GS-AKA can reduce the signaling overhead between SNs and the HN, decrease authentication delay and speed up the handoff process. Meanwhile, the proposed GK-AKA and GS-AKA can still guarantee the same secure level as UMTS-X-AKA does.

Chapter 4

Fast Authentication Schemes

In the previous chapter, the proposed two G-AKA schemes can reduce the authentication time and speedup the handoff process. Furthermore, in the chapter, we will present a fast authentication schemes to shorten the inter-AP handoff process in multi-hop network, such as WLAN Mesh network.

4.1 An Integrated Security Domain Scheme

To reduce the overhead of authentication and encryption processing in WLAN Mesh, we present a security mechanism to integrate the security domains of WLAN Mesh. An MPP and the MAPs connected to this MPP form an integrated security domain (ISD). An MN only performs 802.1X authentication while first time connects to an MAP within the ISD. Authentication latency is removed from the following handoffs in the same ISD. Furthermore, an end-to-end security channel between an MN and an MPP is established without exchanging any extra message. The security channel can improve the performance of WLAN Mesh in routing the encrypted frame.

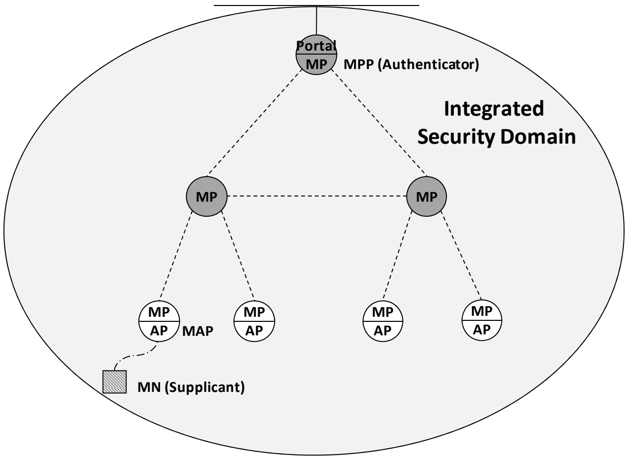


Figure 4.1: WLAN Mesh security architecture with ISD

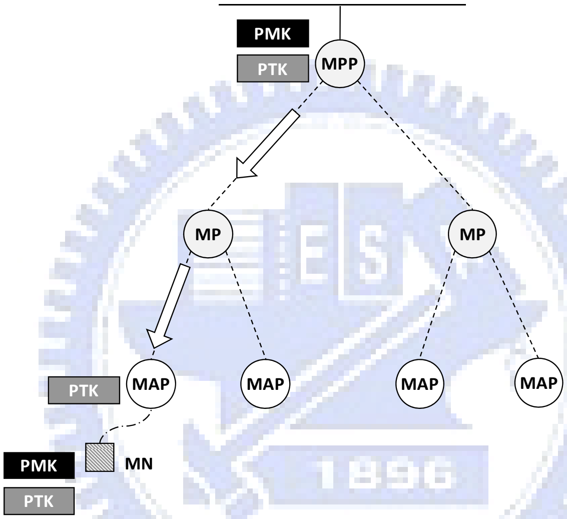


Figure 4.2: PTK distribution

4.1.1 The proposed ISD scheme

The Proposed Security System Architecture

With ISD, security functions of the AP services, such as 802.1X authentication and robust security network association (RSNA) [48] key management, are implemented in the MPP. As shown in Fig. 4.1, the role of 802.1X authenticator is adopted by the MPP instead of the serving MAP.

MAP is the edge of WLAN Mesh and responsible for blocking malicious MNs from accessing the network. In order to provide the ability for MAP to verify frame

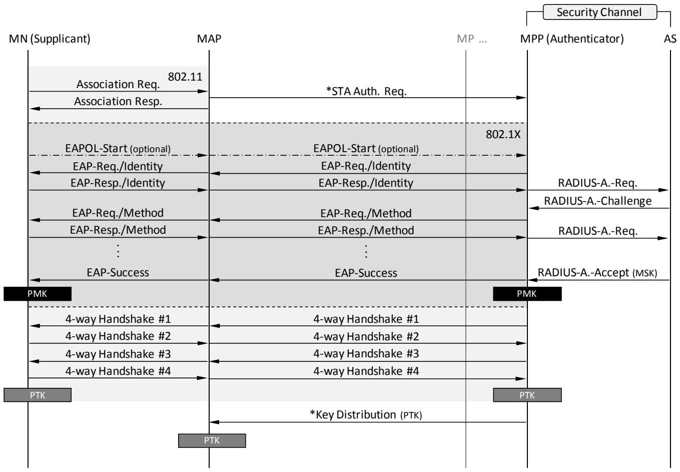


Figure 4.3: RSNA establishment with ISD

integrity, pairwise transient key (PTK) and group traffic key (GTK) are distributed from MPP to the serving MAP via secured mesh links right after 4-way handshake. Fig. 4.2 shows the PTK distribution.

RSNA Establishment

While an MN initially associated to any MAP within the ISD, it is required to perform 802.1X authentication and 4-way handshake to establish the security association with the MPP. For being compatible with conventional MNs, the message flows in the MN portion are identical to ISD and 802.11i in the RSNA establishment.

Since MPP is an authenticator, serving MAP participates in neither 802.1X authentication nor 4-way handshake but forwards all authentication messages between MN and MPP. Fig. 4.3 illustrates the procedures of RSNA establishment for an MN initially authenticating with an MAP within the ISD.

1. The serving MAP checks the Association Request frame to see is any PMKID included. If not, an MN Authentication Request message is sent to the MPP to initialize 802.1X authentication

2. The MN and the MPP perform 802.1X authentication and 4-way handshake, and all messages are forwarded via the serving MAP.

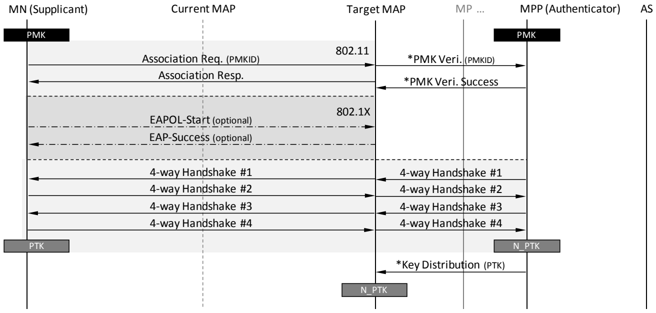


Figure 4.4: Intra-MPP handoff with ISD

3. The MPP distributes the PTK to the serving MAP for integrity verifying.

4. Once the serving MAP obtains the PTK, it will switch the port to the authorized state, and thus the MN is able to access the network.

5. If a GTK is assigned by the MPP in 4-way handshake, it will be distributed to the serving MAP as well.

Handoff Procedures

802.11s allows multiple MPPs reside in one WLAN Mesh, and thus the handoff behaviors with ISD are categorized into intra-MPP handoff and inter-MPP handoff. Moreover, the authentication procedures vary in the two types.

1. Intra-MPP handoff

   Intra-MPP handoff means that an MN drops current connection and re-associates with another MAP connecting to the same MPP.

   Since MPP is the authenticator, MN does not change the authenticator in the intra-MPP handoff. If the PMK is cached by the authenticator, 802.1X authentication will be skipped. Fig. 4.4 illustrates the message flows of intra-MPP handoff.

   (a) The MN re-associates with the target MAP. The PMKID is passed to the MPP for verifying the PMK cached in the MN.

    (b) The PMKID is compared with the PMK cached in the MPP. If the PMKID is valid, the MPP will inform the target MAP with a PMK Verification Success message.

    (c) Some implementations of the supplicant use the EAPOL-Start message to initialize 802.1X authentication. If the target MAP receives an EAPOL-Start message, it will reply an EAP-Success message to skip the EAP authentication.

    (d) Following 4-way handshake and PTK distribution are identical to the RSNA establishment mentioned before.

2. Inter-MPP handoff

   Inter-MPP handoff is performed while an MN moves from one MAP to another MAP connecting to the different MPP. The MN will switch to another ISD in the inter-MPP handoff. If the ISD has not been visited by the MN or the cached PMK is expired, pre-authentication will be performed. However, the MN may fail to pre-authenticate with the new MPP, and thus the overhead of 802.1X authentication is introduced.

   There are many factors cause pre-authentication to be failed, such as the moving speed of the MN, the size of the overlapping coverage area, the target AP prediction, the latency of EAP authentication, etc.

   Fig. 4.5 illustrates the message flows of inter-MPP handoff and RSNA establishment. Detail procedures are as follows:

    (a) The MN re-associates with the target MAP. The PMKID is forwarded to the MPP for verifying the PMK cached in the MN.

    (b) Since the new MPP does not cache the PMK, the PMKID verification is failed, and a message will be sent to the target MAP for informing that following authentication messages should be forwarded to the MPP.

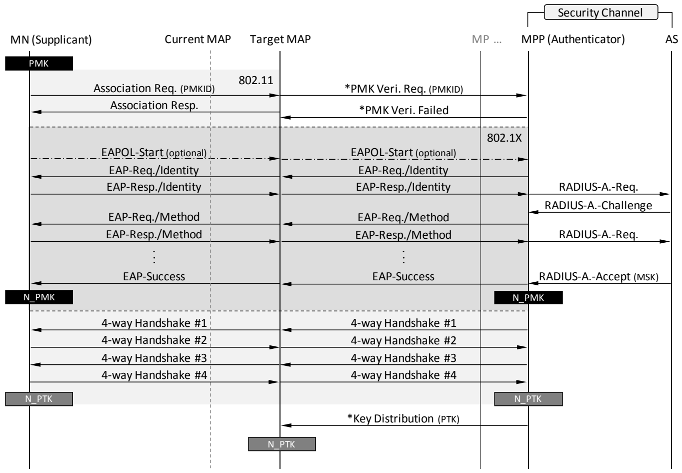


Figure 4.5: Inter-MPP handoff with ISD

(c) 802.1X authentication and 4-way handshake are performed, followed by the PTK distribution. The procedures are the same as the RSNA establishment described in section 4.1.1.

Encapsulation

To mitigate the routing overhead incurred by the hop-by-hop encryption in the multi-hop network, the proposed mechanism establishes an end-to-end security channel between MN and MPP. Therefore, if the correspondent host is outside the WLAN Mesh, encryption and decryption operations will be only performed by serving MAP and MPP.

We construct a bidirectional MAC tunnel between serving MAP and MPP to avoid the MAC header used as the input of the frame encryption processing being modified. Fig. 4.6 gives an instance to explain the encapsulation processing of ISD.

The MN transmits a WLAN frame to the destination which is outside the WLAN Mesh, e.g., the default gateway (GW). Detail procedures are as follows:

1. The MN constructs a WLAN frame (H1 + P, where H1 is the header of the WLAN frame, and P is the payload) and encrypts the frame with the PTK.

2. The WLAN frame is transmitted to the serving MAP via an 802.11 link.

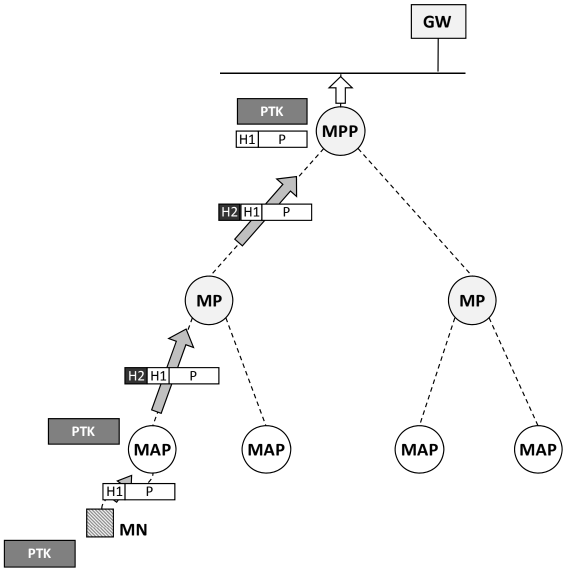


Figure 4.6: Encapsulation processing (external destination)

3. The MAP verifies the MIC code of the frame with the PTK. If the MIC code is invalid, this frame will be discarded, otherwise the destination will be examined.

4. If the destination is outside the WLAN Mesh, the MAP will encapsulate the WLAN frame into a WLAN Mesh frame (H2 + H1 + P, where H2 is the header of the WLAN Mesh frame) and forward the frame to the next hop. Thus, the inner header (H1) will not be altered in the routing.

5. The MP forwards the frame to the next hop. No further operations are needed.

6. The MPP removes the WLAN Mesh header (H2) and decrypts the WLAN frame (H1 + P) with the PTK.

7. Finally, the MPP encapsulates the payload (P) into an Ethernet frame and forwards the frame to the destination.

Fig. 4.7 illustrates the encapsulation processing for the source which is outside the WLAN Mesh. For example, the GW transmits an Ethernet frame to the MN.

1. The MPP receives an Ethernet frame and translates into the WLAN Mesh format (H2 + P). The frame is encrypted by the PTK and encapsulated into

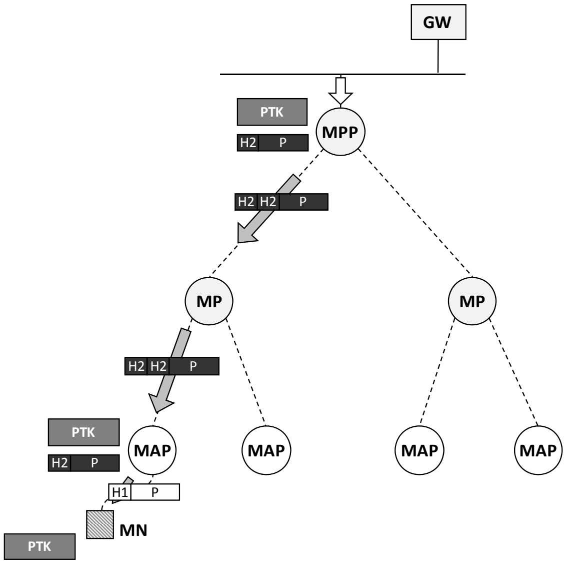


Figure 4.7: Encapsulation processing (external source)

another WLAN Mesh Frame (H2 + H2 + P). Two identical WLAN Mesh headers can keep the inner header intact in the routing. After encryption and encapsulation processing finished, the MPP forwards the frame to the next hop.

2. The MP forwards the frame to the next hop.

3. The MAP removes the outer WLAN Mesh header (H2) and decrypts the inner WLAN Mesh frame (H2 + P) with the PTK.

4. The MAP encapsulates the payload (P) into a WLAN frame (H1 + P) and encrypts the frame with the PTK. Finally, the MAP forwards the WLAN frame to the MN.

To improve the routing performance, if destination and source are both reside the WLAN Mesh, 802.11s will apply the shortcut routing path instead of the regular routing path while. For example, as shown in Fig. 4.8, D→B→A→C→G is replaced by D→B→C→G.

