WIRELESS IP NETWORK MOBILITY MANAGEMENT: ADVANCING FROM MOBILE IP TO HIP-BASED NETWORK

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

JOSEPH YICK HON SO
B.Eng

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
SCIENCE, ENGINEERING AND TECHNOLOGY PORTFOLIO
RMIT UNIVERSITY
AUGUST 2008
Dedicated to my parents
Statement of Original Authorship

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; and any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Signature: _________________________
Joseph Yick Hon So

Date: 28/08/2008
Acknowledgments

This research project would not have been possible without the support of my supervisors. I am deeply indebted and grateful to my supervisor, Dr Jidong Wang for his patient guidance. His encouragement and advice throughout this study are important for me. Special thanks to him for correcting my technical writing style.

I would like to express my greatest thanks to Dr Peterjohn Radcliffe for supporting me in this project and helping me to solve the problems and difficulties which I faced during the period of this research.

I am thankful to both Dr Jidong Wang and Dr Peterjohn Radcliffe for editing my thesis writing, and Prof Andrew Jennings and Dr Heiko Rudolph for their generous assistance during the time.

Last, but not least, this dissertation is dedicated to my parents, Mr Chau Ping So and Mrs Kia Ping Kam, for their support and encouragement.
Keywords

Host Identity Protocol (HIP), Micro-HIP (mHIP), Micro-mobility Management, Hybrid SIP-HIP (SHIP), Handover, Intra-Domain Handover, Mobile IP, Mobility Management, Public Key based Network, Session Initial Protocol (SIP), 3G
Abstract

Wireless networking introduces a whole range of challenges to the traditional TCP/IP network. In particular, IP address the issue of overloading because IP addresses are used as a network locator and an end point identity in the different layers in an OSI model. Even though Mobile IP is widely deployed, it has significant problems relating to performance and security. The Host Identity Protocol (HIP) provides secure mobility management by solving the IP address overloading from another angle. It restructures the TCP/IP model and introduces a new layer and a new namespace. The performance of HIP has proven to be better than Mobile IP and also opens a range of new research opportunities.

This dissertation proposes and analyses a new step-stone solution from the Mobile IP-based network into a HIP-based network. The main advantage of this new solution is that much less change is required to the operating system kernel of the end point compared to a full HIP implementation. The new step-stone solution allows Mobile IP to use some HIP features to provide better security and handover performance. This dissertation also proposes several new and novel HIP-based wireless communication network architectures. An HIP-based heterogeneous wireless network architecture and handover scheme has been proposed and analysed. These schemes limit the HIP signalling in the wireless network if no communication to external networks is needed. Beside the network architecture modification, the hybrid Session Initial Protocol (SIP) and HIP-based Voice over IP (VoIP) scheme is proposed and analysed. This novel scheme improves the handover latency and security.

This dissertation also proposes and analyses a new and novel extension to HIP, a HIP-based micro-mobility management, micro-HIP (mHIP). mHIP provides a new secure framework for micro-mobility management. It is a more complete HIP-based micro-mobility solution than any other proposed in existing studies. mHIP improves the intra-domain handover performance, the security, and the distribution of load in the intra-domain handover signalling.
These concepts have been the basis of four conference and two journal papers as listed below:


The new work presented opens up a number of very interesting research opportunities.
# Table of Contents

Statement of Original Authorship .............................................................................i  
Acknowledgments ........................................................................................................ii  
Keywords .......................................................................................................................iii  
Abstract .......................................................................................................................iv  
Table of Contents .........................................................................................................vi  
List of Figures ..............................................................................................................x  
List of Tables ...............................................................................................................xiii  
List of Abbreviations ...................................................................................................xiv  

1 CHAPTER 1: INTRODUCTION ................................................................................. 1  
1.1 Background ..........................................................................................................1  
1.2 Thesis Goals & Significance ................................................................................2  
1.3 Thesis Outline .......................................................................................................4  

2 CHAPTER 2: MOBILITY MANAGEMENT OVERVIEW .......................................... 7  
2.1 Wireless Network History ....................................................................................7  
2.2 Wireless Technology ............................................................................................9  
2.2.1 3G standard ....................................................................................................9  
2.2.1.1 3GPP Overview .......................................................................................9  
2.2.1.2 3GPP2 Overview ..................................................................................11  
2.2.2 IEEE 802 Wireless Standard ........................................................................12  
2.2.2.1 IEEE 802.11 ..........................................................................................12  
2.2.2.2 IEEE 802.16 ..........................................................................................13  
2.2.2.3 IEEE 802.20 ..........................................................................................13  
2.3 Mobility Management in IP Network ................................................................14  
2.3.1 IETF Mobility Solution ................................................................................14  
2.3.1.1 Mobile IP ...............................................................................................14  
2.3.2 Session Initial Protocol (SIP) ........................................................................16  
2.3.2.1 Micro-mobility Management ................................................................17  
2.4 Summary .............................................................................................................18  

3 CHAPTER 3: HOST IDENTITY PROTOCOL (HIP) ................................................... 19  
3.1 HIP Overview ......................................................................................................20  
3.2 HIP Architecture ................................................................................................20  
3.2.1 Rendezvous Server (RVS) ............................................................................21  
3.2.2 HIP Packet .....................................................................................................22  
3.3 HIP Base Exchange ..............................................................................................24  
3.4 HIP mobility management and mulithoming .....................................................25  
3.4.1 Pre-session Mobility Management ..................................................................25  
3.4.2 Mid-session Mobility Management .................................................................26  
3.4.3 Multi-homing ..................................................................................................26
3.5 HIP security discussion........................................................................................................ 27
3.6 Summary.............................................................................................................................. 28

4 CHAPTER 4: MOBILE IP WITH HIP STYLE HANDSHAKING AND READDRESSING. 29
4.1 Mobile IP Handover Performance Problem....................................................................... 29
  4.1.1 Triangle Routing............................................................................................................. 30
  4.1.2 Route Optimization security problem and Return Routability ...................................... 31
4.2 Secure Mobile IP (SMIP).................................................................................................... 32
  4.2.1 Comparison between HIP and Mobile IP RO mobility support................................. 32
  4.2.2 Secure Mobile IP with HIP Style Handshaking and Readdressing............................ 33
4.3 Discussion and Analysis..................................................................................................... 35
  4.3.1 Handover performance Analysis ................................................................................... 35
4.4 Summary.............................................................................................................................. 39

5 CHAPTER 5: HIP ENABLED 3G WIRELESS NETWORKS .............................................. 41
5.1 GSM and 3GPP HIP Integration......................................................................................... 42
5.2 CDMA2000 HIP Integration............................................................................................... 46
5.3 Interconnection between WCDMA and WLAN................................................................. 46
5.4 Analysis and Discussion..................................................................................................... 49
5.5 Summary.............................................................................................................................. 51

6 CHAPTER 6: HYBRID SIP-HIP (SHIP)................................................................. 52
6.1 SIP Mobility Management and its limitation..................................................................... 52
  6.1.1 Limitation of SIP Terminal Mobility Management.......................................................... 55
  6.1.2 Hybrid Mobile IP SIP .................................................................................................... 55
6.2 Hybrid SIP-HIP (SHIP) Solution...................................................................................... 57
  6.2.1 Similarity between SIP and HIP .................................................................................... 57
  6.2.2 Hybrid SIP-HIP (SHIP) Architecture ......................................................................... 57
    SHIP scheme (HIP Based Media Session Only)................................................................. 58
    SHIP Scheme (HIP Based SIP Signalling and Media Session) ........................................ 59
    SHIP Handover mechanism.............................................................................................. 60
6.3 Discusss AND ANALYSIS ................................................................................................. 61
  6.3.1 SIP security Enhancement.............................................................................................. 61
  6.3.2 Handover Delay Analysis .............................................................................................. 61
    Binding update signalling delay analysis........................................................................... 61
    Binding Update Signalling Load ....................................................................................... 63
    Binding Update Signalling Overhead ................................................................................ 64
    Handover processing delay .............................................................................................. 64
  6.3.3 Comparison of different SHIP schemes......................................................................... 68
6.4 Limitation of SHIP.............................................................................................................. 69
6.5 Summary.............................................................................................................................. 69

7 CHAPTER 7: HIP AND MICRO-MOBILITY MANAGEMENT ..................................... 70
7.1 Macro-mobility management vs Micro-mobility management........................................ 70
7.2 Objectives of HIP-based Micro-mobility Management.................................................... 71
7.3 Existing works of HIP related micro-mobility management.............................................. 72
  7.3.1 Novaczki’s Scheme ........................................................................................................ 72
  7.3.2 Ylitalo’s Scheme............................................................................................................ 73
    Lamport One-Way Hash Chain ......................................................................................... 74
Table of Contents

Ylitalo’s scheme Overview ........................................................................................................ 75
Ylitalo’s scheme operation ........................................................................................................ 75
Analysis of Ylitalo’s scheme ..................................................................................................... 77

7.4 Improvement of Ylitalo’s Scheme – ILHC Scheme .......................................................... 78
  7.4.1 Infinite Length Hash Chain (ILHC) ............................................................................. 78
  7.4.2 New ILHC Scheme .................................................................................................... 80

7.5 Summary ............................................................................................................................ 81

8 CHAPTER 8: MICRO-HIP (MHIP) ..................................................................................... 82

8.1 mHIP Overview .................................................................................................................. 82

8.2 mHIP Agent ...................................................................................................................... 83
  8.2.1 mHIP Router ............................................................................................................. 84
  8.2.2 mHIP Gateway ......................................................................................................... 85

8.3 Signature Schemes in mHIP .............................................................................................. 85
  8.3.1 Shared Private Key Scheme ...................................................................................... 87
         Applying Share Private Key Scheme in mHIP .............................................................. 87
  8.3.2 Group Signature Scheme ........................................................................................ 87
         System Initialization Phase ......................................................................................... 88
         Partial Secret Key Generation Phase ........................................................................ 89
         Group Signature generation and verification phase ................................................... 89
         Applying Group Signature Scheme in mHIP .............................................................. 90
  8.3.3 Ring Signature Scheme ............................................................................................ 91
         Applying Ring Signature Scheme in mHIP ................................................................. 91
  8.3.4 Proxy Signature Scheme ......................................................................................... 91
  8.3.5 Summary of Signature Schemes ............................................................................... 92

8.4 Overlay Routing ............................................................................................................... 93
  8.4.1 mHIP ESP Scheme .................................................................................................. 95
         New mHIP SPINAT ..................................................................................................... 97
  8.4.2 mHIP SRTP Scheme .............................................................................................. 98
         SRTPNAT .................................................................................................................. 99
  8.4.3 Overlay Routing Table Storage ............................................................................... 99
         Distributed Hash Table ................................................................................................. 99
         Cellular IP style routing table ................................................................................... 101
  8.4.4 Summary of Overlay Routing ................................................................................. 101

8.5 mHIP Protocol Operation ............................................................................................... 102
  8.5.1 mHIP Router Registration ..................................................................................... 102
  8.5.2 MN Registration in mHIP Gateway ...................................................................... 103
  8.5.3 mHIP Idle Handover ............................................................................................ 103
  8.5.4 mHIP Connection Establishment and Tear Down .................................................. 104
  8.5.5 mHIP intra-domain handover process and multi-homing support ......................... 105

8.6 Analysis and Discussion ................................................................................................. 107
  8.6.1 Propagation Delay Analysis .................................................................................... 107
         Pre-session Handover Delay ...................................................................................... 108
         Mid-session Handover Delay ...................................................................................... 109
  8.6.2 Signalling Cost Analysis .......................................................................................... 111
         Location Update Cost ................................................................................................. 112
         Binding Update Cost .................................................................................................. 113
         Lookup Cost ............................................................................................................. 114
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Signalling Cost</td>
<td>115</td>
</tr>
<tr>
<td>8.6.3 Security Analysis</td>
<td>119</td>
</tr>
<tr>
<td>mHIP Agent Security Analysis</td>
<td>119</td>
</tr>
<tr>
<td>Impersonation attacks</td>
<td>120</td>
</tr>
<tr>
<td>8.7 Application of mHIP and future work</td>
<td>121</td>
</tr>
<tr>
<td>8.7.1 mHIP enhanced SHIP</td>
<td>121</td>
</tr>
<tr>
<td>8.7.2 mHIP integrated 3G network</td>
<td>121</td>
</tr>
<tr>
<td>8.7.3 Future Work</td>
<td>123</td>
</tr>
<tr>
<td>8.8 Summary</td>
<td>124</td>
</tr>
<tr>
<td>9 CHAPTER 9: CONCLUSIONS</td>
<td>126</td>
</tr>
<tr>
<td>9.1 Contribution</td>
<td>126</td>
</tr>
<tr>
<td>9.2 Future work</td>
<td>128</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>129</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>138</td>
</tr>
<tr>
<td>Appendix A: Publication lists</td>
<td>138</td>
</tr>
</tbody>
</table>
List of Figures

Figure 2-1 Mobile IPv4

Figure 3-1 Comparison of a Traditional Network and an HIP-based Network Architecture

Figure 3-2 HIP Architecture

Figure 3-3 a) HIT-IP mapping with DNS only
   b) HIP-IP mapping with DNS and RVS

Figure 3-4 HIP Packet Structure

Figure 3-5 HIP Base Exchange

Figure 3-6 HIP Pre-session mobility – HIP Base Exchange via RVS

Figure 4-1 Triangle Routing

Figure 4-2 Mobile IP Route Optimization

Figure 4-3 Mobile IP RR mechanism

Figure 4-4 Simplified SMIP Base Exchange initiated by CN via SMIP HA

Figure 4-5 SMIP Multi-homing Scenario

Figure 4-6 Mobile IPv6 Handover Performance Analysis in terms of RTT

Figure 4-7 SMIP Handover Performance Analysis in terms of RTT

Figure 5-1 HIP-based GPRS Network Architecture

Figure 5-2 HIP Integrated 3GPP Release 4 Distributed Network Architecture

Figure 5-3 HIP Integrated 3GPP Release 5 Multimedia Network Architecture

Figure 5-4 Basic HIP Base Exchange process in UMTS network

Figure 5-5 HIP Integrated CMDA2000 Network Architecture

Figure 5-6 Integrated UMTS/WLAN HIP-based Network Architecture

Figure 5-7 HIP-based Vertical handover in UTMS/WLAN Architecture

Figure 6-1 SIP Pre-call Mobility

Figure 6-2 SIP Mid-call Mobility
Figure 6-3 SIP Session Mobility using call transfer
Figure 6-4 SIP Personal Mobility
Figure 6-5 Simplified M-SIP Handover Message Sequence
Figure 6-6 SHIP based SIP INVITE (with SDP)
Figure 6-7 SHIP Scheme (HIP Based Media Session Only)
Figure 6-8 SHIP Scheme (HIP protected SIP Signalling and Media Session)
Figure 6-9 Handoff signalling flow
Figure 6-10 Binding Update Signalling Delay
Figure 6-11 Handover signalling overhead (H=50)
Figure 6-12 Handover delay vs delay between MN and CN
Figure 6-13 Handover delay vs. delay between MN and HA
Figure 7-1 Ylitalo’s scheme
Figure 7-2 Ylitalo’s scheme micro and macro-mobility exchanges using common packet structure
Figure 7-3 ILHC scheme micro- and macro-mobility exchanges use common packet structure.
Figure 8-1 mHIP Architecture
Figure 8-2 Group Signature mHIP micro-mobility handover packets
Figure 8-3 Modified HIP Conceptual Header for HIP Data Packet
Figure 8-4 HIP Packet Structure
Figure 8-5 SPINAT
Figure 8-6 Modified ESP header for mHIP scheme
Figure 8-7 mHIP Router Registration Process
Figure 8-8 mHIP idle handover
Figure 8-9 State Diagram of a connection in mHIP Agent
Figure 8-10 mHIP intra-domain handover with the same SA
Figure 8-11 Balanced Binary Tree Structure with three-level depth
Figure 8-12 Simulation results for average cross over distance of HIP UPDATE packet transfers in different depth levels of Balanced Binary Tree structure of different schemes
Figure 8-13 Total signalling cost vs. number of MNs with different $T_s$ value

Figure 8-14 Signalling cost improvement over HIP vs. number of MN in different schemes with different $T_R$

Figure 8-15 Total signalling cost vs. Average number of CNs per MN with different $T_s$

Figure 8-16 Signalling Cost Improvement over HIP vs. Average Number of CNs per MN in different schemes with different $T_R$

Figure 8-17 Total Signalling Cost vs. Residence Time with different $N_{MN}$

Figure 8-18 Signalling Cost Improvement in Total Signalling Cost vs. Residence Time in different schemes with different $N_{MN}$

Figure 8-19 3GPP Release 5 with mHIP enhanced

Figure 8-20 mHIP enhanced CDMA2000 architecture
List of Tables

Table 4-1  RTT required for RR pocesss
Table 6-1 Input parameters for handover
Table 6-2 Parameteres for handover processing delay
Table 6-3 Handover process delay for single protocol
Table 6-4 Handover process delay for multi-layer protocol stacks
Table 8-1 Comparison of different signature schemes
Table 8-2 Handover Performance Summary of Different Schemes
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1G</td>
<td>First Generation Cellular Network</td>
</tr>
<tr>
<td>2.5G</td>
<td>2.5 Generation Cellular Network</td>
</tr>
<tr>
<td>2G</td>
<td>Second Generation Cellular Network</td>
</tr>
<tr>
<td>3G</td>
<td>Third Generation Cellular Network</td>
</tr>
<tr>
<td>3GPP</td>
<td>3rd Generation Partnership Project</td>
</tr>
<tr>
<td>3GPP2</td>
<td>3rd Generation Partnership Project 2</td>
</tr>
<tr>
<td>4G</td>
<td>Fourth Generation Cellular Network</td>
</tr>
<tr>
<td>AC</td>
<td>Address Checking</td>
</tr>
<tr>
<td>ACR</td>
<td>Address Checking Reply</td>
</tr>
<tr>
<td>AMPS</td>
<td>Advanced Mobile Phone Service</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>API</td>
<td>Application programming interface</td>
</tr>
<tr>
<td>B3G</td>
<td>Beyond 3G network</td>
</tr>
<tr>
<td>BC</td>
<td>Binding Cost</td>
</tr>
<tr>
<td>BEET</td>
<td>Bound End-to-End Tunnel</td>
</tr>
<tr>
<td>BU</td>
<td>Binding Update</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>CDMA2000</td>
<td>Code Division Multiple Access 2000</td>
</tr>
<tr>
<td>CN</td>
<td>Corresponding Node</td>
</tr>
<tr>
<td>CoA</td>
<td>Care-of-Address</td>
</tr>
<tr>
<td>CoT</td>
<td>Care-of Test</td>
</tr>
<tr>
<td>CoTI</td>
<td>Care-of Test Init</td>
</tr>
<tr>
<td>CPU</td>
<td>Central Processing Unit</td>
</tr>
<tr>
<td>DH</td>
<td>Diffie-Hellman</td>
</tr>
<tr>
<td>DHT</td>
<td>Distributed Hash Table</td>
</tr>
<tr>
<td>DNS</td>
<td>Domain Name System</td>
</tr>
<tr>
<td>DoS</td>
<td>Denial of Service</td>
</tr>
<tr>
<td>ESP</td>
<td>Encapsulating Security Payload</td>
</tr>
<tr>
<td>FA</td>
<td>Foreign Agent</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency-Division Duplex</td>
</tr>
<tr>
<td>FQDN</td>
<td>Fully Qualified Domain Name</td>
</tr>
<tr>
<td>GHz</td>
<td>Giga Hertz</td>
</tr>
<tr>
<td>GPRS</td>
<td>General Packet Radio Service</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communications</td>
</tr>
<tr>
<td>HA</td>
<td>Home Agent</td>
</tr>
<tr>
<td>HI</td>
<td>Host Identity</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>HIP</td>
<td>Host Identity Protocol</td>
</tr>
<tr>
<td>HIT</td>
<td>Host Identity Tag</td>
</tr>
<tr>
<td>HLR</td>
<td>Home Location Register</td>
</tr>
<tr>
<td>HoT</td>
<td>Home Test</td>
</tr>
<tr>
<td>HoTI</td>
<td>Home Test Init</td>
</tr>
<tr>
<td>HSDPA</td>
<td>High-Speed Downlink Packet Access</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>IETF</td>
<td>Internet Engineering Task Force</td>
</tr>
<tr>
<td>IKE</td>
<td>Internet Key Exchange</td>
</tr>
<tr>
<td>ILHC</td>
<td>Infinite Length Hash Chain</td>
</tr>
<tr>
<td>IMS</td>
<td>IP Multimedia Core Network Subsystem</td>
</tr>
<tr>
<td>IMT-2000</td>
<td>International Telecommunication Union-2000</td>
</tr>
<tr>
<td>IMTS</td>
<td>Improved Mobile Telephone Service</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>IPSec</td>
<td>Internet Protocol Security</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>LRVS</td>
<td>Local RVS</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MAN</td>
<td>Metropolitan Area Network</td>
</tr>
<tr>
<td>MAP</td>
<td>Mobile Application Part</td>
</tr>
<tr>
<td>Mbps</td>
<td>Megabit per second</td>
</tr>
<tr>
<td>mHIP</td>
<td>micro-HIP</td>
</tr>
<tr>
<td>MHz</td>
<td>Mega Hertz</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-Input Multiple-Output</td>
</tr>
<tr>
<td>MitM</td>
<td>Man-in-the-Middle</td>
</tr>
<tr>
<td>MN</td>
<td>Mobile Node</td>
</tr>
<tr>
<td>MRP</td>
<td>Mobile Routing Point</td>
</tr>
<tr>
<td>MSC</td>
<td>Mobile Softswitch Center</td>
</tr>
<tr>
<td>M-SIP</td>
<td>Hybrid Mobile IP and SIP</td>
</tr>
<tr>
<td>NAPT</td>
<td>Network Address Port Translation</td>
</tr>
<tr>
<td>NAT</td>
<td>Network Address Translation</td>
</tr>
<tr>
<td>NGN</td>
<td>Next generation wireless communication network</td>
</tr>
<tr>
<td>NROP</td>
<td>The Nearest MRP in the Old Path</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>ORCHID</td>
<td>Overlay Routable Cryptographic Hash Identifier</td>
</tr>
<tr>
<td>OSI</td>
<td>Open System Interconnection</td>
</tr>
<tr>
<td>PAN</td>
<td>Personal Area Network</td>
</tr>
<tr>
<td>PAP</td>
<td>Previous anchor point</td>
</tr>
<tr>
<td>PDA</td>
<td>Personal Digital Assistance</td>
</tr>
<tr>
<td>PDSN</td>
<td>Public Data Serving Node</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>PSTN</td>
<td>Public Switched Telephone Network</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RA</td>
<td>Routing Area</td>
</tr>
<tr>
<td>RAN</td>
<td>Radio Access Network</td>
</tr>
<tr>
<td>RFC</td>
<td>Request for comment</td>
</tr>
<tr>
<td>RNC</td>
<td>Radio Network Controller</td>
</tr>
<tr>
<td>RNS</td>
<td>Radio Network Subsystem</td>
</tr>
<tr>
<td>RO</td>
<td>Route Optimization</td>
</tr>
<tr>
<td>RR</td>
<td>Return Routability</td>
</tr>
<tr>
<td>RTT</td>
<td>Round Trip Time</td>
</tr>
<tr>
<td>RVS</td>
<td>Rendezvous Server</td>
</tr>
<tr>
<td>SA</td>
<td>Security Association</td>
</tr>
<tr>
<td>SDP</td>
<td>Session Description Protocol</td>
</tr>
<tr>
<td>SEET</td>
<td>Stripped End-to-End Tunnel</td>
</tr>
<tr>
<td>SGSN</td>
<td>Serving GPRS Support Node</td>
</tr>
<tr>
<td>SHIP</td>
<td>Hybrid SIP-HIP</td>
</tr>
<tr>
<td>SIMA</td>
<td>Simultaneous Multi-Access</td>
</tr>
<tr>
<td>SIP</td>
<td>Session Initial Protocol</td>
</tr>
<tr>
<td>SMIP</td>
<td>Secure Mobile IP</td>
</tr>
<tr>
<td>SPI</td>
<td>Security Parameter Index</td>
</tr>
<tr>
<td>SRNS</td>
<td>Service Radio Network Subsystem</td>
</tr>
<tr>
<td>SRTP</td>
<td>Secure Real-Time Transport Protocol</td>
</tr>
<tr>
<td>SS7</td>
<td>Signalling System 7</td>
</tr>
<tr>
<td>SSRC</td>
<td>Synchronization Source</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>TD-CDMA</td>
<td>Time Division – CDMA</td>
</tr>
<tr>
<td>TDD</td>
<td>Time-Division Duplex</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time Division Multiple Access</td>
</tr>
<tr>
<td>TD-SCDMA</td>
<td>Time Division-Synchronous Code Division Multiple</td>
</tr>
<tr>
<td>ULA</td>
<td>User agent</td>
</tr>
<tr>
<td>UDP</td>
<td>User Diagram Protocol</td>
</tr>
<tr>
<td>ULI</td>
<td>Upper Layer Identifier</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunications System</td>
</tr>
<tr>
<td>U-NII</td>
<td>Unlicensed National Information Infrastructure</td>
</tr>
<tr>
<td>UWB</td>
<td>Ultra-Wideband</td>
</tr>
<tr>
<td>VLR</td>
<td>Visitor Location Register</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over IP</td>
</tr>
<tr>
<td>WAN</td>
<td>Wide Area Network</td>
</tr>
<tr>
<td>W-CDMA</td>
<td>Wideband Code Division Multiple Access</td>
</tr>
<tr>
<td>WG</td>
<td>Working Group</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>WiMAX</td>
<td>Worldwide Interoperability for Microwave Access</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless LAN</td>
</tr>
<tr>
<td>xDSL</td>
<td>Digital Subscriber Line</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 BACKGROUND

Wireless networking is evolving from the traditional homogeneous wireless network into a heterogeneous wireless network. Mobility management only in the Physical Layer and Data Link Layer are insufficient for a heterogeneous wireless network, and mobility management in the Network Layer or above are required to provide a seamless handover.

The IP mobility problem is due to IP address overloading. In the early design of TCP/IP network architecture, all computers were bulky, and they were impossible to move unless powered off; there was no "mobility" concept at that time. In order to simplify the network architecture design, the IP address is used not only in the Network Layer to locate the node, but also used to identify the nodes in the Transport Layer or above.

Nowadays, computers and other devices, such as mobile phones and personal digital assistances (PDA), come with different wireless network interfaces; they can move from one network into the other network from time to time, even switch between different wireless technologies. IP address overloading is a serious problem in wireless networks.

As a mobile node (MN) moves from one sub-network into the other, the IP address is changed, which means that not only is the network locator in the Network Layer changed, but also the endpoint identity in the Transport Layer or above has been changed. The MN needs to notify its Corresponding Nodes (CN) about the change of IP address in order to provide seamless communication. In practice, all the layers, except the Physical Layer and Data Link Layer, need to update their IP information after the handover of MN.

Different schemes in different layers are proposed to solve the IP mobility problem in the OSI model. Mobile IP [1, 2] is one of the most widely discussed and developed schemes. It uses a Home Address to hide the actual change of the IP address. CNs establish connections with the MN based on the MN’s Home Address. There is a server, called Home Agent (HA) in the MN’s home network. The MN’s HA will forward MN’s packets to the current location of the MN.
when the MN is roaming into a foreign network. Although this mechanism provides the IP mobility, there is a problem in the packet routing performance. Different extensions are proposed in order to solve this problem for different scenarios.

The Host Identity Protocol (HIP) [3, 4] is a new protocol to solve IP mobility from a different angle. HIP re-models the IP network architecture by introducing a new namespace and new layer, Host Identify (HI) and Host Identity Layer. Instead of IP address, HI takes the role of an endpoint identity in the network architecture. IP address is used only in the Network Layer as the network locator, while HI is used in the Transport Layer or above as the endpoint identity. If a MN moves into a new network, its IP address is changed, but the HI remains the same. HIP can provide IP mobility just by updating the network locator, which means that only the Network Layer needs to be updated.

1.2 THESIS GOALS & SIGNIFICANCE

HIP is a new draft macro-mobility management protocol. Although the handover performance of HIP is better than that of Mobile IP [5], there are many uncertainties in deploying HIP into current wireless networks and, therefore, many research opportunities.

The major purpose of this dissertation is to improve the security and performance of the handover process in a wireless IP network. Handover performance of HIP is better than that of Mobile IP. It is impossible to change from the traditional IP-based mobility network (Mobile IP based network) to a public key-based IP network (HIP-based network) overnight. In this dissertation, a new step-stone evolution process is proposed. This step-stone solution allows the network to benefit from some of the advantages of a public key-based network, such as security and performance improvement, while the IP address is still being used as an endpoint identity. The solutions proposed require much less change to the operating system kernel of endpoints than does a full HIP implementation.

Beside the kernel modification, HIP network components also need to be integrated into the wireless network to provide a better performance. New IP-based enhanced wireless communication network architectures are proposed in this dissertation. We have proposed the
HIP-based Universal Mobile Telecommunications System (UMTS) network architectures for different UMTS Releases and also the HIP-based CDMA2000 network architecture. This helps the 3G networks to handle most HIP signalling internally, which can minimize the signalling packets to external networks. Beside the homogeneous wireless network architectures, we have also proposed a new HIP-based heterogeneous wireless architecture for the handover between different wireless technologies. An HIP-based vertical handover scheme is also demonstrated.

Moreover, beside HIP-based network components integration, the interworking between Session Initial Protocol (SIP) and HIP is also important. SIP is the major signalling protocol in the 3G network. HIP and SIP can work independently to provide a reliable service, but some processes are overlapped, which are a waste of the network resource and introduce unnecessary handover process and handover delay. A new HIP and Session Initial Protocol (SIP) interconnection model is proposed. This model improves the performance of SIP in the HIP-based IP network.