To support the shortcut routing path, ISD applies the original hop-by-hop encryption of 802.11s. Fig. 4.8 shows the encapsulation processing for the MN1 transmitting a WLAN frame to the MN2. Detail procedures are as follows:

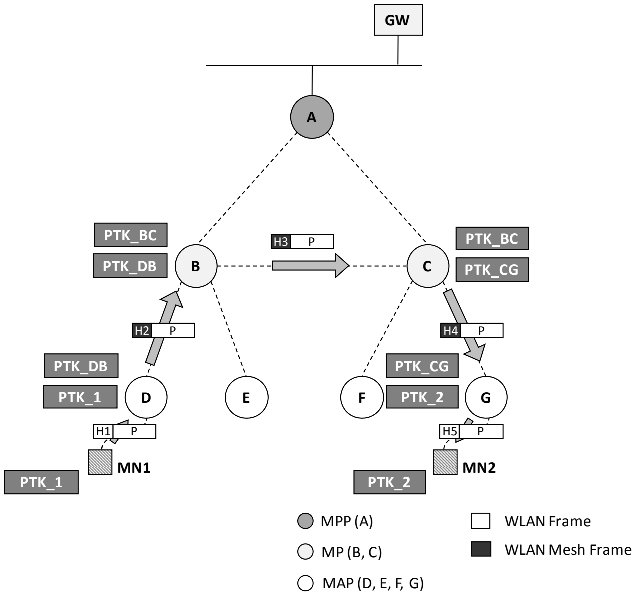


Figure 4.8: Encapsulation processing (internal)

1. The MN1 constructs a WLAN frame (H1 + P). The frame is encrypted with the PTK_1 and transmitted to the MAP D.

2. The MAP D decrypts the WLAN frame with the PTK_1 and encapsulates the payload (P) into a WLAN Mesh frame (H2 + P). The frame is encrypted with the PTK_DB and forwarded to the MP B.

3. The MP B and the MP C decrypt the frame and then re-encrypt it with the PTK of the next-hop. After that, the frame is forwarded to the next-hop.

4. The MAP G decrypts the WLAN Mesh frame with the PTK_CG and encapsulates the payload (P) into a WLAN frame (H5 + P). The frame is encrypted with the PTK_2 and forwarded to the MN2.

5. MN2 decrypts the WLAN frame with the PTK_2.

4.1.2 Security Considerations

To claim that ISD is a secure mechanism, it is necessary to state the security goal as well as the security assumptions. The security goal of ISD is to secure the

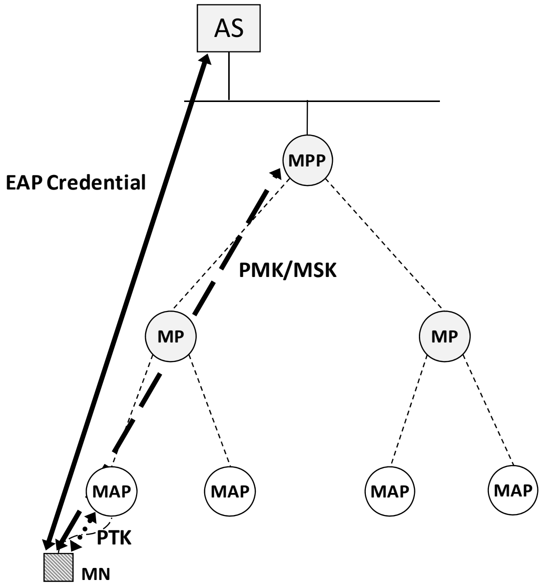


Figure 4.9: Trust Relationships in the ISD

wireless communication between MN and MAP, and the strength of ISD should be equivalent to 802.11i. ISD assumes that MNs and MAPs are 802.11i-based devices and the 802.11i security assumptions should be satisfied. Besides, mesh links among MPs are required to be protected by EMSA services. To present ISD is as secure as 802.11i, we first analyze the trust relationship of ISD, and then threat models are examined.

Trust Relationship

For ISD, as shown in Fig. 4.9, 802.1X authentication and 4-way handshake are performed by the MN and the MPP, and the MN↔AS↔MPP trust chain is established. Since there is the MAP↔MPP trust relationship, the MN↔MPP↔MAP trust chain can be inferred from the former two trust relationships. Therefore, we can claim that the trust relationship provided by ISD is equivalent to 802.11i.

In terms of the handoff, there are three related trust relationships: MN↔AS, MN↔MAP and MN↔MPP. The MN↔MAP trust relationship is destroyed in the intra-MPP handoff and needs to be reestablished. For ISD, since the MN↔MPP and the MAP↔MPP trust relationship are remained, the implicit trust exists between the MN and the new MAP. However, to secure the connection between the MN and the new MAP, a new PTK is necessary to prevent the unauthorized disclosure to the old MAP. Therefore, in the intra-MPP handoff, MN and MPP need to perform 4-way handshake to derive a fresh PTK. Since the old MAP has neither the new PTK nor the PMK, it can not obtain the content encrypted by the new PTK.

For 802.11i, to reestablish the MN↔MAP trust relationship, the MN needs to perform 802.1X authentication with the new MAP. Consequentially, it will introduce significant latency.

Threat Model

The proposed mechanism should avoid introducing any security degradation to the 802.11i RSN. In addition to the threats against 802.11i and 802.11s, there are other threats need to be recognized for ISD.

- PMKID Leakage

  Even though an attacker may obtain the corresponding PMKID from previous eavesdropping and is able to skip 802.1X authentication, it does not result in any security flaw. Due to MSK and PSK are never transmitted via the wireless media, a valid PTK can not be derived by the attacker. Therefore, the attacker can not compute the valid MIC code of message 2 in the 4-way handshake, and the attacker is blocked by the MAP.

- Authenticator Compromise

  In the situation that an authenticator is compromised or stolen, an attacker may obtain all PMKs cached in this authenticator. With ISD, the attacker can access the WLAN Mesh via any MAP connected to this authenticator. However, 802.11r also incurs this vulnerability. The compromised authenticator in 802.11r will expose PMK-R0s to the attacker. Since IEEE 802.11 working group allows this situation to occur, we believe this vulnerability is acceptable.

- Unauthorized Disclosure

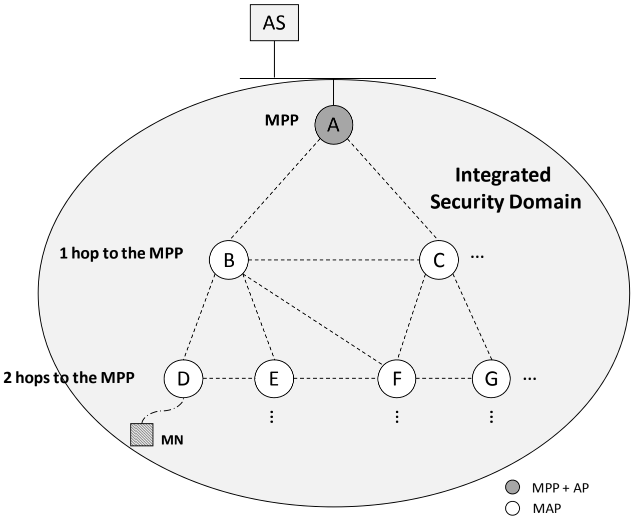


Figure 4.10: WLAN Mesh handoff analytical architecture

Compromised mesh links will result in the unauthorized disclosure of keys. For 802.11i, an MSK is transmitted from the AS to the serving MAP via mesh links. If the security of mesh links is compromised, it is possible that the MSK will be exposed to an attacker. For ISD, only the PTK is transmitted via mesh links. Since the hierarchy of PTK is lower than MSK, the compromised PTK will not introduce further security degradation compared with the compromised MSK.

4.1.3 Handoff overhead analysis

In this section, we analysis the link layer security mechanisms and present the related handoff overhead. For MN, the major concern is whether the handoff latency will damage the quality of real-time applications or not. For WLAN Mesh, the handoff traffic is the main issue. The handoff analytical architecture, as shown in Fig. 4.10, consists of a MPP and MAPs with capability as APs for MN.

For the handoff analytical architecture, the two-dimensional random walk model [3] is applied to capture the movement of MNs in the WLAN Mesh and calculate the number of handoffs. Fig. 4.11 illustrates a 6-subarea cluster, where cells are marked as (x, y). The x represents the layer of the cluster in which the cell resides, and y

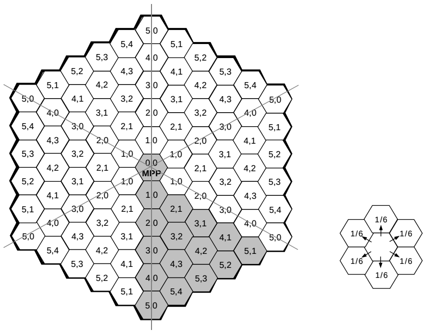


Figure 4.11: 6-subarea cluster analytical model

denotes the type y. Cells with the same set of neighbors' type are classified into one type. MNs in cells with the same type will have the same candidate handoff targets and will leave the cells with the same pattern. Therefore, the gray area shown in Fig. 4.11 can capture the movement of MNs within the cluster.

An analytical model, as shown is Fig. 4.11, is proposed to compute the handoff overhead for an MN roaming within the WLAN Mesh. The estimated handoff overhead of ISD and 802.11i will be compared in the following section. With the handoff pattern, the proposed equations can estimate the average handoff latency.

Whereas 802.1X authentication and 4-way handshake contribute the major part of the handoff latency, the quality of real-time applications of affected by the security mechanism. The latency introduced by the security mechanism can be classified into two types: intra-MPP handoff latency (L_{INTRA}) and latency inter-MPP handoff latency (L_{INTER}).

Intra-MPP handoff latency

L_{INTRA} represents the latency for an MN performing the intra-MPP handoff, which consists of authentication latency (L_{INTRA_AUTH}) and 4-way handshake latency (L_{INTRA_4W}).

There are two two authentication messages are exchanged to verify the PMK cached by the MN. L_{INTRA_AUTH} represents the average latency, where

$$L_{INTRA\_AUTH} = 2 \times T \times H \tag{4.1}$$

$$H = \frac{\sum_{x=0}^{n-1} x \times S}{1 + n(n-1)/2} \tag{4.2}$$

- T is the single-hop transmission time.
- H is the average hop count between MAP and MPP and calculated based on the proposed handoff model.
- x is the hop count between MAP and MPP, i.e. the type of MAP.
- S is the number of MAPs in the gray area with x hops to MPP.
- n is the cluster size. $1 + n(n-1)/2$ is the total number of MAPs.

In the handshake phase, 4-way handshake messages are transmitted between the MN and the MPP. In addition, the PTK is distributed to the target MAP. L_{INTRA_4W} represents the average latency, where

$$L_{INTRA\_4W} = L_{4W} + 5 \times T \times H \tag{4.3}$$

- L_{4W} is the latency for an MN performing 4-way handshake in the single-hop network, i.e., WLAN.

For 802.11i, if the PMK is not cached by the target MAP, 802.1X authentication will be performed in the handoff. L_{INTRA_AUTH} represents the average latency, where

$$L_{INTRA\_AUTH} = L_{1X} + M_{RADIUS} \times T \times H \tag{4.4}$$

- L_{1X} is the latency for an MN performing 802.1X authentication in the single-hop network, i.e., WLAN.

- M_{RADIUS} is the number of RADIUS messages exchanged between the target MAP and the AS in an 802.1X authentication.

In the handshake phase, the latency is the same in WLAN Mesh and WLAN. L_{INTRA_4W} represents the latency, where

$$L_{INTRA\_4W} = L_{4W} \tag{4.5}$$

- L_{4W} is the latency for an MN performing 4-way handshake in WLAN.

Based on equations 4.1, 4.3, 4.4, and 4.5(13), L_{INTRA} is defined as

$$L_{INTRA} = (1 - P_{PMK\_MISS}) \times L_{INTRA\_4W} + P_{PMK\_MISS} \times (L_{INTRA\_AUTH} + L_{INTRA\_4W}) \tag{4.6}$$

$$P_{PMK\_MISS} = (1 - P_{REVISIT}) \times P_{PF} \tag{4.7}$$

- P_{PMK_MISS} is the probability that the PMK is not cached by the target MAP.
- $P_{REVISIT}$ is the probability that an MN moves to a visited cell or cluster.
- P_{PF} is the probability that 802.11i pre-authentication is failed.

Since the PMK is always cached by the authenticator, the intra-MPP handoff with ISD will only introduce L_{INTRA_4W}. However, if an MN handoffs to a new MAP and fails to pre-authenticate with it, L_{INTRA_AUTH} will be introduced to the intra-MPP handoff with 802.11i.

Inter-MPP handoff latency

L_{INTER} represents the latency for an MN performing the inter-MPP handoff, which consists of authentication latency (L_{INTER_AUTH}) and 4-way handshake latency (L_{INTER_4W}).

While the MN moves out of the cluster, if the PMK is not cached by the new MPP, 802.1X authentication will be performed. L_{INTER_AUTH} represents the latency, where

$$L_{INTER\_AUTH} = L_{1X} + M_{1X} \times (n-1) \times T \tag{4.8}$$

- M_{1X} is the number of EAPOL messages exchanged between the target MAP and the MPP in an 802.1X authentication.
- $n-1$ is the hop count between the target MAP and the new MPP.

An MN performing the inter-MPP handoff will re-associate with another boundary MAP in another cluster. Thus, the hop count between the target MAP and the new MPP is definitely $n-1$.

L_{INTER_4W} represents the average latency for 4-way handshake and PTK distribution in the inter-MPP handoff, where

$$L_{INTER\_4W} = L_{4W} + 5 \times (n-1) \times T \tag{4.9}$$

L_{INTER_AUTH} represents the authentication latency, where

$$L_{INTER\_AUTH} = L_{1X} + M_{RADIUS} \times (n-1) \times T \tag{4.10}$$

L_{INTER_4W} represents the 4-way handshake latency, where

$$L_{INTER\_4W} = L_{4W} \tag{4.11}$$

Based on equations 4.7, 4.8, 4.9 and 4.10, L_{INTER} is defined as

$$L_{INTER} = (1-P_{PMK\_MISS}) \times L_{INTER\_4W} + P_{PMK\_MISS} \times (L_{INTER\_AUTH} + L_{INTER\_4W}) \tag{4.12}$$

For ISD, only an MN moves to an unvisited ISD and fails to perform pre-authentication, the authentication latency is introduced to the inter-MPP handoff.

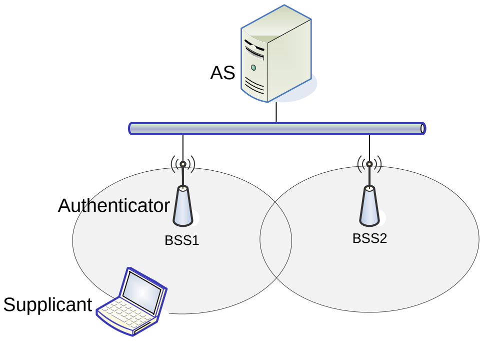


Figure 4.12: Experimental Environment

However, for 802.11i, the MN will perform 802.1X authentication in each handoff in the same condition. Therefore, ISD can greatly reduce the demand for performing 802.1X authentication and provide the equivalent security strength as 802.11i.