Micro-mobility management is the other important, but often neglected, issue in the future heterogeneous wireless network. Subscribers may use more than one wireless access technology to access the communication network. Even though the wireless interfaces are different, subscribers are still under the same domain. Currently, most of the micro-mobility management protocols are Mobile-IP based [6-8]. Without Mobile IP as macro-mobility management protocol, those Mobile-IP based micro-mobility management protocols cannot function properly. At this stage, there is no complete solution to HIP-based micro-mobility management. In this dissertation, a novel HIP-based micro-mobility management protocol, micro-HIP (mHIP), is proposed. mHIP introduces new solutions to micro-mobility management. All the mHIP enhanced middle boxes are treated as one virtual server. All the middle boxes can generate a valid signature on behalf of the domain. MNs are conceptually connected to this virtual server directly and MN’s packets are routed internally inside this virtual server and forwarded to MNs. mHIP provides a good performance and load-balancing solution to HIP-based micro-mobility management.
Fours conference papers and two journal papers related to this dissertation have been accepted and published; they are listed in the following:

**Journal**


**Conference**


1.3 **THESIS OUTLINE**

The rest of this thesis is organized as follows:

- Chapter 2: Mobility Management Overview
  - An overview of mobility management
  - Review of different major wireless technologies
    - 3G,
    - IEEE802.11
Review of different mobility protocols

- Mobile IP and
- SIP
- Micro-mobility management

Chapter 3: Host Identity Protocol (HIP)

- Overview of HIP
- Introduction of HIP architectures and HIP network components
- HIP Base Exchange
- HIP Mobility and Multi-homing support
- Security of HIP

Chapter 4: Mobile IP with HIP Style Handshaking and Readdressing

- Discussion of the Mobile IP problem
- Propose a new step-stone solution – Secure Mobile IP with HIP style handshaking and readdressing (SMIP)
- Analysis and discussion of SMIP

Chapter 5: HIP Enabled 3G Wireless Networks

- Propose a new HIP-based GPRS network architecture
- Propose a new HIP-based Release 4 and Release 5 network architecture
- Propose a new CDMA2000 network architecture
- Propose a new HIP-based UMTS-WLAN interconnection architecture and HIP-based vertical handover process.
- Discussion and analysis of the HIP-based vertical handover process

Chapter 6: Hybrid SIP-HIP (SHIP)

- Overview of SIP mobility management and its problems
- Propose Hybrid SIP-HIP (SHIP) model and its architecture
- Discussion and analysis of SHIP handover performance

Chapter 7: HIP and Micro-Mobility Management

- Micro-mobility overview
• Setting up the objectives of HIP-based micro-mobility management
• Review two existing studies which are related to HIP-based mobility management
• Discussion of the limitation of these existing studies
• Propose an improvement scheme for one of the existing schemes

- Chapter 8: Micro-HIP (mHIP)
  • Overview of mHIP
  • mHIP network architecture
    - mHIP Agent
      • mHIP Gateway
      • mHIP Router
  • Review of different signature scheme can be applied in mHIP network
    - Implementation in mHIP Agent
  • Overlay routing method and overlay routing table
  • mHIP operations
  • Discussion and analysis of mHIP
    - Propagation Delay Analysis
    - Signalling Cost Analysis
    - Security
  • Application of mHIP

• Chapter 9: Conclusions
Chapter 2: Mobility Management Overview

Mobility Management is an important issue for wireless IP networks. In traditional wired computer networks, all nodes are bulky and unable to move until they power off. In order to simplify this design, IP addresses play dual roles in the architecture. An IP address is used as a host identifier in the Transport Layer or above in the OSI model, while it is used as the network locator in the Network Layer. The limitations of this design are obvious, as wireless technology has advanced.

In the wireless IP network, a Mobile Node (MN) may attach to different access points (APs) from time to time, which means that the IP address may change when it roams into a new subnet. In the homogenous wireless IP network, it is still possible to handle mobility management in the Physical Layer and Data Link Layer only, in order to re-direct communication to the updated location of the MN. However, this solution is insufficient if there is more than one wireless technology used in the network (e.g. heterogeneous wireless networks). It is impossible to handle overall mobility management by means of the Physical Layer and Data Link Layer only; the Network layer or above need to be involved in order to provide mobility management.

In this chapter, different wireless technologies are reviewed. We introduce the major protocols of mobility management protocols in use today.

2.1 WIRELESS NETWORK HISTORY

The early development of the wireless communication network focused only on voice communication. The Improved Mobile Telephone System (IMTS) was the first commercially available wireless telephone system in 1946[9]. However, the performance of IMTS was unsatisfactory. Operators had to erect a very tall IMTS transmitter tower near the centre of a metropolitan area. There were several channels assigned to the transmitter tower. Mobile users could use one of those channels to complete a call but, because there were not enough channels
available, the demand was always greater than the supply. In addition, more power was required for the transmitter, or receiver, if it was located in a larger metropolitan area. This became problematic when a city was developing around its periphery. In order to solve this problem, the concept of a cellular network was introduced.

This first generation cellular network, Advanced Mobile Phone Service (AMPS), has been available since the early 1980s in North America [9, 10]. The AMPS was based on analogue communication, that is, all the voice traffic channels were analogue. The Second Generation Cellular Network (2G), such as the Global System for Mobile communications (GSM: originally from Groupe Spécial Mobile) [11, 12], was the next system to be introduced. 2G uses digital transmission technology; all communication channels are digitalised. It is also equipped with error detection and correction, so the quality of voice communication is a dramatic improvement on the first generation cellular network. Moreover, each frequency channel can be shared by many connections using different technologies, such as Code Division Multiple Access (CDMA) and Time Division Multiple Access (TDMA) [10, 13]. Data communication in the cellular network was later added in the 2G network and was called ‘2.5G’, an example of which is the General Packet Radio Service (GPRS) [14]. This enabled users to have data connections on top of the 2G network. However, the data rate was very low.

In order to overcome data rate limitation, third generation cellular network (3G) was introduced. The aim of 3G is to provide multimedia services and increase data rates. IP architecture is a new feature in 3G system and this network targets universal roaming. Compared to the 2G system, there are fewer air interface standards. There are currently only two key players in the market; Wideband CDMA (WCDMA) from 3GPP [13, 15] and CDMA2000 from 3GPP2 [10]. There is also the Time Division Synchronous Code Division Multiple Access (TD-SCDMA) model, developed in China; however, it is not yet commercially available. In the telecommunication network evolution, the network has changed from a circuit-switched network (1G) into a mix of circuit and packet-switched networks (3G provides packet and circuit-switched networks) [16]. In future standards, such as Beyond 3G (B3G) and 4G systems, the architecture will be IP-based networks [10, 16, 17]. No single technology will serve as standard.
Instead, it will be a combination of different wireless technologies. Mobile devices will have more than one network interface to connect to the communication network.

In addition to telecommunication organisations working on the wireless standards, the data network is being developed simultaneously in different areas, from the Personal Area Network (PAN) to the Wide Area Network (WAN). The IEEE 802 Working Group (WG) is also working on wireless technologies for computer networks, such as IEEE 802.11 [18] and IEEE 802.16 [19] standards.

### 2.2 WIRELESS TECHNOLOGY

#### 2.2.1 3G standard

3G standard is one of the most important wireless technologies in mobile communication. Compared to other wireless technologies, 3G technologies are the only technologies that have been commercialised for high speed movement communication. 3G standards are under the aegis of the IMT-2000 program of International Telecommunication Union (ITU), which aims to provide improved voice capacity, high-speed Internet and multimedia services. IMT-2000 is intended to provide these services “anytime, anywhere” to a worldwide market. ITU also demands that 3G standards support a data transfer rate of at least 144Kbps in outdoor environments and 2Mbps in indoor environments [10].

### 3GPP Overview

The 3GPP was established in December 1998 as a collaborative agreement between groups of telecommunication associations who work together on 3G specifications. They focus on evaluation of the GSM network, one of the most popular 2G systems in the world [10, 13, 20-22]. Members of this group are the European Telecommunications Standards Institute, the Association of Radio Industries and Businesses/Telecommunication Technology Committee, the China Communications Standards Association, the Alliance for Telecommunications Industry Solutions and Telecommunications Technology Association.

UMTS is the 3G standard proposed by 3GPP. There are different versions of the proposed UMTS architecture, which are called Releases. UMTS contains two modes of
operation, WCDMA for Frequency-Division Duplex (FDD) and Time Division-Code Division Multiple Access (TD-CDMA) for Time-Division Duplex (TDD). Different frequencies spectrums are used for transmission in uplinks and downlinks in FDD, while the same frequency but multiplexed in time, is used for uplinks and downlinks in TDD [10, 13, 20]. WCDMA is a pure CDMA-based technology. It requires two 5MHz frequency bands for operation in FDD mode. TD-CDMA combines TDMA and CDMA technology. Aside from the 3G radio system, which can provide a faster data transfer rate, UMTS also changes in the core network. Instead of the Signalling System 7 (SS7) network, IP networks take on a more important role in the code network. The core network of UMTS is IP-based [10, 13, 16, 20, 23, 24]. Soft switching and voice gateway are added into the UMTS core network. In addition, IMS is also introduced to provide Session Initial Protocol (SIP) based service in the UMTS network [25].

Handovers can be categorised in two ways in the 3GPP UMTS solution: soft handovers and hard handovers. There are two variants of soft handover – ‘soft’ handover and ‘softer’ handover [10, 26]. Soft handover uses the make-before-break strategy. Soft handovers can be the handovers between different base stations, but managed by the same Radio Network Subsystem (RNS) (inter Node B/intra RNS), between different base stations and RNSs, but connect to same Serving GPRS Support Node (SGSN) (inter Node B/inter RNS/intra SGSN) and between different based stations, RNSs and SGSNs (inter Node B/inter RNS/inter SGSN) [10, 13, 20]. In the UMTS architecture, an SGSN manages one or more Radio Network Controller (RNC), while an RNC manages many base stations, which are also co-called ‘Node Bs’. In different scenarios of soft handovers, the corresponding components databases are updated. In an inter Node B/inter RNS scenario, the Service Radio Network Subsystem (SRNS) is relocated upon soft handover completion, while in the inter SGSN scenarios, the SGSN is relocated. Softer handover is a special case of soft handover. The handover is conducted between the cells which are served by the same base stations. This means that the MN performs the handover between different antennas (e.g., directional antennas) of the same base station [10, 22, 26]. About 5-15% of handovers are softer handovers and 20-40% are soft handovers [26]. Soft handovers are not supported in High-Speed Downlink Packet Access (HSDPA) [10]. Hard handovers use break-
before-make strategies to perform the handover between cells. A Hard handover is similar to handovers in the 2G network but without the change of frequency. The old radio link is removed before adding a new radio link. This occurs when the MN is moving between cells that are using different carrier frequencies or when the base stations are connected to different RNCs when there is no Iur interface between the RNCs.

**3GPP2 Overview**

3GPP2 is an alternative to the 3G solution. The members of 3GPP2 include the Telecommunications Industry Association, the Association of Radio Industries and Business, the Telecommunication Technology Committee, the Telecommunication Technology Association and the China Wireless Telecommunications Standards Group. The 3GPP2 solution is an evolution of the Qualcomm CDMA solution (which is cdmaOne or IS-95) [10, 13, 15]. Unlike the 3GPP solution, 3GPP2 has created a new architecture for 2G and 3G services. The evolution from IS-41 to an all IP network in the architecture is more direct. IS-41 is a standard similar to a Mobile Application Part (MAP) in the GSM network, which is used for identifying and authenticating users, and routing the calls on mobile phone networks [10, 13, 15, 27]. 3GPP2 reuses accepted IETF protocols where possible. CDMA2000 is the radio interface used in the 3GPP2 solution. There are two main versions of CDMA2000: CDMA2000 1x and CDMA2000 3x, however, CDMA2000 3x is no longer in development [10]. CDMA2000 1x operates on 1.25 MHz carriers. There are different variants in CDMA2000 1x. CDMA2000 1xRTT makes the CDMA network with IP packet support and is similar to GPRS in the GSM network. The CDMA 1xRTT is considered a 2.5G technology. CDMA2000 1xEVDO, which is also known as IS-856 or CDMA2000 HRPD, separated the data and voice into different channels to provide a data rate of 2.4Mbps on the downlink. Data and voice traffic are integrated on the same 1.25 MHz carrier. The data speed of CDMA2000 1xEVDO can be up to 4.8Mbps. Soft-switch and voice gateways are added into the CDMA 1xEVDO network [10, 13, 15].

Similar to the 3GPP UMTS solution, 3GPP2 CDMA2000 solutions also support hard handovers, soft handovers and softer handovers in 1xRTT. However, there is no soft handover in the EVDO network [10, 15]. In general, an MN runs the mobility management procedure in idle
mode, while assisting the Radio Access Network (RAN) to perform handovers in active mode. In
the idle state, it is the MN’s responsibility to update the network components. It registers a new
Mobile Softswitch Center (MSC) or Public Data Serving Node (PDSN). The MN updates the
circuit switch network in order to receive the circuit voice calls. Hard handovers and soft
handovers are processed for mobility between base transmission stations. A8/A9 and A10/A11
interfaces support the mobility when the MN is moving around the Base station controller and
packet control function [10, 13, 15]. Mobile IP supports CDMA2000 when the MN is switching
between a PDSN/Foreign Agent (FA) under the same Home Agent [10].

2.2.2 IEEE 802 Wireless Standard

IEEE 802 has defined many standards that are related to Local Area Networks (LANs)
and Metropolitan Area Networks (MANs); for example, IEEE 802.3 is Ethernet and IEEE 802.5
is Token Ring. Compared to ITU, the IEEE 802 working group is focusing on computer data
communication, mainly the lowest two layers in the OSI network model. In the following section,
different wireless standards of IEEE 802 are reviewed.

IEEE 802.11

IEEE 802.11 WG is developing a wireless LAN network standard for data
communication. The scope of IEEE 802.11 [18, 28] is “to develop a Medium Access Control
(MAC) and Physical Layer (PHY) specification for wireless connectivity for fixed, portable and
moving stations within a local area.” [29]. It works on two different frequency spectrums:
2.4GHz and 5GHz band [13, 18, 22, 28, 30-33]. The IEEE 802.11 standard [18, 28] is also
known as “Wi-Fi”, which was a reference to IEEE 802.11b [32] in the past. There are numerous
amendments to IEEE 802.11. Originally, it was defined only to support 1 Mbps transmission
rates (mandatory) and 2Mbps (optional) and used 2.4GHz radio frequency for transmission
media [28]. In the amended IEEE 802.11a [31], 5GHz is supported in the IEEE 802.11 standard.
IEEE 802.11 supports the infrastructure mode and ad-hoc mode. The former is similar to the
traditional network model, while the latter is a “peer-to-peer” operation. With the succession of
IEEE 802.11b, IEEE 802.11 becomes the dominant player in the wireless LAN.
IEEE 802.11 does not specify the implementation of the handover process in the most widely used amendments, IEEE 802.11b/g [32, 33]. Manufacturers of IEEE 802.11b/g hardware implement a suitable handover mechanism. The most common mechanism currently being used consists of two phases, the discovery phase and the re-authentication phase [34]. Recommend Practice IEEE 802.11F [35], which was redrawn in 2006, discussed the interoperability between multi-vendor access points, but only for data use. An amendment to IEEE 802.11r [36] proposed a mechanism for fast roaming for mobile Voice over IP (VoIP) in the IEEE 802.11 environment, but it is still in draft form and not expected to be published until 2008.

**IEEE 802.16**

Although the IEEE 802.11 standard is successful in industry, many companies use a hot spot to provide commercial wireless connection. IEEE 802.11 can provide only a wireless LAN connection; users still depend on other technologies to connect to external networks, such as Digital subscriber line (xDSL). IEEE 802.16 [19, 37] is a standard for wireless broadband connection access. The official name of IEEE 802.16 is WirlessMAN, but it is also known as “WiMAX”.

IEEE 802.16 provides wireless broadband internet access by using 10-66GHz [19, 30]. With the amendment IEEE 802.16a, wireless broadband internet access can also be provided by using 2-11GHz [38]. The original aim of IEEE 802.16 was to only provide only the fixed wireless broadband internet access However, the new amendment, IEEE 802.16e [39], enhances IEEE 802.16 to provide mobility.

The IEEE 802.16 handover progress is divided into two parts, the Handover Pre-registration Phase and the Real Handover Phase [40]. The target base station is selected and pre-registered with the MN in the Handover Pre-Registration phase. The connection to the current base station is maintained. In the Real Handover Phase, the connection between the MN and the serving base station is broken; MN re-associates with the target base station after this.

**IEEE 802.20**

The aim of IEEE 802.20 [41] is similar to IEEE 802.16e in that it targets the mobility communication. The target customers of IEEE 802.20 are similar to those of 3G. The draft
standards proposed claim that it can provide mobility service up to a speed of 250km/hr. It operates in licensed bands which are below 3.5GHz [10, 30, 41].

2.3 MOBILITY MANAGEMENT IN IP NETWORK

Mobility management of wireless technologies is well defined only in the Physical Layer and Data Link Layer. The network components in the wireless network forward packets to the new location after the MN has attached to a new AP. However, in heterogeneous wireless networks such as, e.g. 4G, the MN may handover vertically between different wireless technologies. So, conducting handover signalling in the Physical Layer and Data Link Layer is not sufficient. Handover progress in the Network Layer or above is needed for vertical handover.

The fundamental problem of IP mobility is the overloading of IP addresses; i.e., an IP address identifies the device’s location on the network topology in the Network Layer and identifies the host in the Transport Layer or above. If the MN changes the IP address after handover, the IP address is changed. This means that not only is the location of the network changed, but so also is the host identity of the MN. The MN needs to notify its CNs to update its IP address, which means that not only does the network locator need to be updated, but also the endpoint identity needs to be changed. From the Corresponding Node’s (CN) point of view, it communicates with a new peer. This design creates a mobility management problem in an IP network.

In the remainder of this section, different protocols, which handle IP mobility problems, from IETF are reviewed.

2.3.1 IETF Mobility Solution

IETF is an organisation that develops and promotes standards for the protocols used on the Internet. They have different solutions for mobility management issues in IP networks.

Mobile IP

Mobile IP [1, 2] is one of the most popular solutions for IP network mobility. It is a Network Layer solution. There are two versions of Mobile IP: Mobile IPv4 [1] and Mobile IPv6 [2], for IPv4 [42] and IPv6 [43] networks representatively. Mobile IP keeps any changes to the
upper layer model in TCP/IP architecture to a minimum. Mobile IP still uses the IP address as the end-point identifier; however, it tries to hide the location of the MN from the CN in order to solve the dual roles problem of IP addresses. Some important components of Mobile IP network include:

- **Home Network**: A network that assigns a Home Address to the MN.
- **Home Address**: IP address assigned to an MN in the Home Network. The IP address will not change when the MN is roaming.
- **Foreign networks**: Any networks other than the Home Network.
- **Home Agent (HA)**: The router on an MN’s Home Network. This router keeps the record of the MN and redirects packets of the MN to its foreign network when the MN is roaming in foreign networks.
- **Foreign Agent (FA)**: The router on the MN’s foreign network, which receives packets from the HA and forwards to the MN. It exists only in Mobile IPv4.
- **Care-of-Address (CoA)**: The IP address that is assigned to the MN (Mobile IPv6) or the FA (Mobile IPv4). An HA is based on an MN’s CoA to forward its packets.

![Figure 2-1 Mobile IPv4](image)

Every MN gets a Home Address from its Home Network. The CN uses an MN’s Home Address instead of its current IP address to establish the connection with the MN. In the Mobile IPv4 scenario, when the communication packets arrive, the HA located in the MN’s Home Network forwards these packets to MN directly (if MN is still in the Home Network.). When the MN roams into a foreign network, as depicted in Figure 2-1, it registers itself in the FA in the foreign network and uses the IP address of the FA to update its CoA record in HA. HA forwards
communication packets to the MN via the FA when the MN roams in a foreign network. However, this structure leads to a performance problem in Mobile IP. The drawback of this structure is that no matter how close the MN and CN are, all packets need to be routed via HA, and therefore the performance is downgraded. This problem is called “Triangle Routing”. Even though different extensions are proposed to improve the routing performance of Mobile IP, no extension is clearly better than others [44].

In Mobile IPv6, the FA is removed from the architecture. The MN uses its current IP address in the foreign network as a CoA. In order to solve the triangle routing problem, Mobile IP with Route Optimisation Extension, an extension of Mobile IPv4 [45], is part of the Mobile IPv6 standard. In the Mobile IP with Route Optimization, after handover, an MN not only updates the CoA in HA, but also updates the cache in the CN. The MN sends a binding update packet to the CN to notify the change of IP address. The CN updates its cache and instead sends the packet to the MN’s new CoA.

In heterogeneous wireless network mobility management studies, Mobile IP is assumed for use in upper layer mobility management [46, 47]. Compared to other mobility management protocols, Mobile IP is the most widely adopted model to date.

**Session Initial Protocol (SIP)**

SIP [48] is an application protocol. The idea of SIP is to provide a framework for multi-media connection. SIP does not define anything relating to the under layer connection. It is used only to create or tear down the multi-media sessions. SIP is used as a signal protocol for VoIP service. 3GPP and 3GPP2 also selected SIP instead of Mobile IP as the signalling protocol for the multi-media connections in the 3G service [25, 49].

Although SIP was not originally designed to handle mobility management in the IP network, a mobile user agent (UA) can use re-INVITE to perform application layer mobility management [50, 51]. Users and computer nodes are identified by the SIP URI in the SIP environment. When a caller wants to create a multi-media session with a mobile UA which is roaming in the foreign network, the SIP Registrar and Redirect Server notifies the CN about the update location of the mobile UA or redirects the request to the mobile UA. After the multi-
media session is established, the mobile UA can use re-INVITE to notify its CN about the change of IP address, in order to request the CN to send the multi-media stream to the new location so that handover can be performed [50-52].

Unlike Mobile IP, the mobility management of SIP is centred on the user mobility. The mobility management of SIP is focused on how to reach the SIP user, which can register more than one device in the network. As SIP is an application layer protocol, it is in the lowest priority protocol in the TCP/IP network structure – it introduces a longer delay during the handover when the network is busy. Additionally, SIP is a plain text protocol; it introduces a larger overhead signalling. Furthermore, SIP does not provide overall mobility management. It can provide only mobility management for the SIP-based session, which is UDP-based. For those sessions which are not created by SIP, e.g. ongoing TCP, non-SIP based UDP connections will be lost [50-52]. For details of the limitations of SIP, please refer to Chapter 6.1. In order to provide overall mobility management, SIP has to co-operate with other protocols [53, 54].

2.3.2 Micro-mobility Management

Micro-mobility management is a niche topic in mobility management. Mobile IP and SIP, which have been discussed previously, are “macro-mobility management” protocols. The macro-mobility management protocols handle mobility in any scenario, e.g. handover across different domains [55]. In order to simplify the design, the handover signalling packet is normally handled by the CN in most of the macro-mobility protocols. After the MN has roamed into a new subnet, it notifies its CN about the change of IP address. The CN updates its caches and redirects the communication packets to the new location of the MN [2, 50, 56, 57]. However, this strategy is not efficient for intra-domain handovers. In a common scenario, most of the MNs perform only intra-domain handover, which is the handover between different APs within the same domain. The path from the gateway of the domain to the CN remains unchanged. Adopting macro-mobility management for intra-domain handover is therefore not efficient.

Micro-mobility management targets intra-domain handover [55, 57, 58]. The intra-domain handover signalling is handled within the domain. As the CN or other external network components are not notified for the intra-domain handover, the signalling is handled by other
network components inside the domain such as gateways or routers. Middle boxes in the domain need to be enhanced by micro-mobility management protocols, such as Cellular IP Gateway and Cellular Node of Cellular IP [6], Mobile Anchor Point of Hierarchical Mobile IPv6 [8], and HAWAII Gateway of HAWAII [7, 59]. Micro-mobility management protocols provide a fast handover with lower latency for the intra-domain handover. Additionally, they provide the location privacy of the MN, as the signalling is bounded only within the domain. Micro-mobility management is achieved by the tunnel-based (Hierarchical Mobile IPv6) or routing-based (Cellular IP and HAWAII) models to forward packets to the MN’s updated location [55].

Currently, most of the micro-mobility protocols are Mobile IP based [6-8]. They must use Mobile IP as the macro-mobility management protocol and cannot operate without Mobile IP.

2.4 SUMMARY

This chapter has reviewed some of the dominant models in the wireless technologies used in practice. Heterogeneous wireless technology is one of the characteristics of the 4G network and no single wireless technology will be used as standard. The mobility management in the Physical Layer and Data Link Layer of OSI model is not adequate. The Network Layer or other upper layers need to be involved in mobility management in order to provide seamless handovers. Currently, Mobile IP and SIP are the major players in the macro-mobility management protocols; the former provides an all-round solution for the host and the latter provides mobility management of the multi-media services for the user. Most of the well-known micro-mobility management protocols are Mobile IP based. In the next chapter, Host Identity Protocol (HIP) is introduced. It solves the IP mobility management problem from the other angle of Mobile IP.
This chapter will discuss the key concepts of Host Identity Protocol (HIP) [3, 4] and how to solve IP mobility management problems from an alternate angle to Mobile IP.

Mobility management is one of the most important issues in mobile networks. The fundamental problem of IP mobility is the overloading of IP addresses as discussed in Chapter 2. In the current IP network architecture, when a host is moving into another network, its IP address will change. Mobile IP solves the problem by hiding the new IP address and using a home address and Home Agent (HA) for communication with other hosts. This hides the actual movement of the Mobile Node (MN). However, this structure leads to performance problems in Mobile IP which is covered in Chapter 2[44].

Host Identity Protocol (HIP) [3, 4] is designed to address a number of architectural and security problems in the current TCP/IP architecture. HIP makes IP-based mobility much easier [56, 60]. In order to solve the fundamental problem of IP mobility, HIP is not going to “hide” the change of IP address. HIP re-structures the TCP/IP network architecture by introducing a new namespace and a new layer.

![Figure 3-1 Comparison of a Traditional Network and an HIP-based Network Architecture](image-url)
3.1 HIP OVERVIEW

HIP [3, 4] is a newly drafted secure mobility management protocol. It aims to handle IP mobility and security by using a new approach [56, 60]. Unlike Mobile IP, it solves the IP mobility problem fundamentally. It re-models the current TCP/IP model by introducing a new namespace, Host Identity (HI), and a new layer, Host Identity Layer, which is seen as a 3.5 layer, i.e. a layer between the Network Layer and the Transport Layer in the OSI model. HI is used instead of an IP address as the end host identifier. The concept of HI is similar to SIP URI, which is used to identify the host in the SIP application. The handover performance of HIP is proving to be better than Mobile IPv6 [5]. Furthermore, HIP also supports multi-homing [56, 61], which allows the host to communicate with more than one interface. Multi-homing is not supported in Mobile IP. HIP will be discussed in detail in the rest of this chapter.

3.2 HIP ARCHITECTURE

The dual roles of an IP address become more problematic with the increase of mobility and multi-homing. A new namespace and layer, HI and Host Identity Layer, are proposed in HIP to solve the IP address roles problem [3, 4]. The Host Identity Layer, a 3.5 layer in the OSI model, is introduced in between the Network Layer and the Transport Layer, which is depicted in Figure 3-1. The IP address is only a network topological locator in the HIP network architecture. HI replaces the IP address as the endpoint identifier. HI is a new cryptographic public key namespace which is added to current TCP/IP stacks. Each node has at least one HI, which is either public or anonymous [3]. As HI is the cryptographic public key, meaning the lengths of different HIs vary, it is not practical to apply HI directly. In order to adopt HIP in the current IPv6 application programming interface, a 128-bit long Host Identity Tag (HIT), which contains 28 bits for the Overlay Routable Cryptographic Hash Identifier (ORCHID) [62] and 100 bits for the hash of HI, is used to represent HI in practice [3].
In this new network structure, Transport Layer sockets are named with HIs/HITs instead of IP addresses. HI binds to at least one IP address. This simultaneous binding relationship enables HIP to support mobility and multi-homing [56, 60]. The bindings between HIs and IP addresses are not exactly defined in the HIP architecture. The mapping can be done by a Domain Name System (DNS) [63] or HIP Rendezvous Server (RVS) [64]. The DNS stores the mapping between the Fully Qualified Domain Name (FQDN) and IP addresses in the current Internet model. The DNS does not store the direct mapping between HI/HIT and IP addresses, but does so between FQDN and HI/HIT. When a host looks up an FQDN, the DNS replies with the IP address and HIT [63]. The operation of HIP DNS is depicted in Figure 3-3a.

3.2.1 **Rendezvous Server (RVS)**

As mentioned previously, DNS can provide the mapping between HI/HIT and IP addresses. However, using DNS is not an efficient solution. A DNS server stores only the mapping of FQDN to HIT and FQDN to an IP address. Furthermore, the time required to update all of the DNS on Internet is lengthy. Based on what most ISPs suggest to their customers, it requires 48-72 hours to update all DNS records on the Internet. This is not an acceptable solution in mobility management.
In order to solve this problem, RVS [64] is introduced. The role of RVS is similar to that of the HA in Mobile IP. The DNS no longer holds the mapping between the FQDN and the IP address of the host; it carries the mapping between the FQDN and the host’s RVS IP address. Direct mapping between HI/HIT and the IP address of the host is stored in the RVS. A host registers itself in its RVS and updates its record in the DNS to update the mapping to the FQDN and IP of RVS [64]. The CN establishes the HIP-based connection via the RVS, as depicted in Figure 3-3b. Pre-session mobility management can be achieved by the RVS (for detailed information, please refer to Section 3.4.1.).

### 3.2.2 HIP Packet

HIP is a framework for secure connections. In theory, any secure communication schemes can be applied to HIP. There are two types of HIP packets: HIP signalling packets and HIP data packets. There is a HIP header following the IP header in the HIP signalling packet. HIP contains the HITs of the source and receiver. Other HIP parameters are included in the HIP signalling packets, based on different scenarios. There are eight basic HIP signalling packets, which are:

- **I1 – the HIP Initiator packet**
- R1 – the HIP Responder packet
- I2 – the second HIP Initiator packet
- R2 – the second HIP Responder packet
- UPDATE – the HIP Update packet
- NOTIFY – the HIP Notify packet
- CLOSE – the HIP Association Closing packet
- CLOSE_ACK – the HIP Closing Acknowledgment packet

These eight signalling packets are used to set up, dismantle or update the HIP connection. The first four packets (I1, R1, I2, and R2) are packets for the HIP Base Exchange, which is the process for setting up the HIP connection. The CLOSE and CLOSE_ACK are used to dismantle the connection. The UPDATE packet is used to update the security association information or performs the handover.