Based on equations 4.1, 4.6, and 4.12, for an MN roaming within the WLAN Mesh, the expected handoff latency contributed by the security mechanism is defined as

$$L_S = \frac{\sum_{x=0}^{n-1}\sum_{y=0}^{x-1}\sum_{k=1}^{\infty}\sum_{j=0}^{n-2} P_{k,(x,y),(n,j)} \times [\frac{(k-1)}{k} \times L_{INTRA} + \frac{1}{k} \times L_{INTER}]}{1+n(n-1)/2} \tag{4.13}$$

In order to obtain parameters of the equations, an experimental platform is built to measure the handoff latency, transmission time, the number of messages, etc. The experimental environment is shown in Fig. 4.12, where the AS, two authenticators and the supplicant reside in a LAN.

The supplicant is a laptop installed Windows XP SP2, and the supplicant software is the build-in Windows Zero Configuration Service. Two authenticators are laptops controlled by the hostapd-0.5.7. The FreeRADIUS-1.1.4 is installed in the AS to provide the authentication services. The encryption mechanism is WPA2/AES , and the EAP method is PEAP/EAP-MSCHAPv2.

Parameters are measured in the experimental platform. Table 4.1 presents the average measurement with 20 experiments.

Fig. 4.13 presents the relationship between P_{PF} and L_S at n=3. Estimated

Table 4.1: Parameters measured in the experimental platform

T	2.44 ms
L_{1X}	401.63 ms
L_{4W}	20.76 ms
M_{1X}	22 messages
M_{RADIUS}	18 messages
R	1.049180328

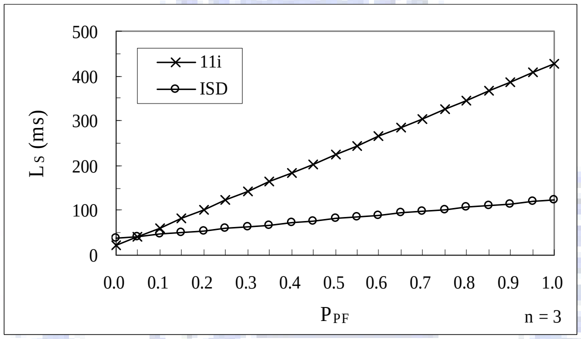


Figure 4.13: Handoff latency with different P_{PF}

results show that ISD remarkably reduces the handoff latency. At P_{PF}=1.0, i.e., MN does not perform preauthentication, ISD can improve the handoff latency up to 245%. Therefore, even though most of current 802.11i devices do not support preauthentication , MNs can still take advantage of ISD. However, at P_{PF} <0.05, due to 4-way handshake messages are forwarded between MAP and MPP, ISD introduces larger L_S, than 802.11i.

Fig. 4.14 presents L_S with different cluster sizes at P_{PF}=1.0. At n=5, ISD approaches the minimal L_S. Actually, the handoff latency of ISD is almost stable at n >3.

For ISD, the burden incurred by the multi-hop transmission in the 4-way hand-

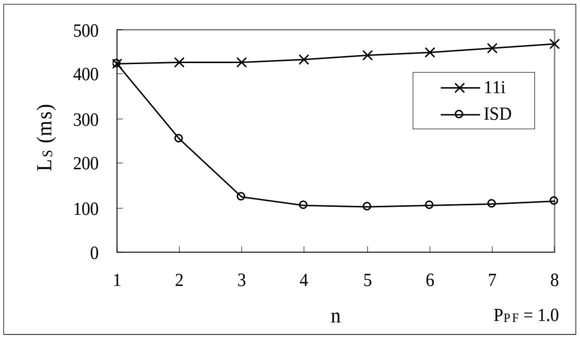


Figure 4.14: Handoff latency with different n

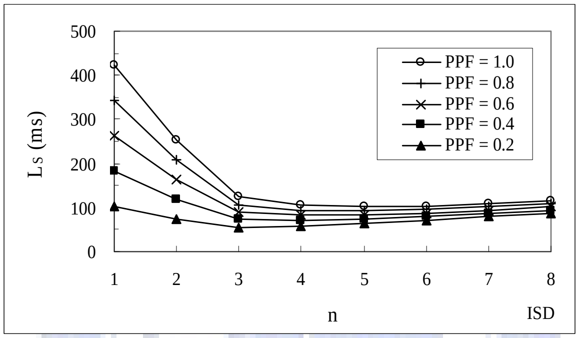


Figure 4.15: Handoff latency of ISD with different n and P_{PF}

shake counteracts the benefit of the larger cluster size. For 802.11i, EAP authentication is also delayed by the multi-hop transmission, and thus L_S increases with the growing cluster size.

L_S of ISD with different cluster sizes and P_{PF} are shown in Fig. 4.15. Results indicate that the larger cluster size avail the handoff latency in all kinds of P_{PF}. Besides, the influence of P_{PF} is decreasing with the growing cluster size.

According to the estimated results, we can conclude that ISD provides great improvement in the handoff latency when the cluster size is around 3 layers, i.e., 37 MPs connect to one MPP. This number accords with the scale of the 802.11s standard.

ISD is practical to use in current wireless environments. In terms of the power consumption, ISD estimates 802.1X authentication, and thus the battery-powered MN can balance the power consumption and the handoff performance. Furthermore, since the AS mostly resides in the core network, L1X would be longer than 400 ms. Estimated results indicate that ISD can further improve the handoff latency in this environment.

Even though ISD is the centralized architecture and forwards 4-way handshake messages to MPP, it does not result in the extra overhead in the handoff traffic. Actually, ISD can reduce the handoff traffic in all kinds of mesh networks at $P_{PF} = 1.0$.

4.1.4 Summary

The authentication latency is a key factor for supporting the seamless handoff. To improve the handoff latency, ISD is proposed to remove 802.1X authentication from the handoff.

Another problem is the routing performance of WLAN Mesh. The hop-by-hop encryption delays the routing processing of MPs. An end-to-end security channel is provided by ISD to solve this problem.

Another advantage of ISD is the compatibility to current 802.11i/11s devices. MNs can apply the proposed mechanism without any modification. Besides, ISD is an optional feature to WLAN Mesh. Original security and routing mechanism of the 802.11s standard can cooperate with ISD.

To evaluate the handoff latency introduced by the link layer security mechanism, we propose a handoff model to estimate the handoff latency for an MN roaming within the WLAN Mesh. Results indicate that ISD improves the handoff latency up to 245% and provides 80%-90% successful preauthentication probability without any assistance.

Chapter 5

Mobile Resource Reservation Schemes

In order to resolve excessive resource reservations problem, we first present an Intelligent Agent-based ReSource reserVation aPproach (IARSVP) for MN to avoid redundant resource reservations made in common routes, support route optimization and registration naturally, and discover alternative routes dynamically. Next, we further present an Mobile Bandwidth-Aggregation (MBA) reservation scheme to overcome mobility unawareness and excessive signals problems for all nodes inside a NEMO. The MBA can makes an Mobile Router (MR) the proxy of all nodes insides a mobile network and has the MR aggregates and reserve the bandwidth required for all nodes inside a NEMO [20].

5.1 An Agent-based Resource Reservation Scheme

In this section, we present an Intelligent Agent-based ReSource reserVation aPproach (IARSVP) to support QoS aware packet transmissions for MIP [37, 29] networks. Mobile Intelligent Agents (MIAs) are characterized by their ability to move across wide-area networks, operate autonomously on foreign hosts, and perform tasks on behalf of the originating hosts. With MIAs, IARSVP can allocate resources in advance for neighbor locations an MN may visit next. Because MIAs carries an MN's mobility security association, QoS requirement and administration specification, and associated executable codes, they can perform location updates on behalf of MNs, and adjust autonomously in accordance with the network topology and resource usage when locating forwarding points for the MN. Therefore, IARSVP can avoid redundant resource reservations made in common routes, support route optimization and regional registration naturally, and discover alternative routes dynamically. Simulation results show that IARSVP outperforms Mobile RSVP [42] and Hierarchical Mobile RSVP [45] in terms of forced termination, reservation blocking, and session completion probabilities.

5.1.1 Mobile Intelligent Agent technology

In this approach, we adopt mobile intelligent agent technology [18, 36, 38] to support QoS-guaranteed resource reservations in MIP networks. MIAs are characterized by their ability to roam across wide-area networks, operate autonomously on foreign hosts (within the running environments for the MIAs), and perform tasks on behalf of the originating hosts. Several efforts at standardization are presently working in the real world, namely The Foundation for Intelligent Physical Agents (FIPA) and Object Management Group (OMG). Many researchers have proposed MIA technologies in various applications or research areas, such as IP telephony [21], routing [10, 15, 33, 30, 32], network management [35, 23], monitoring systems [19], information retrieval [11] and Internet-wide collaborative systems [44]. Although a number of general arguments exist as to why MIA is potentially useful [18], the overhead of MIA is just slightly more than that in a non-MIA network architecture [15, 33].

By adopting MIAs techniques, IARSVP can dynamically determine the forwarding points (FPs), subject to the MN's QoS requirement, on the route to an MN's current location for the routing paths to the locations where the MN may visit next.

In IARSVP, each MN can generate MIAs that carry the MN's mobility association, QoS requirements and QoS strategies, as well as the alteration part of the execution code in the Call Admission Control (CAC) module. The alteration part is adaptable in accordance with QoS strategies such as the guaranteed service, the control-load service and the best-effort service specified by IPv6. Therefore, MIAs can operate autonomously in locating FPs, making resource reservations and sending binding update messages to CNs on behalf of MNs. Due to the autonomy capabilities of MIAs, IARSVP can offload mobility management and resource reservation processes from an MN and the MN's HA to MIAs situated on the routes from a CN to the locations where the MN is currently visiting or may visit next. Consequently, with MIAs, IARSVP can use resources more effectively and can thus fulfill more reservation requests as explained later.

5.1.2 The IARSVP approach

IARSVP adopts MIAs to support QoS guaranteed services for MIP networks. As mentioned previously, MIAs is capable of moving across the Internet and operating autonomously on behalf of MNs. Therefore, with MIAs, IARSVP can avoid redundant reservation, support route optimization and local mobility, and discover alternative routes dynamically.

FPs in IARSVP is very similar to the receiver-anchor nodes in MRSVP except that their locations are dynamically determined by MIAs in IARSVP, rather than stationary as in MRSVP. In order to locate FPs, IARSVP defines two types of MIAs, namely Forwarding Point MIAs (FP-MIAs) and Resource Reservation MIAs (RR-MIAs). Both FP-MIAs and RR-MIAs can move on the Internet and perform autonomous operations on behalf of MNs. FP-MIAs are responsible for locating FPs and performing MIP functions while RR-MIAs for making resource reservations. In IARSVP, an FP of an MN is a crossover router that provides sufficient QoS services and is nearest to the MN in the common portion of both the current and the next anticipated routes to the MN. An MN issue an FP-MIA to locate FPs for the locations the MN may visit next. The FP-MIA will travel reversely along the active reservation path toward the CN. When the FP-MIA finds a candidate router where exists some routes to some anticipated locations, it will create one RR-MIA, for making a passive reservation, for each route to an anticipated location. Each RR-MIA attempts to make resource reservation along the route hop-by-hop toward an anticipated location until either the RR-MIA reaches the anticipated location or fails in making resource reservation on some router. In either case, each RR-MIA will report to the FP-MIA the status of the attempt. If any RR-MIA succeeds in resource reservation, the FP-MIA will stay in the candidate router, which becomes a passive FP, and fork a new FP-MIA to locate the FPs for other anticipated locations remaining if any. Otherwise, if all RR-MIAs fail in making resource reservation, the FP-MIA will move to the next upstream router to locate FPs for the anticipated locations.

When an MN enters a new location, it will report its appearance to the passive FP-MIA of the new location. The passive FP-MIA will then become the active FP-MIA and start intercepting and forwarding packets destined for the MN. Besides, the active FP-MIA also sends binding updates to the MN's HA and the CNs with an active reservation on the FP.

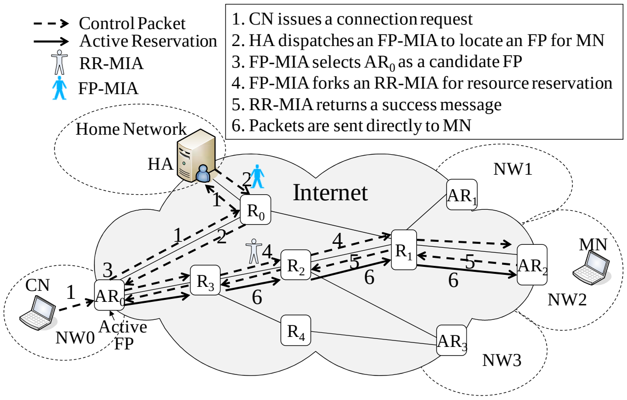
Because MIAs carry an MN's mobility security association, QoS requirement and administration specification and associated executable codes, MIAs can operate autonomously on behalf of MNs. With the autonomous operation capability, MIAs can thus adjust their own behaviors in accordance of the network topology and bandwidth utilization, and perform necessary mobility functions on behalf of MNs.

In the following subsections, we will explain how IARSVP works in detail.

Session Initialization Phase

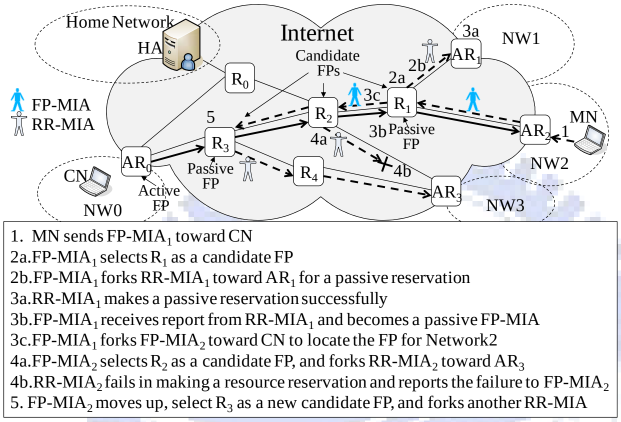
Fig. 5.1 shows the session initialization in IARSVP scheme. Suppose a CN sends a connection request to an MN away from home. The MN's HA will intercept the request and dispatch an FP-MIA to locate an FP for the new session of the MN. The FP-MIA will move reversely along the path the request travelled until the FP-MIA reaches the access router of the CN, AR_0 in Fig. 5.1.