![Figure 3-4 HIP Packet Structure](image)

In theory, there is an HIP header in between the IP header and the upper layer header (such as the Internet Protocol Security (IPSec) header and the TCP/UDP header) for the HIP data packet (Figure 3-4b). However, in order to minimize any change to the current data packet structure, the HIP header is removed in practice – the HIP header exists only conceptually. HIP daemon uses the security scheme header, such as an IPSec Encapsulating Security Payload (ESP) [65] header, to identify the HIP-based connection (Figure 3-4c). Although any security scheme can run on top of HIP, to date, only the HIP ESP scheme is well defined [66]. The application of a Secure Real-Time Transport Protocol (SRTP) [67] to HIP is being discussed [68]. The HIP ESP scheme is the focus of the rest of this dissertation.
3.3 HIP BASE EXCHANGE

HIP Base Exchange, depicted in Figure 3-5, is a four-way handshake process for the establishment of HIP connections between hosts. It is the host-to-host authentication protocol. The hosts, which are involved in the handshake process, are called Initiator and Responders.

- **I1** is the first packet from an Initiator to a Responder. It is a trigger packet, which contains the HIT of the Initiator and the HIT of the Responder, if known.

- **R1** is the second packet in the Base Exchange and it is from the Responder to the Initiator. R1 starts the actual exchange. It contains a cryptographic challenge, which is called a ‘puzzle’ [3]. The Initiator must solve the puzzle before continuing the Base Exchange. The puzzle is difficult to solve but easy to verify, so the Initiator needs to expend some computation power to solve it. This puzzle makes the Base Exchange resistant to Denial of Service (DoS) attacks. The Responder does not need to create any state, and processing is minimal because the I1 can come from any host in the Internet. If state is created in I1, it may create opportunities for various DoS attacks[3]. Beside the puzzle, R1 also contains Diffie-Hellman parameters and a signature.

- **I2** is the third packet in the process and it is sent to the Responder by the Initiator with the solution to the puzzle. Some Central Processing Unit (CPU) time is used to solve this puzzle. However, the Responder can easily verify the accuracy of the solution. I2 is discarded by the Responder if the solution is incorrect. I2 also contains the Diffie-Hellman parameter signed by the Initiator. The Initiator also encrypts and sends its public key (HI) in this packet.

- **R2** is the final packet in this handshake process. Once the Responder has verified the I2 successfully, it will use the Diffie-Hellman parameters to create the session and decrypt the I2 public key. R2 contains the Responder signature, which allows the Initiator to authenticate the Diffie-Hellman session key with the identity of the Responder.
This HIP Base Exchange can be considered as a lighter weight version of IPSec’s Internet Key Exchange (IKE) [60]. In the HIP ESP scheme, a pair of IPSec Security Associations (SA) is created during a HIP Base Exchange; the SPI numbers are exchanged in the I2 and R2. The HIP ESP connection is established after the HIP Base Exchange is complete. The data packet structure of the HIP ESP scheme is no different from the normal IPSec ESP data packet.

![Figure 3-5 HIP Base Exchange](3)

#### 3.4 HIP MOBILITY MANAGEMENT AND MULTITHOMING

HIP is designed to decouple the host identifier and network locator for the current IP network architecture. In theory, HI/HIT maps with IP addresses automatically. In the IPSec ESP scheme, SAs are no longer bounded IP addresses, but are HIs/HITs. An MN can receive packets that are protected by an SA from any IP addresses [56, 60]. It enables a host to change its IP address and continue to communicate with its peers. The following section discusses pre-session mobility management and mid-session mobility management of HIP.

#### 3.4.1 Pre-session Mobility Management

An MN updates its information in its RVS via the normal HIP Registration process [69] after it has booted up. When a CN intends to establish an HIP-based connection with the MN, it requests the MN’s information from the DNS. The DNS replies to the CN with MN’s RVS IP address and then starts the HIP-based four-way handshake. The CN only knows the IP address of the RVS. It sends the I1 to the MN’s RVS with the MN’s HIT as destination HIT if known. The RVS forwards this I1 to the MN. The MN replies with the R1 to the CN directly; the rest of the
HIP Base Exchange is processed between the MN and the CN without the use of any other network server. This process is depicted in Figure 3-6.

When an MN roams into a new subnet, it updates its information in the RVS. When other peers intend to start a new HIP connection with the CN, the RVS server forwards the I1 packet to the updated location of the MN. Pre-session mobility can be achieved by using RVS via this mechanism.

3.4.2 Mid-session Mobility Management

As SAs are no longer bounded to IP addresses, the host can receive the ESP-protected packet from any IP addresses. If an MN changes its IP address during communication with other peers, besides the pre-session mobility management process mentioned above, the MN also sends an HIP UPDATE packet with a LOCATOR parameter to notify its peers of the change of IP address [56]. The LOCATOR parameter contains the new IP address and the SPI associated with this connection. The whole handover process is protected by ESP, which prevents third party bomb attacks. During the handover process, an MN may decide to rekey the SAs. It can notify the CN of the new SPI parameter in the handover packet. Regardless of whether the CN re-keys the SA pairs, it will always process the address, checking to verify that the new address is reachable [56].

3.4.3 Multi-homing

There are two types of multi-homing: host multi-homing and site multi-homing [56]. A host has more than one interface or global address that is called host multi-homing; while a host
has an interface that has multiple globally reachable IP addresses and is called ‘site multi-homing’. HIP is considered a candidate to enable multi-homing in IPv6 networks [61].

When a multi-homing node communicates with other hosts, it uses the LOCATOR parameter to notify its peers of the additional interface. Furthermore, the internet-draft recommends using asymmetric numbers of SAs between hosts in multi-homing situations (e.g. each interface of a multi-homing node has an inbound SA, while the peer has one interface with one inbound SA only) [56]. The multi-homing host will tell its peer which interface is the preferred address for communication. Simultaneous multi-access support is currently being researched in relation to HIP [70]; however the load balancing issue is outside the scope of the development of HIP.

3.5 HIP SECURITY DISCUSSION

HIP is designed to improve the security level of the current IP architecture. HIP provides the authentication of hosts in communication. In addition, HIP also limits the exposure of the host to Denial of Service (DoS) and Man-in-the-Middle (MitM) attacks [3].

HI is a cryptographic public key namespace. As HIT is the hash of HI, peers can check if the HI is valid by running a hash function on it. If the host keeps its private key secret and does not disclose it to any third parties, according to the Public / Private Key security scheme, the peer can verify whether a third party has modified the packet to prevent a MitM attack. Furthermore, any proper public key authentication protocols can be used to check whether a party holds the private key corresponding to a public key and entitle it to the name. Therefore, HIP is a self-certificated protocol that does not require any external certification.

The puzzle in the HIP Base Exchange increases the cost of the Initiator in order to protect the Responder from a number of DoS attacks [3]. The Initiator uses some CPU time to solve the puzzle, while the validation of the puzzle requires only a fairly cheap calculation power. Moreover, the Responder does not create any state until it has received a valid I2; it drops any invalid I2 packets. Furthermore, the Diffie-Hellman exchange in R1, I2, and R2 also allows the creation of a session key to encrypt the Initiator public key for privacy protection. After the HIP
Base Exchange, the HIP connection is established. HIP connections are required in the upper layer secure channel. Through the ESP connection, the communication channel is encrypted and protected by the IPSec algorithm.

### 3.6 SUMMARY

In this chapter, an overview of the HIP has been presented together with a discussion of the way that HIP solves the IP mobility management problem from an alternate angle to Mobile IP. HIP offers a new framework that provides a secure environment with mobility and multi-homing support for data communication. HI/HIT is introduced as a namespace in the new network architecture. A 3.5 layer, Host Identity Layer is added in between the IP Layer and Transport Layer in the OSI model. The HIP ESP scheme is currently a well defined scheme for HIP connections. As the Security Associations (SAs) created by the HIP Base Exchange are not bounded to IP addresses, but to HI/HIT, so the MN can receive the HIP ESP data via any IP address. When the MN roams into a new network, it needs only to notify the CN about its updated IP address. As the SAs are bounded to the HI/HIT, the secure connection can be maintained.

The coming chapters will discuss our research achievements in HIP-related mobility management. We will introduce a step-stone solution for the evolution from a Mobile IP based network to HIP-based network.
Chapter 4: Mobile IP with HIP Style Handshaking and Readdressing

IPv6 [43] was proposed in answer to the global shortage of IPv4 [42] addresses around 1992. Although IPv6 has many enhanced features, it has yet to replace IPv4. In many networks, especially local area networks, IPv4 is still the major player. Mobile IPv6 was proposed later in 1998. It is designed to handle the issues related to Mobile IP nodes. HIP architecture is in a similar situation. HIP is a protocol for a public key-based IP network. Apart from providing a more secure environment for the IP network; its mobility management outperforms that of Mobile IP [5]. However, some modifications are required for the current TCP/IP network to support HIP (such as API), and the operation system kernel needs to be updated, especially those IPv4 API. It is not surprising that it will take a longer time for the HIP to be adopted into current network architecture. In this chapter, we proposed a new stepping-stone solution – Secure Mobile IP (SMIP), a HIP style handover Mobile IP solution that can be used in current Mobile IP architecture. SMIP uses the HIP handover mechanism in Mobile IP to improve the handover performance for Mobile IP networks and make the handover independent from the Home Agent (HA). It also provides a secure environment and multi-homing support for the current Mobile IP network. It will be shown that the Round Trip Time (RTT) and Binding Cost (BC) of the new proposed SMIP in handover is less than that of Mobile IP. SMIP can be considered as the first step in Mobile IP network’s evolution to a public key-based network.

4.1 MOBILE IP HANDOVER PERFORMANCE PROBLEM

Mobile IP [1, 2] is a well known mobility management protocol in the network layer; it enhances mobility support for the standard TCP/IP network by hiding the change of IP address. The Mobile Node (MN) gets an IP address, which is called ‘Home Address’, from its Home Network to represent itself globally. HA, a router in the Home Network, tracks the MN’s
movement and re-directs packets to the MN’s current location. When an MN roams into the foreign network, it gets a new IP address in the foreign network. This new IP address is called ‘Care of Address’ (CoA). In Mobile IPv4, there is a router in the foreign network called Foreign Agent (FA). The MN will update its CoA record in the HA by the FA’s IP address. The HA creates a tunnel to the FA and forwards the MN’s packets via the FA to the MN’s CoA. There is no FA in the Mobile IPv6 environment. The MN updates its record in the HA by its CoA and the HA forwards the MN’s packet to the MN directly.

4.1.1 Triangle Routing

Mobile IP provides terminal mobility. However, it does not have good packet routing performance. When the MN roams in foreign networks, no matter how close the CN and MN are, all packets to the MN need first to be sent to its HA and then forwarded to the MN. This mechanism has poor packet routing performance and is labelled ‘Triangle Routing’ which is depicted in Figure 4-1

![Figure 4-1 Triangle Routing](image1)

In order to solve this problem, different extensions of Mobile IP are proposed. Route Optimization (RO) [45] extension is one of them and is widely adopted in Mobile IP. RO is an optional scheme in Mobile IPv4, but it has become part of the standard in Mobile IPv6 [2]. This extension provides better performance by avoiding triangle routing. As shown in Figure 4-2, the MN updates its CoA record in the HA and sends a binding update (BU) packet to the CN to notify the change of IP address. Instead of sending packets to the MN via the HA, the CN sends

![Figure 4-2 Mobile IP Route Optimization](image2)
them directly to the MN’s new CoA. This improves the packet routing after handover, but it creates a security problem [71].

4.1.2 Route Optimization security problem and Return Routability

In the Mobile IP RO extension, the MN sends a BU packet to the CN to notify the change of IP address. The triangle routing is avoided. However, attackers can use spoofed BU packets to corrupt the CN’s binding cache and cause packets to be delivered to an incorrect address [71]. Attackers can use this action to launch denial-of-service (DoS) attacks on the CN or the MN, or cause a third party node to receive unexpected packets. The attackers put an invalid or third party’s IP address in the BU packet and send it to the CN on behalf of the MN. Upon receipt of this fake BU packet, the CN re-directs the communication stream to a third party or an invalid IP. The communication between the CN and the MN is broken and the third party receives many unexpected packets.

Apart from DoS attacks, hackers can also use the same method to ‘steal’ a session. The hacker ‘steals’ the address of an MN by sending a spoofed BU packet with their own current IP address as the new CoA; so the hacker pretends to be an MN and continues the communication with the CN. Hackers may also send two-directional spoofed BU packets to two communicating nodes in order to eavesdrop on their communication. This is known as a Man-in-Middle (MinM) attack.

Figure 4-3 Mobile IP RR mechanism
To deal with the attacks mentioned above, an MN needs to be verified before the BU process. Return Routability (RR) [72], which is depicted in Figure 4-3, is a mechanism to verify the MN. In the basic RR mechanism, four processes, Home Test Init (HoTI), Care-of Test Init (CoTI), Home Test (HoT) and Care-of Test (CoT) need to be undertaken before generating a BU packet. The MN sends the HoTI via the HA and CoTI directly to the CN simultaneously. The CN generates a nonce every two minutes based on the, $K_{cen}$, key, which is generated when the CN boots up. The CN uses the nonce to create two tokens, sending one token in HoT to the Home Address and one in CoT directly to the MN’s CoA. The HA forwards the HoT to the MN inside an IPSec-ESP protected tunnel. The MN uses both tokens to create a $K_{cen}$ key to generate a BU packet which is then sent to the CN. Since the CN has all the information for creating the key, it can reproduce the key and authenticate the BU packet. The lifetime of the state created at the CN for the BU is restricted to a few minutes in order to reduce the threat of the time shifting attack[72].

The Mobile IP RO extension improves the packet routing performance. The RR process improves the security of the BU process; it degrades the handover performance and introduces an additional delay for the handover process. The remainder of this chapter is devoted to the introduction of a Secure Mobile IP (SMIP), which uses an HIP style handover mechanism to improve the performance.

4.2 SECURE MOBILE IP (SMIP)

4.2.1 Comparison between HIP and Mobile IP RO mobility support

The mobility support of Mobile IP RO is similar to that of HIP. Both of them use an ‘agent’ to redirect the initial packet and use an update message to notify the CN directly about the MN’s current IP address. The concepts of Mobile IP RO and HIP handover mechanisms are the same. The difference lies in the detail of their implementation. In Mobile IP RO, the CN initiates a connection by sending packets to the MN’s home address. The MN’s HA in its home network forwards these packets to the MN’s CoA. After the MN has received them, it sends a BU packet to the CN to notify the current IP address and repeats this process upon handover to notify the
change of IP address. Similarly, in HIP architecture, RVS is used to re-direct the first packet in the HIP Base Exchange to establish the connection. The roles of Mobile IP HA and HIP RVS are similar [73]. When the CN sends the I1 to the MN, the I1 is passed to the MN via the MN’s RVS. The MN communicates with the CN directly afterward. Additionally, when the IP is changed, the MN will use the UPDATE packet to notify the change of IP address.