The FP-MIA will then dispatch an RR-MIA to reserve resources along the path



1. CN issues a connection request
2. HA dispatches an FP-MIA to locate an FP for MN
3. FP-MIA selects AR$_0$ as a candidate FP
4. FP-MIA forks an RR-MIA for resource reservation
5. RR-MIA returns a success message
6. Packets are sent directly to MN

Figure 5.1: Session initialization phase



1. MN sends FP-MIA$_1$ toward CN
2a.FP-MIA$_1$ selects R$_1$ as a candidate FP
2b.FP-MIA$_1$ forks RR-MIA$_1$ toward AR$_1$ for a passive reservation
3a.RR-MIA$_1$ makes a passive reservation successfully
3b.FP-MIA$_1$ receives report from RR-MIA$_1$ and becomes a passive FP-MIA
3c.FP-MIA$_1$ forks FP-MIA$_2$ toward CN to locate the FP for Network2
4a.FP-MIA$_2$ selects R$_2$ as a candidate FP, and forks RR-MIA$_2$ toward AR$_3$
4b.RR-MIA$_2$ fails in making a resource reservation and reports the failure to FP-MIA$_2$
5. FP-MIA$_2$ moves up, select R$_3$ as a new candidate FP, and forks another RR-MIA

Figure 5.2: FP-locating procedure

to the MN's current location. The RR-MIA is responsible for performing the resource reservation procedure. If the RR-MIA can reserve resources successfully for the new session, the FP-MIA becomes the active FP-MIA and can start intercept packets destined for the MN. Besides, the MN also needs to perform an FP-Locating procedure for seamless handoff.

FP-Locating Procedure

As shown in Fig. 5.2, the MN initiates an FP-locating procedure by dispatching an FP-MIA to locate FPs for neighbor locations the MN may visit next. The steps of the FP-locating procedure are as follows.

Step 1. MN dispatches FP-MIA$_1$ from the MN's current location, in NW2, toward

CN.

Step 2a. FP-MIA$_1$ selects an intermediate router R$_1$ as a candidate FP for the MN's location in NW1.

Step 2b. FP-MIA$_1$ dispatches RR-MIA$_1$ to make a passive resource reservation for the MN's location in NW1.

Step 3a. RR-MIA$_1$ makes a passive resource reservation successfully from R$_1$ to AR$_1$.

Step 3b. FP-MIA$_1$ receives the success report from RR-MIA$_1$ and decides to stay on R$_1$ as a passive FP-MIA. R$_1$ then becomes a passive FP.

Step 3c. FP-MIA$_1$ forks FP-MIA$_2$ to locate the FP for the MN's location in NW3.

Step 4a. FP-MIA$_2$ moves up, finds a candidate router R$_2$, and forks RR-MIA$_2$ to make a passive resource reservation for the MN's location in NW3.

Step 4b. RR-MIA$_2$ fails in making a resource reservation on the route to AR$_3$ and reports the failure to FP-MIA$_2$.

Step 5. FP-MIA$_2$ moves up further, selects R$_3$ as a candidate FP and dispatches an RR-MIA to make a passive resource reservation for the MN's location in NW3. The reservation succeeds this time and FP-MIA$_2$ becomes a passive FP-MIA staying in R$_3$ for the MN's location in NW3.

Resource Reservation Phase

Similar to the reservations in RSVP, the reservations in IARSVP are receiver-initiated reservations. In RSVP, the sender sends a Path message that is routed, by the legacy routing protocol, hop-by-hop towards the receiver. The Path message carries the necessary information for setting up the path state at the routers along the path from the sender all the way to the receiver. Upon receiving the Path message, the receiver replies with a Resv message to request resource reservations reversely along the path the Path message traversed previously. Both MRSVP and

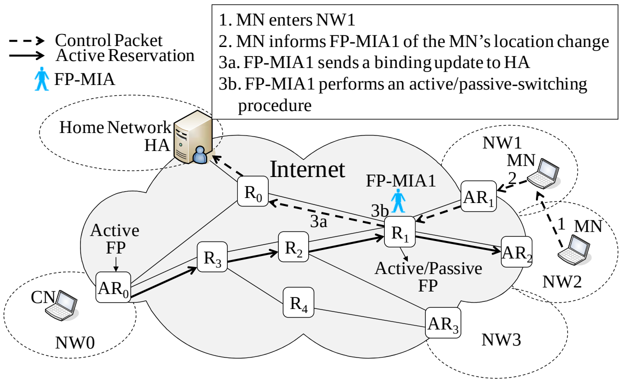


Figure 5.3: Handover procedure

HMRSVP adopt the same reservation principle by having a proxy that is situated in a receiver MN's neighbor location and issues towards the CN a Resv message on behavior of the MN. However, in IARSVP, an RR-MIA of a receiver MN can makes resource reservations when it traverses toward a neighbor location of the MN because the RR-MIA already has the traffic characteristic of the sender and the QoS specification of the MN. The RR-MIA will report to the candidate FP whether it succeeds in making resource reservation. Furthermore, if the RR-MIA can reserve the resource successfully, it will stay in the access router and refresh the reservation periodically.

Handoff Procedure

Fig. 5.3 illustrates the handoff procedure of IARSVP. Assume that a passive FP-MIA, FP-MIA$_1$, situates on R$_1$ for the MN's neighbor location in NW1. When the MN enters the coverage area of AR$_1$, it will notify FP-MIA$_1$ of its location change. After receiving the notification, FP-MIA$_1$ will become an active FP-MIA for the MN's location served by AR$_1$ and start intercepting packets destined for the MN. In addition, FP-MIA$_1$ also sends a binding update message to the MN's HA on behalf of the MN. However FP-MIA$_1$ need not inform the CN of the MN's movement since R$_1$ is now an anchor point for the packets sent by the CN to the MN. Furthermore, FP-MIA$_1$ is also a passive FP-MIA for the MN's old location served by AR$_2$.

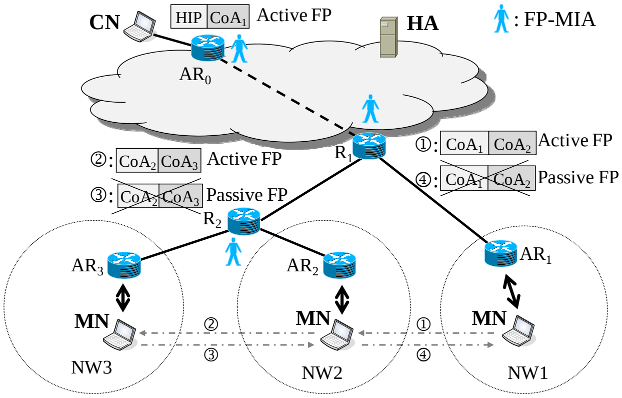

CN
HIP | CoA$_1$ Active FP
HA
: FP-MIA
AR$_0$
R$_1$
①: CoA$_1$ | CoA$_2$ Active FP
④: ~~CoA$_1$ | CoA$_2$~~ Passive FP
②: CoA$_2$ | CoA$_3$ Active FP
③: ~~CoA$_2$ | CoA$_3$~~ Passive FP
R$_2$
AR$_3$
AR$_2$
AR$_1$
MN
MN
MN
② ③
① ④
NW3
NW2
NW1

Figure 5.4: An example of packet forwarding

Packet Forwarding and FPs Maintenance

We use Fig. 5.4 to illustrate the packet forwarding and FPs maintenance processes in IARSVP. Fig. 5.4 shows that a CN initiates a session connection request to an MN when the MN visits a network NW1 served by AR$_1$. The MN then makes several moves from NW1 to NW2, NW3, then back to NW2, and finally returning to NW1. Assume that the MN acquires co-located addresses CoA$_1$, CoA$_2$ and CoA$_3$, respectively, when it first visits NW1, NW2, and NW3, and reuses CoA$_2$ and CoA$_1$ as its care-of address, respectively, when it moves back and re-visits NW2 and NW1.

As mentioned previously in Session Initiation, after the CN establishes the session with the MN, AR$_0$ will be the active FP and starts intercepting and forwarding packets destined for the MN. Therefore, AR$_0$ maintains a binding (HIP, CoA$_1$), which maps the MN's home IP (HIP) to the MN's care-of address CoA$_1$ in NW1, so that it can forward packets destined for the MN's HIP to the MN's CoA$_1$. Further assume that the MN issues an FP-MIA and finds R$_1$ as the passive FP for the MN's neighbor location in NW2, before MN moves to NW2.

Care-of Addresses Renewal

In IARSVP, the MN's FP-MIA in an active FP of an MN needs to renew, on behalf of the MN, the lease time of the MN's CoA that has a binding entry in the FP. When an MN makes several moves, a chain of active FPs may exist and these FPs work together to forward packets destined for the MN from one FP to another to the MN's current CoA. Because each active FP concerns only how to forward packets to the next downstream FP or the MN's AR, it can simply maintain a binding entry that binds the MN's old CoA to the MN's CoA next to the old one. However, an MN can not renew the lease time of the CoA it acquired in a subnet after it leaves the subnet. Therefore the MN's FP-MIA in an active FP of an MN needs to renew, on behalf of the MN, the lease time of the MN's old CoA for which the active FP has a binding entry. For the above example, the MN's FP-MIAs in R_1 and R_2 need to renew the lease times of CoA_1 and CoA_2, respectively, on behalf of the MN.

5.1.3 Qualitative analysis of FP locating

As mentioned previously, IARSVP is superior to MRSVP and HMRSVP in two aspects, namely dynamic FP locating and common route resources sharing. Dynamic FP locating enables IARSVP to find better FPs for MNs, and, together with common route resource sharing, help IARSVP to achieve better resource allocation. We explain the qualitative advantages of dynamic FP locating in this section, and demonstrate the combined effects of dynamic FP locating and common route resource sharing, which make IARSVP achieve better resource allocation, in the next section.

We first analyze the length of the path from an MN to an FP, a Gateway Mobility Agent (GMA) in HMRSVP, or a CN. Fig. 5.5 shows an abstract topology of the network under discussion. Let $Nm\text{-}c$, $Nm\text{-}g$, and $Nm\text{-}f$, represent the number of routers from an MN to a CN, a GMA, and a passive FP, respectively. In general, $Nm\text{-}c$ is much larger than both $Nm\text{-}g$ and $Nm\text{-}f$. Furthermore $Nm\text{-}g$ is normally larger than $Nm\text{-}f$ for an intra-region movement.

Table 5.1 summarizes the qualitative analysis of IARSVP, MRSVP and HMRSVP. First of all, MRSVP and HMRSVP select some specific nodes statically as FPs. MRSVP chooses either the sender's AR or the HA of a receiver MN as the FP for the MN. HMRSVP makes an improvement for intra-region movements by utilizing

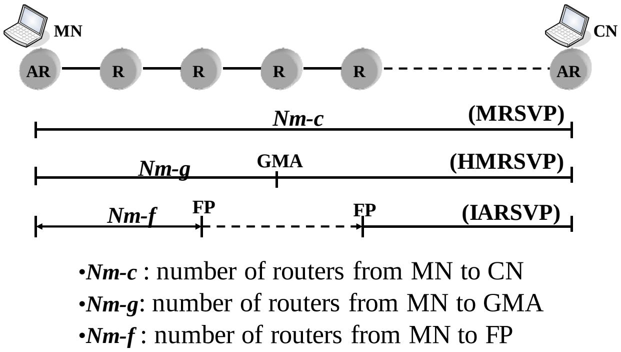

MN
CN
AR
R
R
R
R
AR
Nm-c
(MRSVP)
Nm-g
GMA
(HMRSVP)
Nm-f
FP
FP
(IARSVP)
•Nm-c : number of routers from MN to CN
•Nm-g: number of routers from MN to GMA
•Nm-f : number of routers from MN to FP

Figure 5.5: Illustration of Number of Routers

the hierarchy nature of a region network and choosing the GMA as a regional FP for an MN's intra-region movement. IARSVP takes a step further and makes use of the prefix-based routing of the Internet and locate FPs dynamically in accordance with the network topology and the available resources.

As for the inter-region movement, both MRSVP and HMRSVP choose the sender-anchor point or the MN's HA as the FP, and make passive reservations from the MN's neighbor locations to the FP, independently of the active reservation. By contrast, IARSVP dynamically locates an FP on the active reservation path and makes passive reservations from the FP to the MN's neighbor locations. Therefore, for both intra-region and inter-region movements, IARSVP can locate an FP that is closer to the MN and thus can shorten the handoff latency. Besides, by locating an FP location on the active reservation path, IARSVP can make a passive reservation using the bandwidth that is reserved by the active reservation on the route from the CN to the FP. In other words, it needs to reserve bandwidth explicitly only from the FP to the designated MN's neighbor location for a passive reservation. Therefore IARSVP consumes much less resources than both MRSVP and HMRSVP for passive reservations.

Furthermore, the hierarchical active FPs of an MN can hide the MN's mobility from the MN's HA and CNs. When an MN visits a new location, it sends only a single binding update to the active FP that is closest to the MN's new location, instead of sending a binding update to the HA and to each of the CNs. Finally,

Table 5.1: Qualitative Analysis Results in IARSVP

		MRSVP	HMRSVP	IARSVP
Forwarding Point Location		Static	Static	Dynamic
Number of Routers Involved	Intra-region Handoff $(Nm\text{-}c > Nm\text{-}g \geq Nm\text{-}f)$	High $Nm\text{-}c$	Lower $Nm\text{-}g$	Lowest $Nm\text{-}f$
	Inter-region Handoff $(Nm\text{-}c \geq Nm\text{-}f)$	High $Nm\text{-}c$	High $Nm\text{-}c$	Lowest $Nm\text{-}f$
Triangle routing		Yes	Yes	No

with MIAs, IARSVP can support route optimization whereas both MRSVP and HMRSVP cannot if they select an MN's HA as the anchor point for the MN.

5.1.4 Combined effects and simulation results

In this section, we present the simulation results of IARSVP, MRSVP and HMRSVP to illustrate the combined effects of dynamic FP locating and common route resource sharing. The simulation model adopts an 8×8 wrapped-around mesh topology as shown in Fig. 5.6 to simulate a mobile computing environment with an unbounded number of regions. Without loss of generality, we configure a two-level hierarchical infrastructure with access routers (ARs) in the lower level and gateway routers (GRs) in the higher level. The ARs and GRs represent the MAs and GMAs, respectively, in HMRSVP. However, in order to demonstrate the effects of alternative routes on resource reservations, we adopt the multi-homing concept in our simulation.

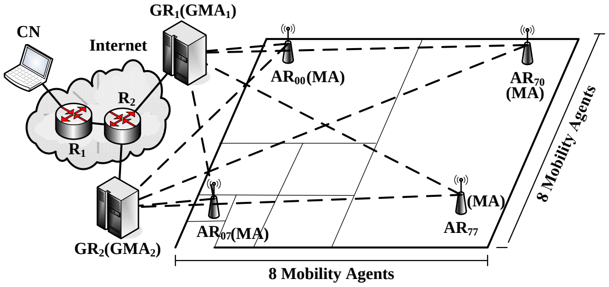


Figure 5.6: The 8×8 mesh simulation models

Therefore, all 64 ARs connect to two GRs and can access the Internet through either GR. (HMRSVP selects only one of the two GRs as the GMA of a region to serve all 64 MAs simultaneously.)

According to the wrapped-around mesh topology, when an MN moves right and away from the cell served by AR$_{70}$, an inter-region handover occurs and the MN will enter the cell served by AR$_{00}$ of a new region. Similarly, when the MN moves down and away from the cell served by AR$_{07}$, an inter-region handover occurs and the MN will enter the cell served by AR$_{00}$ of a new region. As mentioned previously, both MRSVP and HMRSVP designate a specific node as the FP of a data stream for a session. Therefore, MRSVP may select R$_1$ as the anchor point, whereas HMRSVP may select either GR$_1$ or GR$_2$, statically, as the GMA.

According to the mesh topology, each MN can make at most four resource reservations (one active reservation and at most three passive reservations). However, we assume that the bandwidth of a passive reservation in a cell, which is made by an MN in a neighbor cell, can be borrowed by another MN that is making a handoff to the cell.

We compare IARSVP with MRSVP and HMRSVP in terms of the forced termination, reservation blocking, and session completion probabilities with respect to system loads. The system load is denoted as ρ and is equal to $\frac{\lambda}{\mu}$, where the λ is the arrival rate and the μ is service completion rate of resource reservations. In this paper, we assume the inter-arrival time $(\frac{1}{\lambda})$ and the holding-time $(\frac{1}{\mu})$ of resource reservations are both an exponential distribution with a mean of $\frac{1}{\lambda}$ and $\frac{1}{\mu}$,

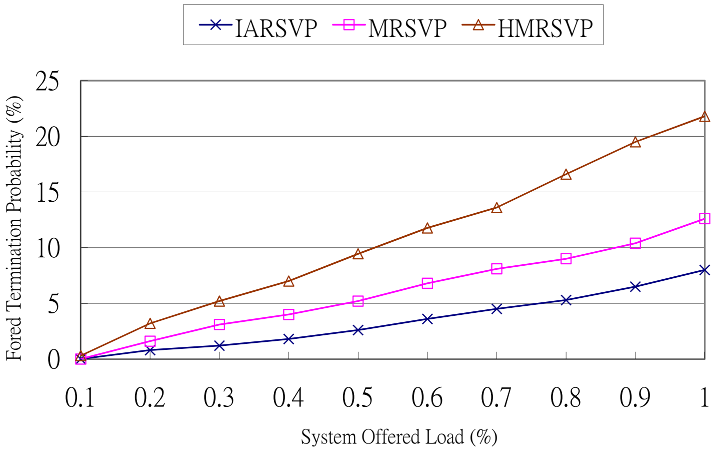


Figure 5.7: Forced termination probabilities

respectively.

Fig. 5.7 shows the forced termination probabilities for the three schemes under discussion. In general, the forced termination probabilities increase in all schemes when the offered load increases. The forced termination probability of IARSVP is lower than both MRSVP and HMRSVP, with a difference of up to 16%. This is because that IARSVP can dynamically relocate FPs and reserve resources on a more optimized route. As a consequence, it can reuse the resource of an active reservation for passive reservation and thus avoid making unnecessary resources reservation on a long-winded routing path, from a CN to an anticipated location of an MN. Therefore IARSVP can spare more bandwidth for fulfilling more resource reservation requests made by other connections. On the contrary, MRSVP and HMRSVP select FPs statically as mentioned above. In particularly, HMRSVP must initially assign a unique GMA to a region, and any packet destined for the region must travel through the GMA since the GMA hides the local movements of the MN from the nodes outside the region.

Furthermore, unlike HMRSVP, both MRSVP and IARSVP schemes can make use of the multi-homing characteristics and configure resource reservation paths through AR_1 (GMA_1) or AR_2 to the Internet. However, MRSVP makes end-to-end resource reservation for each path independently, either passive or active, from the

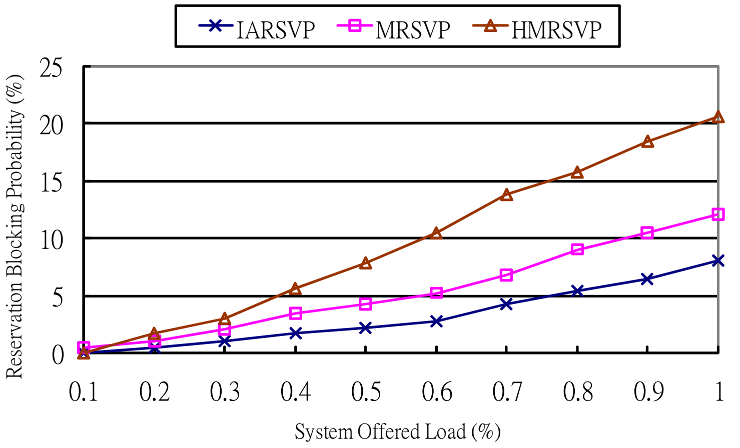


Figure 5.8: Reservation blocking probabilities

access router of a CN or the MN's HA to the current or neighbor proxies of an MN. Therefore, MRSVP may waste bandwidth on redundant reservation. On the contrary, IARSVP can make resource reservation locally from an MN's passive FP to the proxy of an anticipated location that the MN may visit next.

Fig. 5.8 presents the reservation blocking probabilities for the three schemes under discussion. It is clear that when the offered load increases, the reservation blocking probabilities increases in all schemes. According to the same reasons mentioned in the descriptions of Fig. 5.7, we can observe that the reservation blocking probability of IARSVP is lower than both MRSVP and HMRSVP, with a maximum difference of about 13%. Finally, Fig. 5.9 depicts the session completion probabilities for the three resource reservations schemes. We can further observe that the IARSVP also outperforms HMRSVP and MRSVP in terms of session completion probability due to the same reasons.

From the performance results mentioned above, we could conclude that the capability of autonomous operation of MIAs makes IARSVP better in locating FP dynamically. As a consequence, IARSVP is much superior to both the MRSVP and HMRSVP in supporting QoS-guaranteed resource reservations for real-time application in Mobile IP networks.

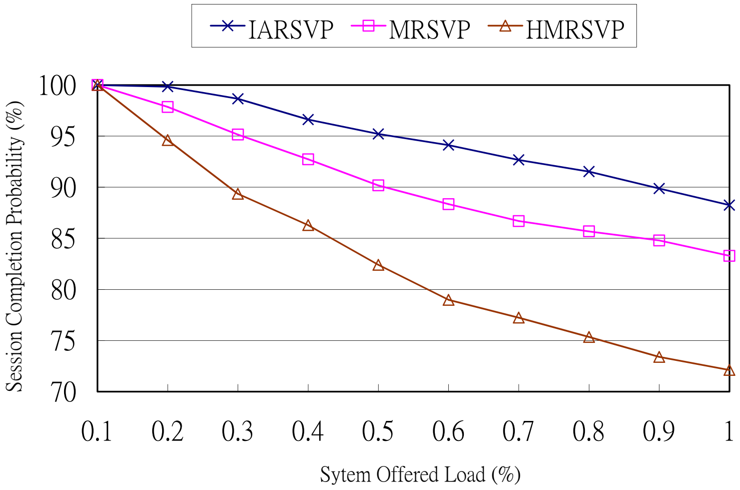


Figure 5.9: Session completion probabilities

5.1.5 Summary

we present an IARSVP scheme to enable the resource reservations and smooth handoff for Mobile IP networks. In IARSVP, MIAs can perform binding updates on behalf of MNs away from their home networks. Furthermore, the capability of autonomous operation of MIAs makes IARSVP more effective in resource reservation because MIA can adjust FPs dynamically to avoid redundant resource reservation, support route optimization and regional registration, and find the alternative routes in accordance with the network topology and resource usages. Simulation results show that the proposed approach, compared with MRSVP and HMRSVP, can significantly improve link utilization and reduce disconnection rate for MNs. In the future, we plan to propose a qualitative analysis model to evaluate the performance of IARSVP.

5.2 Mobile Bandwidth-Aggregation Reservation Scheme

A NEtwork that is MObile (NEMO) [20] usually consists of at least one MR attached to the infrastructure to manage all external communication for of all nodes inside a NEMO (NEMO nodes). Because a NEMO moves as a whole, previous mobile ReSource reserVation Protocols [42, 45] have two problems in supporting QoS for NEMOs; that is, mobility unawareness and excessive signal overhead. In this section, we first address these two problems and then propose a Mobile Bandwidth-Aggregation (MBA) reservation scheme to support QoS guaranteed services for NEMOs. In order to resolve these two problems, MBA makes an MR the proxy of all nodes insides a NEMO and has the MR aggregates and reserve the bandwidth required for all node inside a NEMO. Mathematical analysis and simulation results show that the proposed MBA scheme can significantly reduce the signal overhead for reservation maintenance. Furthermore we also present three hypothetic policies of tunnel reservations for NEMOs, and conduct simulation to evaluate these policies in terms of blocking probabilities and bandwidth utilizations.

5.2.1 Problems in supporting QoS for NEMO nodes

As mentioned previously, mobility unawareness and excessive signal overhead are the two problems in supporting QoS for NEMO nodes. A NEMO moves as a whole and the MR is the only network entity of the NEMO that is aware of the network changes. Other NEMO nodes are not aware of such integral mobility. The unawareness of integral mobility of NEMO nodes makes MRSVP or HMRSVP inapplicable to supporting QoS for NEMO nodes because both MRSVP and HMRSVP rely on the MN's ability to detect the network changes.

Even if a NEMO node could detect the network changes, integral mobility will still cause burst signal problem if each individual NEMO node establishes its own resource reservations. When a NEMO moves from one network to another all NEMO nodes that require QoS will need to make new resource reservations in the new network. The new reservations may be not necessary if the NEMO adopts MRSVP or HMRSVP and make advance reservations beforehand.

However, even with MRSVP or HMRSVP, the maintenance of individual reservation will still cause excessive signal overhead. RSVP requires the sender and the

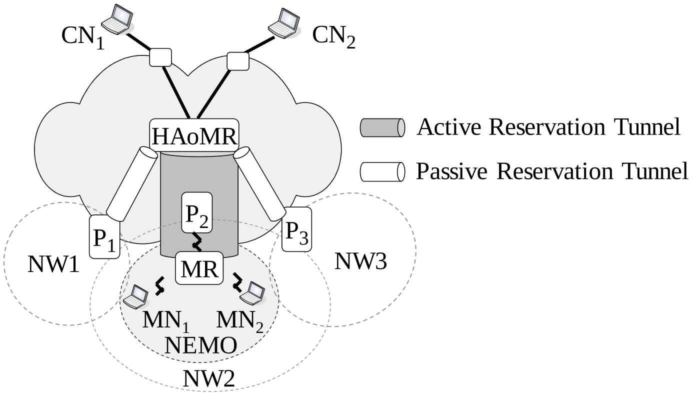


Figure 5.10: An Example of MBA scheme

receiver to issue, respectively, a Path and a Resv message periodically to refresh the reservation states. Such periodical refreshing will generate excessive amount of signals if each NEMO node maintains the resource reservations itself.

The mobility unawareness and excessive signals problems make previous mobile RSVP protocols [42, 45] inapplicable to supporting QoS for NEMO nodes. Therefore, we propose a Mobile Bandwidth-Aggregation (MBA) reservation scheme that can eliminate these two problems and support QoS for NEMO nodes.

5.2.2 The MBA Reservation Scheme

As we mentioned previously, all data traffic of a NEMO goes through the bi-directional tunnel between the MR and HAoMR. Therefore, MBA chooses the MR and HAoMR as the resource reservation proxies that handle mobility and reserve bandwidth for the bi-directional tunnel on behalf of the NEMO nodes. Therefore NEMO nodes need not be aware of the integral mobility. Furthermore, the MR and HAoMR can aggregate the Path and Resv messages to eliminate the excessive signal problem.

Principal Ideas

In this section, we use the example shown in Fig. 5.10 to explain the MBA reservation scheme. As shown in the Fig. 5.10, MN$_1$ and MN$_2$ are two NEMO nodes that attach

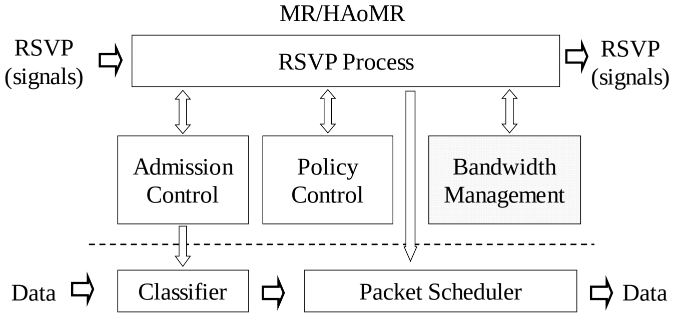


Figure 5.11: MBA Modules of MR and HAoMR

to the MR of a NEMO that is visiting a foreign network NW2. Assume that MN_1 communicates with a corresponding node CN_1 first, and then MN_2 communicates with CN_2. As mentioned previously, all data traffic of MNs goes through the bi-directional tunnel between the MR and HAoMR. In order to provide QoS for MN_1 and MN_2, the MR and HAoMR will establish an active tunnel reservation in NW2 and a passive tunnel reservation in each of the neighbor networks, NW1 and NW3.

The management of the active/passive tunnel reservations in MBA is exactly the same as the one of the active/passive reservations in MRSVP. The passive tunnel reservation of an MR on the new network will change from passive to active upon the MR moves from a network to another. At the same time, the active tunnel reservation on the old network will become a passive tunnel reservation. If the MR moves away further, the passive tunnel reservation will be obsolete and the resources reserved previously for the passive tunnel reservation will be reclaimed by the old network.

MBA Modules of MR and HAoMR

However unlike the active/passive reservations in MRSVP and HMRSVP, the tunnel reservations in the MBA reservation scheme are shared by all NEMO nodes inside a NEMO; the MR and HAoMR of the NEMO are responsible for requesting and managing the resources of the tunnel reservations and forwarding data packets on behalf of the NEMO nodes. Fig. 5.11 shows the MBA modules on an MR or HAoMR. The MBA modules consist of RSVP Message Process, Admission and Policy Control, Tunnel Bandwidth Management, and Packet Classifier and Scheduler.

When receives an RSVP Path/Resv message, RSVP message process performs RSVP Tunnelling protocol to configure a tunnel reservation and a nested end-to-end reservation simultaneously. In addition, it is also responsible for periodical refreshing of the tunnel reservations and the end-to-end reservations on behalf of the NEMO nodes.

Tunnel bandwidth management estimates the bandwidth required to fulfill the request of NEMO nodes, and informs RSVP message process to adjust the tunnel bandwidth dynamically. In addition to the aggregated bandwidth of all active sessions of NEMO nodes, tunnel bandwidth management will also request some extra bandwidth, depending on the reservation policy in use, to serve the NEMO nodes that may visit the MR in the future.

Admission control performs QoS negotiation, and decides whether to reject or grant a new session with a specific QoS class. Policy control determines how to drop the data packets to fulfill the requested QoS if the bandwidth reserved for the tunnel is not sufficient to deliver data packets in time. Packet classifier performs packet classification in accordance with the QoS classes of the data packets. Scheduler determines the packet delivery schedule based on the packet classes.