4.2.2 Secure Mobile IP with HIP Style Handshaking and Readdressing

The RR process combats the security threat of Mobile IP RO but it degrades the handover performance. An experiment conducted by Joekla et al. [5] found that HIP outperforms Mobile IPv6 by 69% in vertical handovers. Although handover performance of HIP is better than that of Mobile IP, it is impossible for an IP-based network (Mobile IP) to be adapted to a public key-based IP network (HIP) in the short term. Taking this into consideration, we propose a combination of the HIP handover mechanism and Mobile IP. It is called Secure Mobile IP (SMIP), a stepping-stone solution in the evolution from IP to HIP.

In SMIP, IP addresses are still used as host identifiers. The home address is generalised as an Upper Layer Identifier (ULI); this is a permanent address of the MN in the network. The ULI is used to identify the host in the upper layer. However, the routing paths between the MN and the CN are based on the current MN’s IP address, which is mapped to the ULI. The binding update mechanism is similar to HIP in that, to date, it is defined only in ESP.

As mentioned previously, the roles of the HA in Mobile IP and the RVS in HIP are similar. They provide the mapping between the ULI (Home Address/HIT) and the current IP of the MN. In SMIP, HA and RVS are integrated to become SMIP HA. This enables the network to process traditional Mobile IP, SMIP and HIP packets at the same time. Thus, it provides the stepping-stone service from traditional IP networks to the public key-based IP network.

In order to maintain the security level for the binding update process, a ‘downgraded’ HIP Base Exchange, called SMIP Base Exchange, between the MN and the CN is added. The operation is depicted in Figure 4-4. An Initiator (CN) wants to establish a connection with an MN. The CN sends the I1 packet with its current IP address and its ULI to the MN based on its ULI. This I1 packet is sent to the MN via SMIP HA. SMIP HA looks up the current location of the
MN and then forwards the I1 packet to the MN. The MN replies to the Initiator with R1, which includes a Diffie-Hellman (DH) value, a public key of the MN and a signature. A puzzle, which is used in HIP Base Exchange to protect the host from DoS attack, is optional. The Initiator replies by I2 packet, which includes the DH values, its public key, its signature and the solution of the puzzle if needed. The MN acknowledges with its signature via the R2 packet. This handshaking process can be treated as a downgraded version of IKE exchange. In the ESP mode, the only mode that is well defined so far, Security Parameter Indexes (SPIs) - are exchanged in I2 and R2 in SMIP Base Exchange. An ESP protected connection is created after SMIP Base Exchange is complete.

The handover mechanism of SMIP is similar to HIP. An SAs pair is created for communication in SMIP; this pair verifies the host. The MN gets a new IP address when it roams into a new subnet. The process of using SMIP before sending BU packets is inherited from Mobile IP. However, SMIP uses HIP style handovers, whereas Mobile IP RR processes with binding updates. The MN sends an update packet with the corresponding SPI, the new locator and the sequence number to its CN. The MN also needs to sign this packet. If the MN wants to perform SA re-keying, it should also put the D-H value in this packet. The CN requests an address check to test if the new address is reachable. This action is similar to CoT in the RR process. After the MN has replied to the address check, the handover process is complete.

Multi-homing is another feature of SMIP, but it is not currently supported in Mobile IP. As shown in Figure 4-5, when an MN wants to notify a CN about a new interface, it sends an UPDATE packet to the CN (this is similar to the handover process). However, it contains
different parameters. This UPDATE packet contains a New SPI (NES), a LOCATOR, a SEQ and a DH. Both the new SPI and the old SPI in the NES parameter are the same, meaning the new SPI does not replace the old SPI. This is a multi-homing case, instead of a handover with SPI re-keying. The remainder of the process is same as the handover process.

![Figure 4-5 SMIP Multi-homing Scenario](image)

### 4.3 DISCUSSION AND ANALYSIS

SMIP can be considered as a temporary solution for the IP network in its upgrading to future secure public key-based architecture. An SMIP Base Exchange is necessary for creating an SAs pair before a connection is established. Compared to Mobile IP, its security level is higher. Secure connection is optional only in Mobile IP, but is compulsory in SMIP. This requires a higher computation power in order for hosts to perform the encryption and decryption on SMIP-based communication. Moreover, SMIP supports multi-homing, which is not supported in Mobile IP. In the following section, the handover performances of Mobile IPv6 and SMIP are compared.

#### 4.3.1 Handover performance Analysis

The analysis in this section is based on the Round Trip Time (RTT) and Binding Cost (BC). RTT is defined as the elapsed time from transmitting data over a closed path. Let $RTT_{A,B}$ represent the RTT between A and B. In order to simplify the analysis, it is assumed that the time required from A to B is the same as B to A.

In Mobile IPv6, the RR process is needed to obtain tokens from the CN so that the MN can generate a valid binding update. There are four different sub-processes in RR, HoTI, CoTI, HoT and CoT. We group HoTI and HoT into one combined sub-process, HT. CoTI and CoT are
also grouped into one combined sub-process, CT. Table 4-1 shows the RTT of different processes in RR.

Table 4-1  RTT required for RR processes

<table>
<thead>
<tr>
<th>Process</th>
<th>RTT required</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT</td>
<td>RTT_{MN,HA} + RTT_{HA,CN}</td>
</tr>
<tr>
<td>CT</td>
<td>RTT_{MN,CN}</td>
</tr>
<tr>
<td>RR Total</td>
<td>Max {(RTT_{MN,HA} + RTT_{HA,CN}), RTT_{MN,CN}}</td>
</tr>
</tbody>
</table>

Figure 4-6 Mobile IPv6 Handover Performance Analysis in terms of RTT

Figure 4-6 depicts the Mobile IPv6 handover process. The MN sends the HoTI and the CoTI packets to the CN simultaneously. These two processes run at the same time. The CN generates a home nonce and care-of nonce after it has received the HoTI and CoTI respectively, and sends these to the MN via HoT and CoT. The MN needs to receive HoT and CoT before generating the binding update packet, so the overall RTT delay of the RR process is dependent on the longer process of HT or CT, which is max \{(RTT_{MN,HA} + RTT_{HA,CN}), RTT_{MN,CN}\}.
The binding update is sent from the MN to the CN and then an acknowledgement is sent from the CN back to the MN. The RTT delay of the binding update process is $RTT_{MN,CN}$. The overall delay of the handover in Mobile IP is $\max\{(RTT_{MN,HA}+RTT_{HA,CN}),RTT_{MN,CN}\} + RTT_{MN,CN}$.

The handover process in SMIP is simplified, as depicted in Figure 4-7. Handover is protected by SAs and the signature. As discussed previously, the handover delay in SMIP is only from the binding update and address checking process. The overall delay is $1.5 RTT_{MN,CN}$. It is at least $0.5 RTT_{MN,CN}$ shorter than the Mobile IP handover process. The improvement is obvious.

![Figure 4-7 SMIP Handover Performance Analysis in terms of RTT](image)

Next we consider the Binding Cost (BC) which is defined as the cost of handover handling. BC includes the binding packet transmission and the binding computation conducted at the nodes. Compared with RTT, BC is considered from the overall network point of view. All processes in the handover are considered - even those processes that do not introduce any handover delay to the MN. There are two main costs in BC: processing costs and transmission costs. Processing cost is based on the cost of computation required at the node, while transmission cost is based on the cost of transmission in the network. Before proceeding with a detailed discussion, some notions are defined as follows:

- Let $BC_X$ be the total binding cost for scheme $X$,
- Let $PBC_Y$ be the binding cost incurred in process $Y$,
- Let $CP_{i,A}$ be the processing cost for process $i$ at node $A$, and
- Let $CT_{i,A,B}$ be the binding packet transmission cost in process $i$ between node $A$ and $B$. 
The BC of Mobile IP is the sum of the cost of the RR process and the cost of the Binding Update. We also group the sub-processes of RR into two classes: HT and CT. The MN sends the HoTI to the CN via the HA and waits for the reply. The BC of HT is:

$$PBC_{HT} = CT_{HoTI,HAMN} + CP_{HoTI,HA} + CT_{HoTI,HCN} + CP_{HoTI,CN} + CT_{HoT,HA,CN} + CP_{HoT,HA} + CT_{HoT,HA,MAN}$$

(4.1)

As the HA only forwards the packets to the MN and CN, so $CP_{HoTI,HA}$ is equal to $CP_{HoT,HA}$. Similarly, the transmission costs of HoTI and HoT packets are equal, so the formula can be simplified as follows:

$$PBC_{HT} = 2\left(CT_{HT,HA,MN} + CT_{HT,HA,CN}\right) + 2CP_{HT,HA} + CP_{HoTI,CN}$$

(4.2)

At the same time that HoTI is sent out, the MN sends a CoTI to CN directly. The BC of CT is:

$$PBC_{CT} = CT_{CoTI,MN,CN} + CP_{CoTI,CN} + CT_{CoT,MAN,CN}$$

(4.3)

Similar to the HT process, the cost of CoTI packet transmission between MN and CN is the same as the cost of CoT. Therefore, the cost of CT can be simplified as follows:

$$PBC_{CT} = 2CT_{CT,MCN,CN} + CP_{CoTI,CN}$$

(4.4)

The total cost of RR can be summarised as the sum of $PBC_{HT}$ and $PBC_{CT}$. The cost of the home token and care-of token generation in CN is similar, so $CP_{HoTLCN}$ = $CP_{CoTLCN}$ = $CP_{RR,CN}$.

$$PBC_{RR} = PBC_{HT} + PBC_{CT}$$

$$= 2\left(CT_{HT,HA,MN} + CT_{HT,HA,CN}\right) + 2CP_{HT,HA} + CP_{HoTLCN} + 2CT_{CT,MCN,CN} + 2CP_{CoTLCN}$$

$$= 2\left(CT_{HT,HA,MN} + CT_{HT,HA,CN} + CT_{CT,MCN,CN}\right) + 2\left(CP_{HT,HA} + CP_{RR,CN}\right)$$

(4.5)

The cost of the Binding Update process is equal to the cost of generating the BU packet by the home token and care-of token in the MN. The MN sends the BU packet to the CN; the CN checks the validity of the packet and replies to the MN. The BC of this process is:

$$PBC_{BU} = CT_{BU,MCN,CN} + CP_{BU,MCN} + CT_{BU,CN,MCN} + CP_{BU,CN}$$

$$\therefore CT_{BU,MCN,CN} = CT_{BU,CN,MCN}$$

(4.6)

$$\therefore PBC_{BU} = 2CT_{BU,MCN,CN} + CP_{BU,MCN} + CP_{BU,CN}$$

The BC of the Mobile IPv6 handover process is the sum of $PBC_{RR}$ and $PBC_{BU}$, that is:
The BC of SMIP is less complex than that of Mobile IP. The MN sends the UPDATE packet with the LOCATOR parameter to the CN. The CN requests the address checking and the MN replies to it; the SMIP handover process is then complete. Since all processes are based on SA, each node processes only the packet and replies with correct parameters. The BC of SMIP is given below:

\[
BC_{SMIP} = 2CP_{SMIP,CN} + CP_{SMIP,MN} + 3CT_{SMIP,MN,CN}
\]  

(4.8)

It is clearly shown that SMIP requires less BC than Mobile IP. Furthermore, in the circumstance of frequent handover, the difference is significant. To avoid an eavesdropping attack and time shifting attack in RR, the key and state have a limited lifetime of a few minutes. Binding updates for an MN that frequently changes its IP address have a higher BC. SMIP relies on SAs; nodes are not required to do any extra computation when an MN is moving from one sub network to another until it requires the re-keying of the SAs pair. Clearly, SMIP requires less processing in binding updates.

In Mobile IP RR, HoTI and HoT are processed via HA. This shows the handover process if there are many MNs requesting a handover simultaneously. The independence of the HA in SMIP leads to its shorter handover delay and lower binding cost.

In short, from both RTT and BC points of view, SMIP outperforms Mobile IP. Furthermore, SMIP handover is independent of other network components for the handover process.

### 4.4 SUMMARY

In order to move from the current IP network architecture into future public key-based IP network architecture, a new stepping-stone solution SMIP is proposed. SMIP provides a secure connection while retaining the use of an IP address as an endpoint identifier. The handover
mechanism in SMIP is similar to that of HIP. It has been shown that SMIP provides a shorter RTT and a lower Binding Cost (BC) in the handover process. Moreover, handover of SMIP is independent from a Home Agent (HA). There is no need to modify the upper layer protocols in SMIP. It can still benefit from the advanced features of the public key-based IP network, such as the strong security mechanism, good handover handling and support of multi-homing. To conclude, SMIP can be considered as an initial step in the migration from Mobile-IP-based networks to future HIP-based IP networks.
Chapter 5: HIP Enabled 3G Wireless Networks

The previous chapter introduced a new step-stone solution for the evolution from Mobile IP based networks to HIP-based networks. Mobile IP adopted ideas from HIP but it is still using Home Address as the endpoint identity. This is only a partial solution and a full implementation of HIP would be more effective for mobility management. This chapter introduces novel network architectures that avoid problems that may plague an HIP implementation.

In this chapter, we consider the deployment of HIP in a 3G wireless network. Host Identity Protocol (HIP) is directed at future public key-based IP networks. In theory, only mobile nodes (MNs) need to update their kernels in order to support HIP, while the other network components can remain unchanged. However, with MN upgrading only, the pre-session handover delay in wireless communication networks is not satisfactory. If HIP RVS is integrated into the wireless network, it can reduce the signalling to external networks, so that the pre-session mobility management can be improved. In this chapter, solutions for HIP support in the 3G wireless networks are proposed. Modifications of various 3GPP and 3GPP2 network architectures in order to support HIP are discussed. With these new architectures, pre-session mobility management is shown to be improved. Furthermore, the existing policy and management mechanism can be applied to HIP connections directly. We also present an integrated HIP-based heterogeneous wireless network. By applying the HIP multi-homing feature, the vertical handover delay in heterogeneous wireless networks can be reduced to a minimum.
5.1 GSM AND 3GPP HIP INTEGRATION

The network architectures of 3GPP UMTS of different releases are slightly different [23, 24, 74]. Generally, circuit-switched signalling and packet-switched signalling are handled separately. We will focus only on the packet-switched signalling and data in this section.

General Packet Radio Service (GPRS) [14] is introduced to GSM networks support packet-switched data. GPRS provides the packet-switched data channel on the air interfaces and packet-switched transport network, which is separate from the standard GSM circuit-switched transport network. GPRS introduces several special additions to GSM networks. Gateway GPRS Support Node (GGSN) and Serving GPRS Support Node (SGSN) are the most important additions to the GPRS network [10, 13, 20].

The functionality of SGSN is similar to the Mobile Switching Centre (MSC) / Visitor Location Register (VLR) in the circuit-switched domain. An SGSN connects to several base stations; the service area of an SGSN is divided into routing areas (RAs). When an MN is moving between RAs under the same SGSN, the SGSN will manage the path update. Moreover, the SGSN also interfaces with a Home Location Register (HLR). An HLR contains subscription information related to the number of subscribers.

The GGSN is an interface to the external packet-switched network (such as the internet). It acts as a gateway for the GPRS network and connects to the SGSN. In the GPRS network, it can also have a signalling interface with the HLR. In the Mobile IP solution of the interconnection model between UMTS and other networks, the Mobile IP FA connects to the GGSN in order to forward the Mobile IP packet to the mobile phone [10, 13, 22].

The UMTS solution also inherits its structure from GPRS, with enhancements and modifications [23, 24, 74]. In the most simplified model, the base stations are connected to the SGSN, while the SGSN connects to the GGSN, which provides a gateway to the other UMTS packet data networks or Internet [10, 20, 23, 24].

In order to enhance the 3GPP UMTS solution to support HIP, an HIP RVS needs to be added into the UMTS network. The functionality of HIP RVS is similar to HLR and HA/FA in Mobile IP [73]. The RVS provides the early state mapping between the HI/HIT to the current IP.
address of the MN. From the external network point of view, the RVS should be integrated in the
GGSN or directly interface with the GGSN. When an external node tries to establish an HIP-
based connection with a mobile phone, the GGSN will be the first network node in the UMTS
network to receive the I1. The GGSN can co-operate with the RVS to map HI/HITs with their IP
addresses.

Besides the GGSN, an SGSN also has a communication interface with the RVS. Mobile
phone users always communicate between each other. If both mobile users are under the same
UMTS, the SGSN will communicate with the RVS directly instead of via a GGSN to process the
I1 packet. Figure 5-1 to Figure 5-3 show three different HIP integrated 3GPP network
architectures.

An RVS in the UMTS needs to be modified to meet the requirement of the UMTS. As
the functionality of an RVS is similar to that if an HLR or a VLR in the UMTS network, the
HI/HIT also needs to map to the mobile subscriber. The RVS in UMTS not only stores the
mapping to the IP address, but also the identity of the subscriber of the SIM/USIM.

The Initiator asks the DNS to get the IP address of the RVS. If the UMTS network is not
enhanced by HIP RVS, the RVS may locate this anywhere in the Internet. The I1, which is sent
by the Initiator, goes to the RVS and is then forwarded to the UMTS network. In the HIP
enhanced UMTS network, which is shown in Figure 5-4, the Initiator sends the I1 to the RVS,
which is located in the UMTS network. When the GGSN receives the I1 packet, it looks up the
mapping in the RVS and then forwards it directly to the mobile phone, as per the normal process
of HIP.

This HIP-enhanced UMTS network can reduce the delay in establishing the HIP
connection. When the MN moves around in the UMTS network, it needs to send only the
UPDATE packet to the RVS. All the pre-session mobility signalling does not need to send to the
external network. This can eliminate the external factors which affect the pre-session mobility
handover process.
Figure 5-1 HIP-based GPRS Network Architecture

Figure 5-2 HIP Integrated 3GPP Release 4 Distributed Network Architecture
Figure 5-3 HIP Integrated 3GPP Release 5 Multimedia Network Architecture

Figure 5-4 Basic HIP Base Exchange process in UMTS network
5.2 CDMA2000 HIP INTEGRATION

Compared with a UMTS network, a CDMA2000 network reuses a lot of existing IETF protocols in the core packet-switched network [10]. The Packet Data Serving Node (PDSN) supports both simple IP and Mobile IP packet services natively. Mobile IP HA exists in the CDMA core network. As previously discussed, the functionality of HIP RVS is similar to Mobile IP HA, so the RVS server can be placed in the same location as the HA to make CDMA2000 HIP supported, which is depicted in Figure 5-5.

5.3 INTERCONNECTION BETWEEN WCDMA AND WLAN

All future wireless communication networks will be IP-based networks. Heterogeneous wireless platforms will coexist in next generation wireless communication networks (NGN). UMTS and IEEE 802.11 are the most popular wireless networks today. UMTS was developed
from the widely deployed GSM networks. The successful IEEE 802.11 has dominated in the Wireless LAN (WLAN) market. UMTS and IEEE 802.11 are providing different types of wireless service to users. UMTS provides a high mobility environment with a lower data rate transmission, while IEEE 802.11 provides a high data rate transmission with less mobility support. It is clear that they – or their successor technologies – will be part of the NGN.

As the NGN is a heterogeneous wireless network, the handover is more complex. It is no longer under the same wireless interface. The overall handover process cannot be completed by the Physical Layer and Data Link Layer alone, but requires intervention from the Network Layer or above. In other words, vertical handover is needed.

The HIP-enabled UMTS architecture proposed in the previous section has been considered as an integration of HIP and various wireless technologies. There are two categories of the interconnection between UTMS and WLAN [22, 46, 47, 75]: tight coupling and loose coupling. In the tight coupling scenario, a WLAN gateway is connected with a SGSN. WLAN can reuse the authentication, mobility management and billing infrastructures of UMTS directly. Loose coupling interconnects the UMTS and WLAN via the Internet. The handover in tight coupling is similar to a micro-mobility management, the route of packets from the CN to the network by using the same path during handover, while the handover in loose coupling is similar to the macro-mobility management. Most of the existing work is based on the assumption that Mobile IP is the upper layer mobility management protocol. In this section, we focus only on the tight coupling model. The integrated HIP-enabled UMTS/WLAN network architecture is shown in Figure 5-6.

An MN has at least two interfaces for the vertical handover in this UMTS and WLAN integrated network. The multi-homing feature supported in HIP [56, 60, 61] can be used to reduce the vertical handover delay.
When an MN is moving from one network to another, the handover is conducted based on the information of Layer 1 (Physical Layer) and Layer 2 (Data Link Layer), i.e. the handover decision is based on signal strength. In the situation where the MN moves from UMTS to WLAN, it gets a new IP address after it is attached to the wireless access point (AP). The MN can use the LOCATOR parameter in the UPDATE packet to notify the CN about the additional interface information. The MN can also set that WLAN interface to be the primary address at the same time in order to reduce the handover process. As the HIP also supports simultaneous multi-access (SIMA) [70], the MN uses a SIMA_FLOW_BINDING parameter in the UPDATE message for SIMA, so does the connection between CN and the UMTS interface remain during the handover process. The MN is still able to receive the communication stream from the CN. After the new communication connection between the CN and MN’s WLAN interface has been established, the MN can break the UMTS connection with the CN if necessary. This strategy is called “Make-before-Break”. The operation of the handover process is shown in Figure 5-7.
The handover from WLAN to UMTS is similar to that of UMTS to WLAN. When the MN starts the handover from WLAN to UMTS, it informs the CN of the changing IP address via the UPDATE packet. When the handover is completed, the MN will disconnect from the wireless access point. Generally, during a HIP-based vertical handover, the MN will use the multi-homing and SIMA feature of HIP to notify the CN to duplicate the data stream and send to both network interfaces. After the new data stream has been established, the MN can drop the old data stream. This can shorten the handover delay. If two data streams are not allowed, a similar process will be carried out excluding SIMA_FLOW_BINDING.

5.4 ANALYSIS AND DISCUSSION

In the proposed HIP-enabled 3GPP and 3GPP2 network architectures, the HIP RVS is integrated into core 3G networks. The HIP pre-session mobility signalling is processed within 3G networks. As a mobile subscriber’s RVS is in the same local network, the HIP pre-session mobility signalling packets can be handled internally. As RVS is integrated into the core network,
management and billings subsystems in 3G networks still work in the HIP-enabled structure without major modification.

Handover performance is one of the main factors in assessing a wireless communication network. Handover delay is defined as the time between switching data streams from one network interface to another. If the MN is physically moving between subnets that share the same wireless technology, the general handover processing time is defined as:

$$\text{Handover}_{\text{process}} = T_t + T_p$$  \hspace{1cm} (5.1)

where $T_t$ is the sum of transmission durations of all handover control packets and $T_p$ is the processing time including packet buffering.

In proposed HIP enabled 3G networks, the $T_t$ is represented as

$$T_t = \sum_{i=1}^{3} \left( \frac{P_s(i)}{B(i)} + P_d(i) \right)$$  \hspace{1cm} (5.2)

where:

- $P_s(i)$ is the packet size of HIP UPDATE packet $i$;
- $B(i)$ is the bandwidth available for transmission of the UPDATE packet $i$;
- $P_d(i)$ is the propagation delay for transmission of the UPDATE packet $i$.

As discussed in Chapter 4, the HIP style handover provides a shorter RTT and lower binding cost to the overall network than does the Mobile IP. The HIP handover is more efficient. HIP uses only three UPDATE packets for the whole handover, while Mobile IPv6 needs to use eight packets (six for Return Routability and two for Binding Update).

HIP-enabled heterogeneous wireless network architecture can also provide improvement in vertical handover by using the HIP feature. The use of the handover from UMTS to WLAN in the proposed HIP-enabled UMTS/WLAN architecture exploits the HIP multi-homing feature and the process adopts the Make-Before-Break strategy. An MN can communicate with the CN through the old network interface (UMTS), while the signalling message (e.g. HIP UPDATE packets) can be passed via the new network interface (WLAN). After the handover process completes and the new data stream starts flowing, the MN switches from the UMTS to the WLAN interface. As the handover delay in this scenario is the time needed for switching
network interfaces, it can be close to theoretical. However, handover delay still exists in practice due to the processing time and buffering strategy in the MN. But it is shorter than that in Mobile IP, which does not support Multi-homing. Handover strategy in Mobile IP has to be Break-Before-Make. Even in a loose coupling UMTS/WLAN network, without the Make-Before-Break strategy, HIP still outperforms Mobile IPv6 in vertical handover by 69% [5], so the advantage of an HIP-based strategy over Mobile IP in future heterogeneous wireless communication networks is clear.

5.5 SUMMARY

In this chapter, several new and novel HIP-enabled 3G network architectures have been proposed. These architectures are the initial start for wireless communications to migrate from current IP-based networks into public key-based IP networks. In our proposals, the RVS is modified to map the HI/HIT, not only to the IP address, but also with the mobile subscriber information. The RVS needs to have communication interfaces with GGSN and SGSN in 3GPP UMTS network architecture. It is in the same location as Mobile IP HA in 3GPP2 CDMA2000 network. RVS becomes a part of 3G core networks, and management and billing policy can be re-used directly without major modification. RVS is integrated into different 3G networks to enable the HIP signalling messages to be bounded in the internal networks, and improves the pre-session mobility. The HIP multi-homing feature and Make-Before-Break strategy can reduce vertical handover delay to virtually zero. In the next chapter, HIP’s adoption in a Voice over IP environment will be discussed.
Chapter 6: Hybrid SIP-HIP (SHIP)

In the previous chapter, we have discussed the application of HIP to current wireless communication networks in order to improve the vertical handover performance. New and novel architectures were proposed to improve the performance. One of the most important applications of the wireless communication system is the voice communication. In the 3G system, or future all IP-based wireless communication system, SIP [48] is and will be the main signalling protocol for Voice over IP (VoIP) or multi-media communications. However, mobility management support of SIP is limited and TCP traffic is not supported unless IP encapsulation is used. As a text based application layer protocol, SIP’s handover delay and overhead are large in some circumstances.

Currently hybrid Mobile IP and SIP (M-SIP) [1, 2] is proposed as a solution to address the limitations of SIP mobility support and efficiency. Mobile IP is used for non-real time communication and SIP is used for the real time communication [53, 54]. SIP and Mobile IP’s handover are jointly processed to reduce the handover delay.

In this chapter, a new alternative hybrid SIP-HIP (SHIP) scheme is proposed and analysed. In SHIP, mobility management for the non-real time communication is supported by HIP and the mobility management for real time communication are processed by the joint HIP and SIP (SHIP). Unlike M-SIP, which has two handover processes, SHIP needs only one HIP handover process in order for an MN to move from one network connection point to another. The proposed SHIP is more significantly efficient than M-SIP in the handover process. Furthermore, SHIP provides a secure environment for VoIP communication.

6.1 SIP MOBILITY MANAGEMENT AND ITS LIMITATION

SIP, a widely adopted application layer signalling protocol, is the main signalling protocol for VoIP application in IP networks including 3G networks [25, 49]. SIP is a simple, scalable text-based protocol for session management. It is used to establish and dismantle the
session between nodes. Unlike Mobile IP and HIP, SIP is not only capable of handling terminal mobility, but also session, personal and service mobility [50-52].

SIP URI is used to identify the SIP user in SIP architecture. When an MN is roaming into a foreign network, MN will re-register itself with the new location in its SIP Registration Server in its home network. SIP Registration Server is always implemented with an SIP Proxy Server or Re-direct Server. The caller can reach the MN when the SIP Proxy Server or Re-direct Server re-directs the SIP INVITE message to MN’s new location. This situation is called ‘pre-call mobility’. Figure 6-1 demonstrates the SIP signalling flow. When the caller sends the SIP INVITE to the recipient of the call, SIP Server replies SIP 302 Moved Temporary with the new location of the recipient. The caller will send the new SIP INVITE to the current location of the recipient.

Besides pre-call mobility, there is the other situation called mid-call mobility. For example, in Figure 6-2, when the MN is moving to a new network during communication, the MN does not only need to re-register itself in its Registration Server, but it also needs to notify its CN about the change of IP address. The MN sends a SIP INVITE message directly to CN with the update IP address. CN communicates with the MN’s new IP address after it has received the SIP INVITE message.

![Figure 6-1 SIP Pre-call Mobility](image1)

![Figure 6-2 SIP Mid-call Mobility](image2)
Besides the aforementioned terminal mobility, SIP can also provide session mobility. SIP allows user to maintain a session while moving from one terminal to another. For example, in Figure 6-3, Bob is using his mobile phone (bob@mobile) to communicate with Alice. Bob wants to use his fixed line phone instead. bob@mobile will send a SIP REFER message to Alice with its fixed line SIP URI (bob@fixedline). Alice will send an SIP INVITE with Referred-by bob@mobile message to bob@fixedline. After the connection between Alice and bob@fixedline has been established, the communication stream between Alice and bob@mobile will be torn down.

SIP also allows a single user to be reached at different terminals via the same logical address. This is called personal mobility. A user may be reachable via a traditional Public Switched Telephone Network (PSTN) phone, a PC and a mobile device at the same time or alternate between them. By using SIP forking proxies, the user can be reached through different devices via the same SIP URI, making his device choice transparent to third parties.

Service mobility allows users to maintain access to their service when users are changing devices or devices are changing network attachment points. SIP provides this function through the “home” server which stores the information of users. Users’ devices will contact the “home” server to obtain the service information when users want to use any particular service.

Figure 6-3 SIP Session Mobility using call transfer
In this chapter, we are focusing on the terminal mobility rather than session, personal and service mobility.

6.1.1 Limitation of SIP Terminal Mobility Management

The terminal mobility management provided by SIP is limited. SIP is restricted to all SIP applications. IP TCP traffic encapsulation is needed in both MN and CN [53].

SIP is a text-based application layer protocol with the lowest processing priority in the network protocol stack. The SIP packets (including SIP based handover signalling packets) will be handled after the under-layer protocols. The overhead of SIP handover signalling is large. Everything leads to a longer handover delay. According to the study in [53], handover performance of SIP will be degraded in a high frequent mobility environment.