MBA Tunnel Reservations

In the following paragraphs, we use Fig. 5.12 as an example to illustrate the reservation procedures of MBA. Assume that CN_1 issues an E2E Path message to MN_1 in the NEMO. Following the NEMO Base Support Protocol [20], the message will be intercepted by HAoMR. Further assume that, MRSVP is used to reserve the resources on the path between HAoMR and CN_1. Therefore, without loss of generality, we describe only the messages flows and procedures MBA uses to reserve resources for the tunnel between HAoMR and MR, henceforth referred to as tunnel HAoMR-MR.

On receiving the E2E Path message, HAoMR will forward the message to MR via IP-in-IP encapsulation. In addition, HAoMR will also send an Active Tunnel

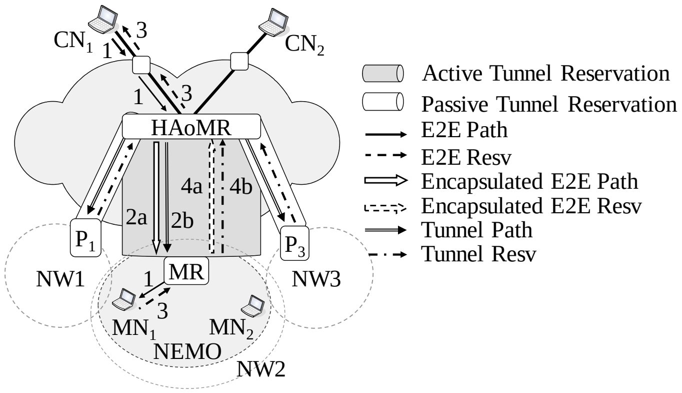


Figure 5.12: MBA Tunnel Reservations (Receiver is a NEMO node)

Path message to MR to setup an active tunnel reservation, and a Passive Tunnel Path to each of the proxies in neighbor networks, NW1 and NW3, to setup a passive tunnel reservation for MR.

Each router on the path of the tunnel forwards the encapsulated E2E Path message, as a normal IP packet, downstream to MR. On the contrary, the routers on the tunnel will recognize the tunnel Path message as an RSVP message and will perform the path finding function as described in the original RSVP protocol [9].

When MR receives the encapsulated E2E Path message, it decapsulates the message and forwards the original E2E Path message to MN_1. However, when MR receives the tunnel Path message, it will not forward the message to MN_1, but creates a soft Path state, instead, for HAoMR-MR tunnel.

In response to the original E2E Path message, MN_1 replies an E2E Resv message to CN via MR. In a similar way, when MR receives the E2E Resv message, it will encapsulate the message and forward the encapsulated E2E Resv message to HAoMR. In addition, MR will also issue a tunnel Resv message to HAoMR. Thus, all routers on the tunnel path and HAoMR, on receiving the tunnel Resv message, can reserve the desired resources for the tunnel HAoMR-MR if the resources requested by MR are available.

Moreover, when HAoMR receives the encapsulated E2E Resv message, it decapsulates the message and forwards the original E2E REsv message to CN. The E2E Resv message will inform the routers on the path from HAoMR to CN_1 to setup

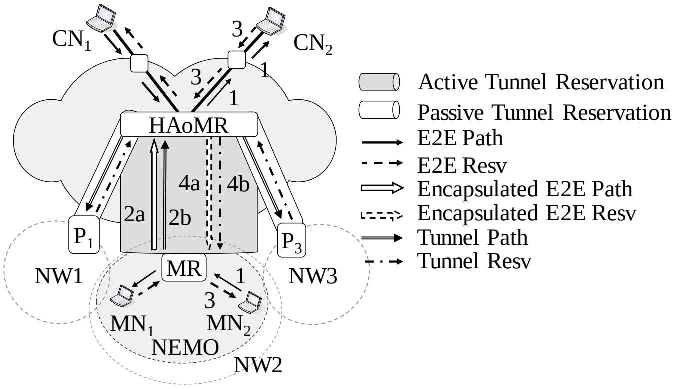


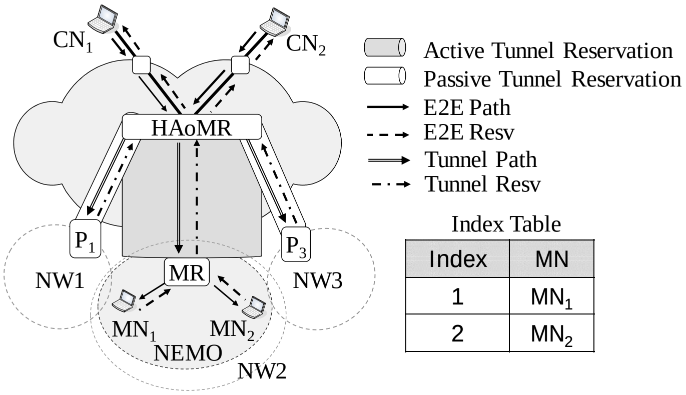
Figure 5.13: MBA Tunnel Reservation (Sender is a NEMO node)

a resource reservation for the CN$_1$-MN$_1$ session. As a consequence, an E2E reservation exists between CN$_1$ and MN$_1$ and is nested by a tunnel reservation between HAoMR and MR. Using the above nested RSVP sessions, MR can reserve resources for the HAoMR-MR tunnel and fulfill the QoS requirements of the NEMO nodes (only MN$_1$ now).

Furthermore, HAoMR will also establish two passive reservations with the proxy agents at neighbor networks of NW2 that the NEMO currently stays. Once the NEMO moves, the E2E reservations do not need to be re-established. Instead, MR and HAoMR need only cooperate to switch the passive tunnel reservation that MR visits to active and the original active tunnel reservation to passive. If MR moves further, a passive reservation may become obsolete and the routers and proxies on the path of an obsolete reservation will reclaim the resources after a predefined period.

MBA Bandwidth Adjustment

Further assume that CN$_2$ initiates a QoS session with MN$_2$ by exchanging a pair of E2E Path/Resv messages through tunnel HAoMR-MR as shown in Fig. 5.13. If MR and HAoMR have reserved bandwidth that is sufficient to fulfill the request, they simply encapsulate the E2E messages and forward the encapsulated message through the tunnel. In this case, they need not exchange tunnel Path/Resv messages



Index	MN
1	MN_1
2	MN_2

Figure 5.14: MBA Reservation Maintenance

to make a new tunnel reservation or modify the original tunnel reservation.

However if MR and HAoMR have not reserved enough bandwidth, they need to exchange a pair of tunnel Path/Resv messages to request more bandwidth for the tunnel reservation. If there is not enough bandwidth existing on the wireless links of the MR or along the path of the tunnel HAoMR-MR, the Admission Control module of MR or HAoMR will deny the request. Therefore, the policies of tunnel bandwidth reservation will affect both the signal overhead and blocking probability of the QoS sessions requested by NEMO nodes. In following, we present three hypothetic reservation policies and discuss the performance of MBA under these policies in section 5.2.3.

MBA Reservation Maintenance

If MBA uses RSVP as the underlying reservation protocol, then the tunnel reservation in MBA are soft reservations; that is, an MR and its HAoMR need to send Tunnel Path/Resv messages periodically to refresh the soft states maintained by the routers along the path of the tunnel. Because a tunnel reservation is shared by all NEMO nodes served by an MR, MBA adopts a batch refreshing scheme to reduce the signal overhead for the maintenance of tunnel reservations.

Fig. 5.14 shows the MBA reservation maintenance operations. In MBA, MRs or HAoMRs serve as proxies that refresh the E2E reservations on behalf of the NEMO

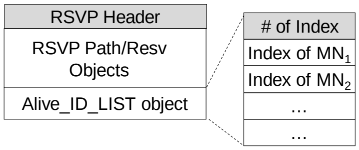


Figure 5.15: Alive_ID_List in a Tunnel Path/Resv Message

nodes and send Tunnel Path/Resv messages periodically to keep tunnel reservations alive. Therefore, in addition to the normal RSVP fields, a tunnel Path/Resv message will also carry an Alive_ID_List object as shown in Fig. 5.15. An Alive_ID_List is an index list of NEMO nodes or an index list of CNs that have an ongoing E2E session inside the tunnel HAoMR-MR.

Each index in an Alive_ID_List object will map to a soft state that contains information for an MR or HAoMR to generate an E2E Path/Resv message on behalf of an NEMO node or a CN. An MR uses the index list to inform its HAoMR the soft states of which NEMO nodes the HAoMR needs to refresh for. Similarly, the HAoMR includes an Alive_ID_List in a Tunnel Path/Resv message to inform the MR the soft states of which CNs the MR needs to refresh for.

The indices are created and refreshed periodically by an MR (or HAoMR) when the MR (or HAoMR) receives E2E Path/Resv messages. The MR (or HAoMR) reclaims the indices if it does not receive matched refreshing messages before the expiration time. In the following paragraphs, we use the above example shown in Fig. 5.12 and 5.13 to explain the creation and usages of an Alive_ID_List. Initially, when MR and HAoMR setup a tunnel reservation and a nested E2E reservation for the session CN_1-to-MN_1, MR and HAoMR will create an index for MN_1 and CN_1, respectively. Similarly, later when MR and HAoMR request more resources for the tunnel reservation and establish a nested E2E reservation for the session MN_2-to-CN_2, MR and HAoMR, respectively, will also create an index for MN_2 and CN_2.

Following the RSVP protocols, CN_1, CN_2, MN_1 and MN_2 will periodically sends E2E Path/Resv messages to refresh the E2E reservations. When MR (HAoMR) receives an E2E Path/Resv message of an ongoing session, it will not tunnel the

E2E messages to HAoMR (MR). Instead, it will simply refresh the corresponding index. The index of an active MN (CN) will appear in the Alive_ID_List object, when MR (HAoMR) sends a tunnel Path/Resv message to HAoMR (MR). Because MR and HAoMR intercept and do not forward the E2E Path/Resv messages, which are sent periodically by MNs or CNs, MBA can reduce the signal overhead significantly.

MBA Bandwidth Reservation Policies

As mentioned earlier, all NEMO nodes served by an MR share the bandwidth of the tunnel HAoMR-MR. When an MR receives a request to establish an E2E QoS session to/from an NEMO node, it needs to acquire more bandwidth to fulfill the request if it has not reserved enough bandwidth in the tunnel HAoMR-MR. Therefore, the policies of tunnel bandwidth reservations will affect both the signal overhead and blocking probability of the QoS sessions requested by NEMO nodes.

In this subsection, we present three bandwidth reservation policies for the tunnel HAoMR-MR. These policies are Static Reservation, Dynamic Reservation and Hybrid Reservation policies.

- Static Reservation Policy (SRP)

  In this policy, an MR reserves a static amount of bandwidth for NEMO nodes. With SRP, an MR reserves a constant amount of bandwidth and does not change the reserved bandwidth no matter a session begins or terminates. If the reserved bandwidth is not enough to fulfill the request of a new session, the MR will simply block the new session. This policy is simple. However, it is not flexible. It may have poor bandwidth utilization if the MR has reserved too much bandwidth or a high blocking rate otherwise. This policy is suitable for a NEMO with frequent joining and leaving of NEMO nodes, such as a bus for example.

- Dynamic Reservation Policy (DRP)

  This policy allows an MR to adjust the reserved bandwidth dynamically when necessary. Initially, an MR does not reserve any bandwidth for the tunnel

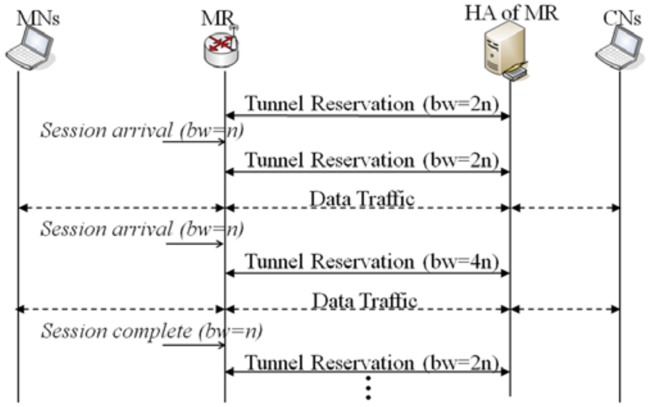


Figure 5.16: Hybrid reservation policy

HAoMR-MR if no sessions request QoS services. When a session that requests n units of bandwidth arrives, the MR would reserve n units for the tunnel to fulfill the requests. On the other hand, the MR releases the reserved bandwidth when a session terminates. If more than one session arrives or terminates at the same time, the MR needs to make only one-time bandwidth adjustment. However, frequent adjustments may still introduce huge signal overhead, and thus the policy is suitable for NEMOs with low variation in bandwidth requests, such as the NEMOs inside trains or airplanes for example. DRP has the best bandwidth utilization, but it may suffer from high blocking rates if the MR can not acquire bandwidth in time.

- Hybrid Reservation Policy (HRP)

  Hybrid Reservation Policy (HRP) is a combination of SRP and DRP. In HRP, an MR requests a static increment of bandwidth dynamically whenever it does not have spare bandwidth to accommodate new requests. Fig. 5.16 illustrates an example of how HRP works with a static increment of $2n$ bandwidth. Initially, the MR reserves a static amount of $2n$ bandwidth, which is enough to fulfill the requests of the first two sessions. After the second requests, MR requests to increase the bandwidth to $4n$ in total. Similarly, the departure of two sessions will trigger MR to release the reserved bandwidth. HRP can provide higher flexibility by altering the value of static increment, and could be proper for both high and low variation NEMOs.

5.2.3 Performance evaluation

As mentioned previously, previous mobile RSVP protocols possess a mobility unawareness problem and are not applicable to NEMOs. Therefore, we proposed the MBA scheme that can resolve the problem. Furthermore, the MBA scheme can also reduce the signal overhead in reservation maintenance. Therefore, we would like to study the effects of MBA on reducing the signal overhead. Besides, we also conduct a preliminary evaluation of the blocking probabilities and resource utilization of the three reservation policies. Furthermore, the MR in MBA is similar to the MN in MRSVP [42] and Hierarchical MRSVP [45]. As a consequence, the forced termination probabilities, during handoff, will be the same as the ones in MRSVP or Hierarchical RSVP, depending on the underlying mobility scheme. Therefore, for the readers who are interested at the forced termination probabilities for handoff, please refer to MRSVP [42] and Hierarchical MRSVP [45] for the performance results.

Mathematics Analysis

In this section, we present an analysis model to evaluate the signal overhead of MBA and use simulation to verify the analytical results. In MBA, an MR and its HAoMR aggregate the bandwidth required for the NENO nodes and reserve bandwidth for the HAoMR-MR tunnel. Furthermore the HAoMR-MR tunnel is transparent to NEMO nodes, and thus NEMO nodes in MBA could behave exactly the same as the nodes with the ordinary RSVP. Therefore, in the following analysis, we focus on the performance analysis of the HAoMR-MR tunnels only. We assume that the inter-arrival time of QoS session requests and the reservation holding time (the connection service time) both follow an exponential distribution with a mean of $\frac{1}{\lambda}$ and $\frac{1}{\mu}$, respectively. We first present the analytical models for the sessions with requests of identical bandwidth, and with requests of multiple classes, in which

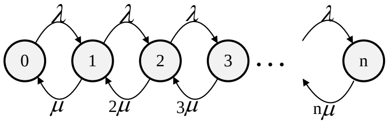


Figure 5.17: State transition diagram for identical requests

each QoS session can request up to a predefined amount of bandwidth. From the mathematic models, we can derive the signal overhead for reservation maintenance in a NEMO and the blocking probabilities that an MR may reject a reservation request, respectively, in sections 5.2.3 and 5.2.3.

- Sessions with Identical Requests

  In this case, we assume all sessions have the same QoS requirement and each session requests a single unit of bandwidth. Furthermore, because the reservation policies make no difference in signal overhead for reservation maintenance, we also assume the MR uses SRP and reserve n units of bandwidth beforehand.

  Fig. 5.17 shows the state transition diagram for the case of identical requests. In the figure, each state represents the number of QoS sessions the MR has already offered. Because the MR has reserved only n units of bandwidth for the tunnel, it will block new requests when there exists n active sessions.

  We could derive the probability of each state as follows.

$$p_k = \begin{cases} \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k p_0 & k \leq n \\ 0 & k > n \end{cases} \tag{5.1}$$

  where P_0 is the probability of the initial state and the equation of P_0 is

$$p_0 = \frac{1}{\sum_{i=0}^{n}\left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}} \tag{5.2}$$

  Next, we can calculate the average number of active sessions, m, in a NEMO as follows.

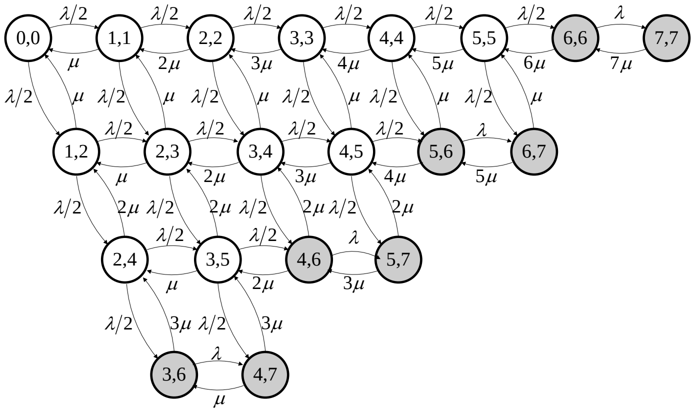


Figure 5.18: State transition diagram $(n = 7, k = 2)$

$$m = \sum_{i=0}^{n} i \times P_i \tag{5.3}$$

- Sessions with Multi-Class Requests

We further analyze the signal overhead for the case of Multi-class requests where active sessions could request different number of bandwidth units. Multi-class requests are more realistic because a session might request a larger bandwidth for video services or a smaller bandwidth for voice or text services. However, this case is more complicate since the number of active sessions is no longer equal to the bandwidth units reserved by an MR. Therefore we can use a two-tuple state transition diagram, as shown in Fig. 5.18, to model the reservation behavior of active sessions. In the diagram, a state (i, j) means that the MR has accepted i sessions and has reserved j units of bandwidths in total for all sessions. For example, the state $(2, 3)$ represents two sessions are currently active and have jointly requested 3 units of bandwidth.

Assume that n is the bandwidth an MR has reserved, and k is the maximal number of bandwidth units a session can request. Fig. 5.18 shows the state transition diagram when $n = 7$ and $k = 2$. The initial state of the system is $(0, 0)$. Let λ and μ represent the arrival and departure rates, respectively, of a new session. Suppose each session can request one or two units of bandwidth

Table 5.2: One-step transition matrix in MBA

	(0,0)	(1,1)	(1,2)	(2,2)	$\cdots$	$(6,6)$	$(6,7)$	$(7,7)$
(0,0)	0	$\frac{1}{2}\lambda$	$\frac{1}{2}\lambda$	0	$\cdots$	0	0	0
(1,1)	μ	0	0	$\frac{1}{2}\lambda$	$\cdots$	0	0	0
(1,2)	μ	0	0	0	$\cdots$	0	0	0
(2,2)	0	2μ	0	0	$\cdots$	0	0	0
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$(6,6)$	0	0	0	0	$\cdots$	0	0	λ
$(6,7)$	0	0	0	0	$\cdots$	0	0	0
$(7,7)$	0	0	0	0	$\cdots$	7μ	0	0

with the same probability. A state (i, j) may transit to states $(i+1, j+1)$ or $(i+1, j+2)$ with an equal transition rate of $\frac{\lambda}{2}$, except when $(j+1)$ is equal to or greater than 7. For example, if the current state is (2,3), the new state will be either (3,4) or (3,5) when the MR accepts a new session. However, if the current state is (5,6) the MR will accept only the session with one unit request. Furthermore if the current state is (5,7), the MR will not accept any new session request since it has already used up all bandwidth. Similarly, when an active session terminates, it will release one or two units of bandwidth with an equal probability of $\frac{1}{2}$.

From the above state transition diagram, we can obtain a corresponding one-step transition matrix as shown in Table 5.2. The one-step transition matrix can be further normalized as Table 5.3 by applying by Jensen algorithm [28].

Let $P^{(1)}$ denote the normalized one-step transition matrix. We can apply $P^{(1)}$ n times to obtain the n-step transition matrix, $P^{(n)}$, as

Table 5.3: Normalized one-step transition matrix in MBA

	(0,0)	(1,1)	(1,2)	(2,2)	$\cdots$	$(6,6)$	$(6,7)$	$(7,7)$
(0,0)	$1-\frac{1}{7}\rho$	$\frac{1}{14}\rho$	$\frac{1}{14}\rho$	0	$\cdots$	0	0	0
(1,1)	$\frac{1}{7}$	$\frac{6}{7}-\frac{1}{7}\rho$	0	$\frac{1}{14}\rho$	$\cdots$	0	0	0
(1,2)	$\frac{1}{7}$	0	$\frac{6}{7}-\frac{1}{7}\rho$	0	$\cdots$	0	0	0
(2,2)	0	$\frac{2}{7}$	0	$\frac{5}{7}-\frac{1}{7}\rho$	$\cdots$	0	0	0
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$(6,6)$	0	0	0	0	$\cdots$	$\frac{1}{7}-\frac{1}{7}\rho$	$\frac{1}{7}\rho$	0
$(6,7)$	0	0	0	0	$\cdots$	0	$\frac{1}{7}$	0
$(7,7)$	0	0	0	0	$\cdots$	1	0	0

$$p^{(n)} = \overbrace{p^{(1)} \times p^{(1)} \times \cdots \times p^{(1)}}^{n\ times} \tag{5.4}$$

Furthermore, let $\pi = (1, 0, \ldots, 0)$ be the vector that represents the state probabilities before any transition, and $\pi^{(n)}$ be the state probabilities after n-step transition. Then, we have $\pi^{(n)} = \pi^{(0)} \times P^{(n)}$. By taking the limit of n, we can obtain the steady state probability vector, π, as follows.

$$\pi = \lim_{n=\infty} \pi^{(n)} = \pi^{(0)} \times \lim_{n=\infty} p^{(n)} \tag{5.5}$$

From the steady state probabilities π, we can derive the average number of active sessions an MR can offer as follows.

$$m = \sum_{i=0}^{n} i \sum_{j=i}^{min(2i,n)} P(i,j) \tag{5.6}$$

Signal Overhead

If we assume each NEMO node runs RSVP itself and each session could either be an outgoing or an incoming session with an equal probability, then the signal overhead for reservation maintenance in RSVP, denoted as O_{RSVP}, can be represented as equation 5.7.

$$O_{RSVP} = \frac{(S_P + S_R) \times \frac{1}{2} \times m \times 8}{T} \ \ bps \tag{5.7}$$

where S_P and S_R are the sizes of RSVP Path and Resv message in octets, respectively; T is the interval of sending RSVP messages in seconds, and m is the average number of active sessions in a NEMO. By replacing m with equations 5.3 and 5.6, respectively, we can obtain the equations of O_{RSVP} for the cases of identical requests and multi-class requests.

On the contrary, in MBA, an MR aggregates the reservations, and issues a single Tunnel Path/Resv, respectively, to refresh the outgoing/incoming tunnel reservations. Therefore the signal overhead the proposed MBA introduces could be represented as follows in both cases.

$$O_{MBA} = \frac{(S_P + S_R) \times \frac{1}{2} \times 1 \times 8}{T} \ \ bps \tag{5.8}$$

Without security and identification requirements, the default value of S_P is 172 byes, S_R is 92 bytes, and T is 30 seconds [31]. We apply these default values and obtain the signal overhead over the system load as shown in Fig. 5.19. From the figure, it is obvious that the signal overhead for establishing a tunnel in MBA is about 35 bps. In addition, we could observe that the signal overhead in RSVP increases as the system load, i.e. average number of active sessions, increases, while the one in MBA retains at the same level, independent of the system load. Furthermore, with the same bandwidth reserved by a NEMO, the NEMO could serve less active sessions for the case of multiple-class requests, and consequently, for RSVP, the signal overhead in this case is comparatively smaller than the one of identical requests.

We also run simulations to verify the above analysis. As shown in the figure, the

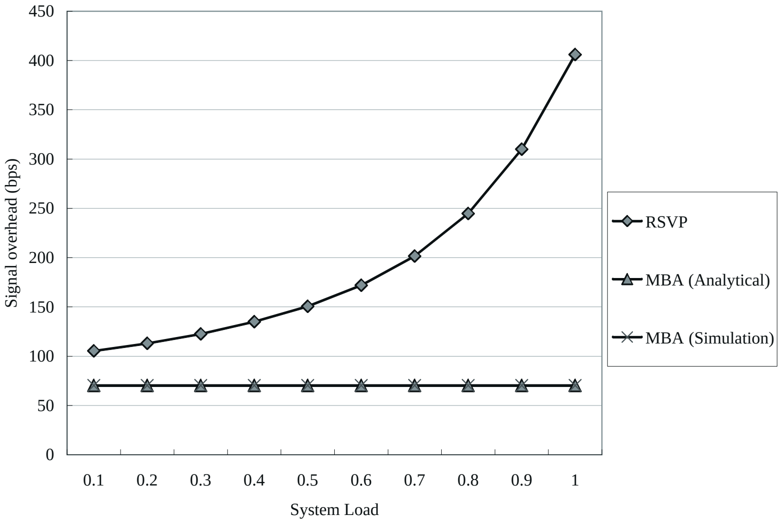


Figure 5.19: Signal overhead for identical requests

simulation result of MBA is identical with the analytical one. The simulation will be explained in section 5.2.3.

Blocking Probability

Because we assume the MR uses SRP for tunnel reservations in the above analysis, the blocking probability, P_{block}, for the sessions with identical requests equals to P_n, and can be derived from Erlang-B serving model as follows:

$$P_{block} = p_n = \frac{\frac{\left(\frac{\lambda}{\mu}\right)^n}{n!}}{\sum_{i=0}^{n} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}} \tag{5.9}$$

For the sessions with multi-class requests, the equation of P_{block} is more complicate. A multi-class request will be blocked by the MR if it has requested more units than the remaining units of bandwidth. Therefore, the shadowed states, in Fig. 5.18, represent the ones where a blocking may occur. The shadowed states and equation of P_{block} can be derived as follows.

$$P_{block} = \sum_{i=0}^{n} p_{(i,n)} + \sum_{i=0}^{n-1} \frac{1}{k} p_{(i,n-1)} + \sum_{i=0}^{n-2} \frac{2}{k} p_{(i,n-2)}$$

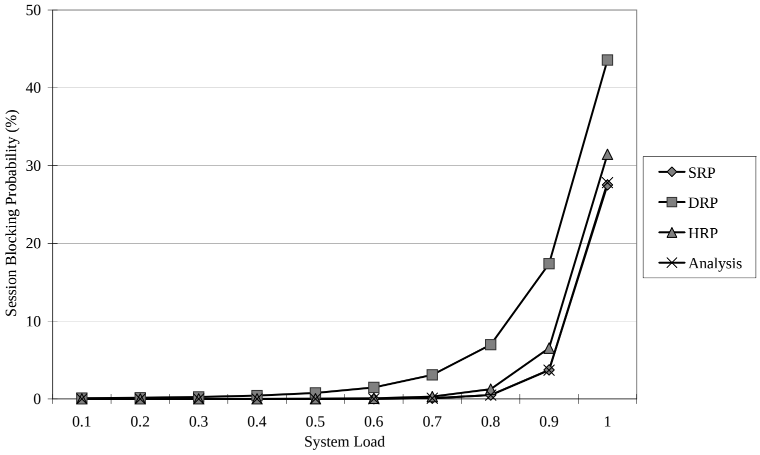


Figure 5.20: Blocking probability for identical requests

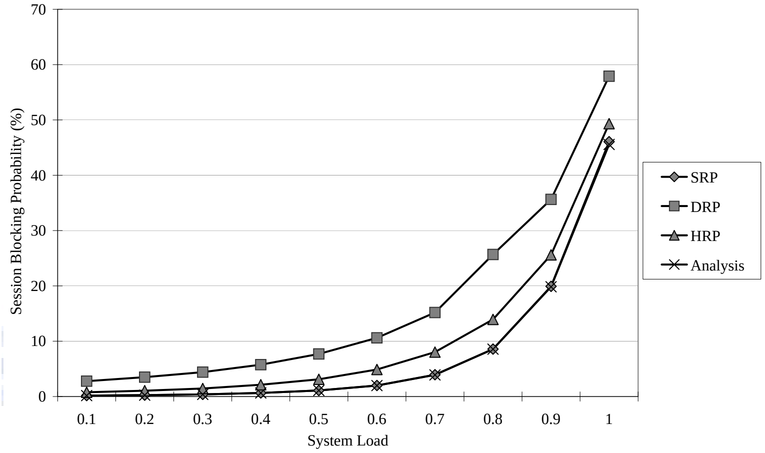


Figure 5.21: Blocking probabilities for multi-class requests

$$+\cdots+\sum_{i=0}^{n-(k-1)}\frac{k-1}{k}p_{(i,n-(k-1))} \tag{5.10}$$

where $\frac{j}{k}\times p(i,n-j)$ represents the probability that a new request will be blocked by the MR in state $(i,n-j)$. By summing the blocking probabilities of all shadowed states, we can obtain the equation 5.10 for the blocking probability, P_{block}. In the equation, we assume the probabilities of the non-existing states are all zeros.

Fig. 5.20 and 5.21 plot the blocking probabilities for different tunnel reservation policies under various system loads $(\frac{\lambda}{\mu})$. Because in DRP, the MR will make a new reservation request when it receives a new reservation request, we use DRP to simulate the behavior of RSVP. In both figures, we can easily observe that the proposed MBA scheme is better than the original RSVP protocol. Furthermore, for the proposed MBA scheme, the blocking probabilities for multi-class requests are higher than the ones for identical requests.

Simulation Results

In addition to the mathematical analysis, we also run simulations and compare the results with the ones of analytical model for SRP tunnel reservations. In this section, we first explain how we conduct the simulations and then present the utilization of the bandwidth reserved by using the three reservation policies, namely SRP, DRP, and HRP.

In the simulation, we generate background traffic randomly with a mean 40% total bandwidth on the external link of the MR. Similar to the above analytical model, we assume the arrival and departure rates follow the Poisson distribution with a mean of λ and μ, respectively. The total bandwidth of the MR's external link is 10 units and we set $k = 2$ for the multi-class requests.

Fig. 5.20 and 5.21, respectively, show the blocking probabilities for the identical requests and multi-class requests under various system loads of the MR. In general, the blocking probability increases as the system load increases. SRP has the lowest blocking probability since it reserves maximum available bandwidth in the beginning. DRP has the worst blocking probability because it reserves bandwidth dynamically when the MR (HAoMR) receives a connection request and may fail in competing bandwidth with the background traffic. Furthermore, different values of the background traffic will shift the curves of blocking probabilities horizontally.

On the contrary, as shown in Fig. 5.22 and 5.23, DRP always has 100% of bandwidth utilization since it reserves bandwidth only when necessary. SRP, oppositely, has the lowest bandwidth utilization, especially during light system load. On average, the bandwidth utilization of multi-class requests is lower than the one of identical requests.

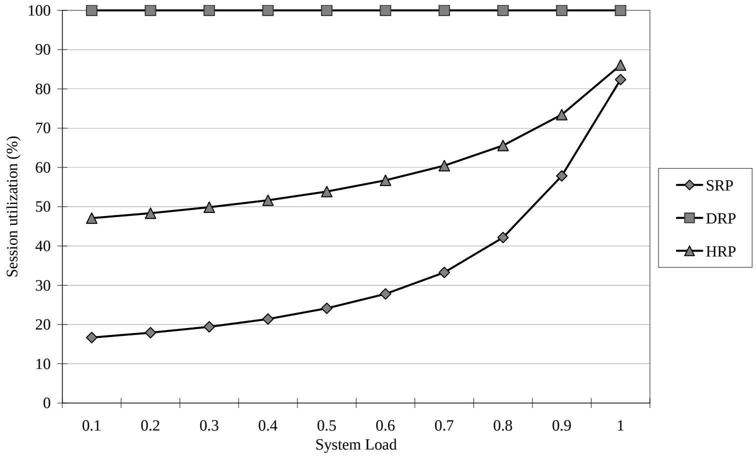


Figure 5.22: Bandwidth utilization for identical requests

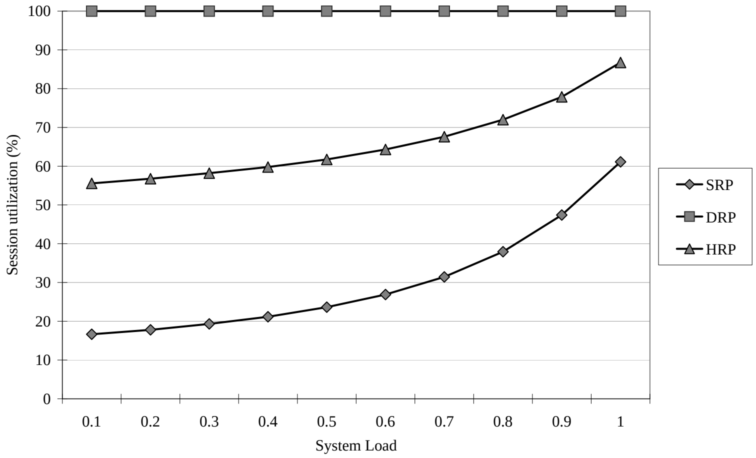


Figure 5.23: Bandwidth utilization for multi-class requests

5.2.4 Summary

We propose an MBA scheme to support QoS for NEMO nodes. The proposed MBA scheme makes the MR and HAoMR the resource reservation proxies that handle mobility and reserve bandwidth on behalf of the NEMO nodes. Therefore the MBA scheme can resolve the mobility unawareness and excessive signal overhead in supporting QoS for NEMOs. Mathematical analysis and simulation results show that the proposed MBA scheme can significantly reduce the signal overhead for reservation maintenance. Although we describe the MBA scheme in the context of RSVP, the concept of MBA can work with other resources reservation protocols as well. Furthermore the policies of tunnel bandwidth reservations may affect the blocking probabilities and bandwidth utilization, and thus need further investigation in the future. Furthermore, MBA could incorporate some administration control to drop some ongoing session of NEMO nodes if it demands more bandwidth than whatever the external link can offer. In the future, we will first design the admission control and administration mechanism for MBA, and then implement the proposed MBA scheme.

Chapter 6

Conclusions

In this thesis, we propose several approaches to improve the fast handoff performance while fulfilling the security and network resource requirement for MN and NEMO. In order to improve the time-consuming AKA process, we propose two Group-based AKA schemes, Group Key-based AKA scheme and Group Signature-based AKA scheme, to shorten authentication process. The G-AKA schemes can accelerate the handoff process to avoid real-time service interruption time caused by long authentication latency while maintaining the same secure level as the other AKA protocols do. Due to the characteristic of sharing group authentication data, the proposed G-AKA also can save storage for authentication data in the SNs.

For the longer re-authentication latency problem, we present an integrated security domain (ISD) for WLAN Mesh to support MN inter-AP fast handoff. The ISD mechanism can integrate the security domain of WLAN Mesh and remove the overhead caused by link layer security protocols.

As for excessive resource reservations, we present an IARSVP approach that can support QoS aware packet transmissions for MIP networks. With MIAs, IARSVP can allocate resources in advance for neighbor locations an MN may visit next. Since MIAs carry MN's mobility security association, QoS requirement and administration specification, and associated executable codes, they can perform location updates for MNs, adjust autonomously in accordance with the network topology and resource usage when locating the forwarding points for the MN. Therefore, IARSVP can avoid redundant resource reservations made in common routes, support route optimization and regional registration naturally, and discover alternative routes dynamically.

In addition, we propose an MBA reservation scheme to support QoS guaranteed services for NEMOs. MBA makes an MR the proxy of all nodes inside a NEMO to aggregate and reserve the bandwidth required for them. Mathematical analysis and simulation results show that the proposed MBA scheme can significantly reduce the signal overhead for reservation maintenance. Furthermore we also present three hypothetic policies of tunnel reservations for NEMOs, and conduct simulation to evaluate these policies in terms of blocking probabilities and bandwidth utilizations.

The adoption of the aforementioned schemes supports QoS of real-time services during MN's fast handoff. While moving to different base station, MN can not only authenticate and establish security association with a short interruption time, but also establish immediately a QoS reservation for real-time services in the target location.

Bibliography

[1] 3rd Generation Partnership Project. *Security threats and requirements.* 3GPP TR 21.133, 1999.

[2] I. Akyildiz, X. Wang, and W. Wang. Wireless mesh networks: a survey. *Computer networks*, 47(4):445–487, March 2005.

[3] I. F. Akyildiz, Y. B. Lin, W. R. Lai, and R. J. Chen. A new random walk model for pcs networks. *IEEE journal on selected areas in communications*, 18(7):1254–1260, July 2000.

[4] J. Arkko and H. Haverinen. *Extensible authentication protocol method for 3rd generation authentication and key agreement (EAP-AKA).* RFC 4187, IETF working group, January 2006.

[5] G. Ateniese, J. Camenisch, M. Joye, and G. Tsudik. A practical and provably secure coalition-resistant group signature scheme. In *Proceedings of 20th annual international cryptology conference on advances in cryptology*, volume 1880, pages 255–270, 2000.

[6] G. Ateniese and G. Tsudik. Some open issues and new directions in group signatures. *Lecture notes in computer science*, 1648:196–211, 1999.

[7] M. S. Bargh, R. J. Hulsebosch, E. H. Eertink, A. Prasad, H. Wang, and P. Schoo. Fast authentication methods for handovers between ieee 802.11 wireless lans. In *Proceedings of the 2nd ACM international workshop on wireless mobile applications and services on WLAN hotspots*, pages 51–60, October 2004.

[8] B. Braden, D. Clark, and S. Shenker. *Integrated services in the internet architecture: an overview.* RFC 1633, IETF working group, June 1994.

[9] B. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. *Resource reservation protocol (RSVP) version 1 functional specification.* RFC 2205, IETF working group, September 1997.

[10] M. Breugst and T. Magedanz. Mobile agent-enabling technology for active intelligent network implementation. *IEEE network*, 12(3):53–60, May 1998.

[11] G. Cabri, L. Leonardi, and F. Zambonelli. Agents for information retrieval: issues of mobility and coordination. *Journal of systems architecture*, 46(15):1419–1433, December 2000.

[12] J. Camenisch and M. Michels. A group signature scheme with improved efficiency. In *Proceedings of the international conference on the theory and applications of cryptology and information security*, volume 1514, pages 160–174, 1998.

[13] J. L. Camenisch. Efficient and generalized group signatures. *Lecture notes in computer science*, 1233:465–479, 1997.

[14] J. L. Camenisch and M. A. Stadler. Efficient group signature schemes for large groups. *Lecture notes in computer science*, 1294:410–424, 1997.

[15] G. D. Caro and M. Dorigo. Mobile agents for adaptive routing. In *Proceedings of 31th annual hawaii international conference on system sciences*, page 74, 1998.

[16] D. Chaum and E. V. Heyst. Group signatures. *Lecture notes in computer science*, 547:257–265, 1991.

[17] L. Chen and T. P. Pedersen. New group signature schemes. *Lecture notes in computer science*, 950:171–181, 1995.

[18] D. Chess, C. Harrison, and A. Kershenbaum. *Mobile agents: are they a good idea?* IBM research report RC19887, December 1994.

[19] D. Dasgupta and H. Brian. Mobile security agents for network traffic analysis. In *DARPA information survivability conference and exposition*, volume 02, page 1332, 2001.

[20] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert. *Network mobility basic support protocol.* RFC 3963, IETF working group, January 2005.

[21] B. Emako, R. H. Glitho, and S. Pierre. A mobile agent-based advanced service architecture for wireless internet telephony: design, implementation, and evaluation. *IEEE transactions on computers*, 52(6):690–705, June 2003.

[22] M. Gast. *802.11 wireless networks: the definitive guide, second edition.* O'Reilly, USA, April 2005.

[23] T. Gschwind, M. Feridun, and S. Pleisch. ADK: building mobile agents for network and system management from reusable components. In *Proceedings of first international symposium on agent systems and applications*, pages 13–21, 1999.

[24] E. Gustafsson, A. Jonson, and C. E. Perkins. *Mobile IP regional registration.* Internet-Draft: draft-ietf-mobileip-reg-tunnel-08.txt, 2003.

[25] H. Haverinen and J. Salowey. *Extensible authentication protocol method for global system for mobile communication (GSM) subscriber identity modules (EAP-SIM).* RFC 4186, IETF working group, January 2006.

[26] C. M. Huang and J. W. Li. Authentication and key agreement protocol for umts with low bandwidth consumption. In *Proceedings of 19th conference on AINA*, volume 1, pages 392–397, May 2005.

[27] R. H. Jan and Y. C. Huang. Fast pre-authentication based on ieee 802.11i. In *The 2nd workshop on wireless ad hoc and sensor networks*, pages 317–324, August 2006.

[28] Jensen. *Markoff chains as an aid in the study of markoff process.* Skandinavisk aktuarietidskrift, 1953.

[29] D. B. Johnson, C. E. Perkins, and J. Arkko. *Mobility support in IPv6.* RFC 3775, IETF working group, June 2004.

[30] I. Kim and K. Kim. Local authentication mechanism for micro mobility in wireless active network environment. In *Proceedings of 8th international conference of advanced communication technology*, volume 2, pages 20–22, February 2006.

[31] J. Manner and X. Fu. *Analysis of existing quality of service signaling protocols.* RFC 4094, IETF working group, May 2005.

[32] N. Migas, W. J. Buchanan, and K. A. McArtney. Mobile agents for routing, topology discovery, and automatic network reconfiguration in ad-hoc networks. In *Proceedings of 10th IEEE international conference and workshop on the engineering of computer-based systems*, pages 200–206, April 2003.

[33] N. Minar, K. H. Kramer, and P. Maes. *Cooperating mobile agents for dynamic network routing.* Springer-verlag, 1999.

[34] G. E. Montenegro. *Reverse tunneling for mobile IP.* RFC 3024, IETF working group, January 2001.

[35] S. Papavassiliou, A. Puliafito, O. Tomarchio, and J. Ye. Mobile agent-based approach for efficient network management and resource allocation: framework and applications. *IEEE journal on selected areas in communications*, 20(4):858–872, May 2002.

[36] M. K. Perdikeas, F. G. Chatzipapadopoulos, I. S. Venieris, and G. Marino. Mobile agents standards and available platforms. *Computer networks*, 31(19):1999–2016, August 1999.

[37] C. E. Perkins. *IP mobility support.* RFC 2002, IETF working group, October 1996.

[38] V. A. Pham and A. Karmouch. Mobile software agents: anoverview. *IEEE communications magazine*, 36(7):26–37, July 1998.

[39] E. Rescorla. *Diffie-Hellman key agreement method.* RFC 2631, IETF working group, June 1999.

[40] W. A. Simpson. *PPP challenge handshake authentication protocol (CHAP).* RFC 1994, IETF working group, August 1996.

[41] D. X. Song. Practical forward secure group signature schemes. In *Proceedings of 8th ACM conference on computer and communications security*, pages 225–234, November 2001.

[42] A. K. Talukdar, B. R. Badrinath, and A. Acharya. MRSVP: a resource reservation protocol for an integrated services network with mobile hosts. *Wireless networks*, 7(1):5–19, 2001.

[43] C. H. Tan and J. C. M. Teo. An authenticated group key agreement for wireless networks. In *Proceedings of wireless communications and networking conference*, volume 4, pages 2100–2105, March 2005.

[44] A. R. Tripathi, T. Ahmed, and N. M. Karnik. Experiences and future challenges in mobile agent programming. *Microprocessors and microsystems*, 25(2):121–129, April 1990.

[45] C. C. Tseng, G. H. Lee, and R. S. Liu. HMRSVP: a hierarchical mobile RSVP protocol. *Wireless networks*, 9:95–102, March 2003.

[46] D. M. Wallner, E. J. Harder, and R. C. Agee. *Key management for multicast: issues and architectures.* RFC 2627, IETF working group, June 1999.

[47] C. K. Wong, M. Gouda, and S. S. Lam. Secure group communications using key graph. *IEEE/ACM transactions on networking*, 8(1):16–30, February 2000.

[48] IEEE 802.11 working group. *Amendment 6: medium access control (MAC) security enhancements.* IEEE standard 802.11i, July 2004.

[49] IEEE 802.11 working group. *Port-based network access control.* IEEE standard 802.1X, December 2004.