6.1.2 Hybrid Mobile IP SIP

It is evident that SIP does not provide mobility management for all applications and SIP is not an efficient type of mobility management for frequently moving MN. For SIP applications,
it is good for real time traffic. Mobile IP is suitable for terminal mobility management. For the non-real time traffic, it works well and it does not need to be greatly modified in the current protocol stack. A Hybrid Mobile IP and SIP solution (M-SIP) is proposed for mobility management. Mobile IP and SIP can work independently, but their performance is not good. In an M-SIP network, Home Agents will be integrated with SIP Registrar Servers [53, 54]. Figure 6-5 shows the simplified M-SIP handover message sequence in IPv4 networks. For the handover to occur in real-time SIP-based communication, MN sends the Registration Request to the FA and the FA forwards this request to its HA. After HA has received the Registration Request, it forwards all the MN packets to MN’s new location via FA. At the same time, once the SIP application in MN has detected the change of IP address, MN sends the SIP INVITE packet to CN to notify the change of IP. The SIP media session will be re-established at the new MN’s location. This can help to avoid the triangle routing problem. This situation also applies to Mobile IPv6 or Mobile IPv4 with Router Optimization Extension. Binding Update (BU) will be used instead of Registration Request.

M-SIP is good for overall mobility management as it inherits the advantages of both Mobile IP and SIP. Considering that HIP can outperform Mobile IP in the terminal mobility management, we have proposed a new hybrid solution, Hybrid SIP-HIP (SHIP) in this chapter.
6.2 HYBRID SIP-HIP (SHIP) SOLUTION

6.2.1 Similarity between SIP and HIP

Both SIP and HIP provide mobility management, but in different layers. Their common architectural characteristics will be discussed in this section.

Both SIP and HIP are signalling protocols. The former is for real time communication, and the latter is for overall secure connection. SIP does not define under-layer protocols in an SIP session. Different protocols, such as UDP and SCTP, can be used in the transport layer. HIP also has similar characteristics, theoretically, since it can use any different protocols for upper layer communication, although only IPSec [65, 76] and SRTP [67] are discussed in IETF HIP Working Group so far [66, 68].

In terms of architecture, both use new namespaces as host identifier. SIP uses SIP URI, while HIP uses HI/HIT as host identifier. They use their own mechanisms to perform mapping between IP addresses and their host identifiers. If the location of a destination node is known, SIP and HIP can perform peer to peer connection to communicate directly. They do not need additional network entities for mobility management. MNs will notify CNs directly about the change of their IP addresses. In the secure VoIP environment, SIP will use IPSec or SRTP connection for secure communication. IPSec and SRTP are the widely discussed protocols for HIP connection.

6.2.2 Hybrid SIP-HIP (SHIP) Architecture

As an alternative multi-layer mobility management solution to M-SIP, an integrated SIP and HIP (SHIP) is presented in the following.

SIP does not support HIP natively. As described in Chapter 2, SIP uses SDP [77] message to exchange the SIP based multi-media connection information. Session Description Protocol (SDP) is a text-based protocol for describing the initial parameters of streaming media sessions and multicast transmission. SDP still uses an IP address to identity the host. We need to modify the SDP message in order to support HI/HIT. SDP contains the parameter of the SIP media session. In the SHIP proposal, the owner parameter (‘o’ parameter in SDP message) needs
to be modified in order to store the owner HIT. A new address type (addrtype) called HIT is
added into the current addrtype. At around the same time of our proposal, IETF also suggested
putting the HIT in the encryption key parameter (‘k’ parameter in SDP message) or ‘a’ parameter
[78]. Each uses a different approach but for the same purposes. Figure 6-6 demonstrates a SIP
INVITE packet with SDP message in the SHIP, which uses ‘k’ parameter to store HIT
information.

```
INVITE sip:SECE@rmit.edu.au SIP/2.0
Via: SIP/2.0/UDP proxy.rmit.edu.au:5060
Max-Forwards: 70
To: sip:SECE@rmit.edu.au
From: sip:student1@student.rmit.edu.au
Call-ID: a50b4c76f46738
CSeq: 218153 INVITE
Contact: student1@student.rmit.edu.au
Content-Type: application/sdp
Content-Length: ...

v=0
o=student1 63914747 4753907367 IN IP4 proxy.rmit.edu.au
s=Session SDP
t=0 0
c=IN IP4 proxy.rmit.edu.au
m=audio 3456 RTP/AVP 0 1 3 99
a=rtpmap:0 PCMU/8000
k=host
```

Figure 6-6 SHIP based SIP INVITE (with SDP)

In an HIP-based network, a node is no longer represented by its IP address, but by its
HI/HIT. One of the problems in a SHIP network is the mapping between SIP URIs, HIs/HITs
and IP addresses. SIP Registrar Servers and Location servers no longer store the map between the
SIP URI and IP address; instead, they store the map between SIP URI, HI/HIT and IP address of
RVS instead. The HIP RVS in the same network should be implemented with a location server or
SIP Registrar Server. Alternatively, the RVS should have a direct connection to SIP Registrar
Server to optimize the performance. HIP RVS also needs to be modified. Users are allowed to
request the mapping between IP and HIT directly without any HIP packets. This is similar to the
normal DNS process.

**SHIP scheme (HIP Based Media Session Only)**

There are two possible session establishment flows in SHIP. One provides protection for
the media session only. The other provides protection for the SIP signalling and media session. In
the former scheme as shown in Figure 6-7, when Bob (CN) wants to communicate with Alice’s
mobile (MN), similar to a pure SIP environment, Bob will send SIP INVITE to Alice. The SDP message in the INVITE will contain Bob’s HIT. The INVITE message will be delivered to the SIP Registrar, Proxy or Re-direct Server if necessary. SIP Registrar Sever will communicate with HIP RVS to obtain Alice’s updated location and then Proxy or Re-direct Server will pass the INVITE message to Alice. Once Alice accepts the phone call, the HIP-based media session is established. HIP Base Exchange is carried out while the call is ringing. This reduces the media session set-up time.

![Figure 6-7 SHIP Scheme (HIP Based Media Session Only)](image)

**SHIP Scheme (HIP Based SIP Signalling and Media Session)**

In the second scheme as shown in Figure 6-8, the same scenario is used for demonstration: Bob is going to call Alice. Bob will send an I1 message without the receiver’s HIT and SIP INVITE message together with Alice’s SIP URI to her SIP Registrar, Proxy or Re-direct Server. SIP Proxy or Re-direct Server obtains her HI/HIT in the location server and looks up Alice’s IP in RVS. The SIP Proxy or Re-direct Server forwards both I1 and SIP INVITE to Alice. Alice’s SIP UA will process I1 to establish an HIP connection to Bob before handling SIP
INVITE message. After the HIP connection has been established, the entire SIP signalling messages and media session will be on top of that HIP connection.

**SHIP Handover mechanism**

The handover mechanism is exactly the same in both schemes. When Alice’s mobile phone is roaming into a new network, Alice’s mobile will send an HIP UPDATE packet with the new IP address to Bob. Unlike the M-SIP solution, no SIP re-INVITE message is necessary. As the media session is HIP-based, once Alice receives the HIP UPDATE packet, Bob will communicate with Alice’s mobile via new IP address. Moreover, the handover process may utilize the multi-homing feature to provide smoother handover of Alice’s mobile handovers between different wireless interfaces.
6.3 DISCUSSION AND ANALYSIS

6.3.1 SIP security Enhancement

SHIP provides a secure multi-media data connection for the VoIP. In traditional SIP, SIP uses IPSec or SRTP for secure communication. However, traditional IPSec connection is bounded to IP address. When SIP UA performs handover, IPSec connection needs to be re-established.

In SHIP, the secure communication is HIP based. It means the IPSec connection is no longer bounded to the IP address, but HI/HIT. When a handover occurs, the MN uses HIP handover mechanism to redirect the communication stream into the new IP address. As the communication stream is bounded to HI/HIT, SIP UA does not require a new IPSec connection to be re-established unless a re-keying is necessary. This reduces the network signalling overhead.

6.3.2 Handover Delay Analysis

Binding update signalling delay analysis

Binding update signalling delay is an important measure of handover performance. SHIP handover signalling will be compared with other protocols in the following part. The methods are similar to those in [79], in which the binding update signalling is the only factor to be compared in this part; the other processes before or after binding update signalling message, such as RR process of Mobile IP and acknowledge of messages, are not be included.

![Handoff signalling flow](image)
Figure 6-9 shows the basic scenario of the handover process. We can find that the handover delay in the basic scenario is equal to the sum of the time of requesting an IP address in DHCP and the time of binding update signalling from MN to CN. It is shown in the following equation $D_{\text{handover}} = D_{\text{dhcp}} + D_{\text{notice}}$ [79], where the symbols implementation can be found in Table 6-1. Mobile IP, SIP and HIP are working at different layers. This analysis focuses only on the notice message of the protocol concerned.

### Table 6-1 Input parameters for handover

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Typical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{\text{dhcp}}$</td>
<td>Delay of DHCP address assignment</td>
<td>1s</td>
</tr>
<tr>
<td>$D_{\text{notice}}$</td>
<td>Delay for MH to notify CH of its new location</td>
<td></td>
</tr>
<tr>
<td>$BW_{\text{wired}}$</td>
<td>Bandwidth of wired links</td>
<td>100Mb/s</td>
</tr>
<tr>
<td>$BW_{\text{wireless}}$</td>
<td>Bandwidth of wireless links</td>
<td>11Mb/s (802.11b)</td>
</tr>
<tr>
<td>$L_{\text{wired}}$</td>
<td>Latency of wired links (propagation delay + link-layer delay)</td>
<td>0.5ms</td>
</tr>
<tr>
<td>$L_{\text{wireless}}$</td>
<td>Latency of wireless links (propagation delay + link-layer delay)</td>
<td>2ms</td>
</tr>
<tr>
<td>$H$</td>
<td>Distance between MH and CH in hops</td>
<td></td>
</tr>
<tr>
<td>$L$</td>
<td>IP packet length of notice message</td>
<td>140 bytes (SIP re-Invite's SDP message)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80 bytes (HIP UPDATE with REA, SEQ parameters)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>56 bytes (Mobile IP binding update)</td>
</tr>
<tr>
<td>$T_s$</td>
<td>Average time for which MH remains in a subnet</td>
<td></td>
</tr>
</tbody>
</table>

$D_{\text{notice}}$ depends on the distance between the MN and CN. The formula can be rewritten as follows:
Figure 6-10 shows the binding update signalling delay. SIP has the largest delay and Mobile IP has the shortest delay. HIP and SHIP (which are using the same method to notify the change of IP) are slightly worse than Mobile IP. The performance of M-SIP is same as for Mobile IP. SIP is a text-based protocol, so the packet size is larger than other protocols. The performance of SHIP is an improvement on SIP. However, due to the various sizes of HIP_SIGNATURE (typically 40 bytes are needed for each HIP UPDATE packet; typically 80 bytes with LOCATOR, SEQ and HIP_SIGNATURE), HIP/SHIP has a longer handover signalling delay than Mobile IP and M-SIP. But the performance of Mobile IP/M-SIP and HIP/SHIP are very close. In a later discussion, we will see that HIP/SHIP is better than Mobile IP/M-SIP in terms of overall handover processing delay.

![Figure 6-10 Binding Update Signalling Delay](image)

**Binding Update Signalling Load**

In a dynamic real world, MNs may move very rapidly from one access point to another access point. The more frequent the movement of MN, the more binding will be the update signalling sent to the network. In the M-SIP, although it provides the shortest binding update signalling delay, it requests two binding update signalling messages, one for Mobile IP and the other one for SIP. When the new SIP session is not established, it uses the Mobile IP mechanism to redirect the communication stream into the MN. It can reduce the delay for SIP applications,
but the SIP re-INVITE process is needed to overcome the triangle routing problem in Mobile IPv4. In this aspect, the handover mechanism in SHIP is more efficient as only a HIP handover signalling is needed and no SIP re-INVITE message is necessary. If there are many frequent moving MNs in the network, the load of the network traffic in SHIP is much lower than that of M-SIP.

**Binding Update Signalling Overhead**

The signalling overhead for handover can be shown by \( \frac{L \times H}{T_s} \). Figure 6-11 shows the overhead of the handover signalling of different protocols. Similar to the result of handover signalling delay, SIP has the largest overhead, while Mobile IP has the smallest one in a single layer protocol environment. However, in a multi-layer protocol environment, SHIP has a smaller signalling overhead than does M-SIP. The overhead of M-SIP is the sum of the Mobile IP and SIP as the two handover processes are needed in the M-SIP scheme. In contrast, in SHIP, only the HIP handover mechanism is used; no SIP re-INVITE is needed. The overhead of binding update signalling of SHIP is the same as for HIP.

![Figure 6-11 Handover signalling overhead (H=50)](image)

**Handover processing delay**

In terms of the binding update signalling delay, Mobile IP and M-SIP outperform other protocols. In the previous sections, we have shown that the overall handover delay of HIP is shorter than that of Mobile IPv6. Here, we analyse the overall handover processing delay by modifying the mechanism used in [80].
To provide a secure environment, Mobile IP RR, which has been discussed in Chapter 4, is needed in Mobile IPv6 or M-SIPv6 before the BU packet is sent.

Table 6-2 lists the symbols for the analysis. $t_{\text{MN->HA}}$, $t_{\text{HA->CN}}$, $t_{\text{MN->CN}}$, $t_{\text{HA->FA}}$ are the end-to-end delays between MN, CN, HA and FA respectively for a successful packet transmission. To simplify the analysis, we assume $t_{A->B}$ is equal to $t_{B->A}$ and also the processing time for a node A to forward a packet from one interface to another is zero. We also assume that FA is somewhere on the same path between HA and MN. We are focusing on the round trip time (RTT) delay in this analysis.

Table 6-2 Parameters for handover processing delay[80]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value or Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_d$</td>
<td>24ms (RTT between the MN and DHCP; Derived from [81])</td>
</tr>
<tr>
<td>$t_{Mc}$ or $t_{Hc}$</td>
<td>0 ms (Time to configure a new IP address and start the Mobile IP or HIP handover process. Negligible since Mobile IP and HIP are implemented in kernel. Equal value assume in Mobile IPv4 and v6)</td>
</tr>
<tr>
<td>$t_{sc}$</td>
<td>100 ms (Time to configure a new IP address and stand SIP handover process.)[80]</td>
</tr>
<tr>
<td>$t_{RR}$</td>
<td>$2 \times \max (t_{\text{MN-&gt;HA}} + t_{\text{HA-&gt;CN}}, t_{\text{MN-&gt;CN}})$ (Mobile IP RR process)</td>
</tr>
<tr>
<td>$t_{BU_{-MIPv4}}$</td>
<td>$2 \times (t_{\text{MN-&gt;FA}} + t_{\text{FA-&gt;HA}}) = 2t_{\text{MN-&gt;HA}}$ (Binding update delay in Mobile IPv4)</td>
</tr>
<tr>
<td>$t_{BU_{-MIPv6}}$</td>
<td>$2t_{\text{MN-&gt;CN}}$ (Binding update delay in Mobile IPv6)</td>
</tr>
<tr>
<td>$t_{BU_{-HIP}}$</td>
<td>$3t_{\text{MN-&gt;CN}}$ (Binding update delay in HIP)</td>
</tr>
<tr>
<td>$t_{BU_{-SIP}}$</td>
<td>$2t_{\text{MN-&gt;CN}}$ (Binding update delay in SIP)</td>
</tr>
</tbody>
</table>

Table 6-3 Handover process delay for single protocol

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Mobile IPv4</th>
<th>Mobile IPv6</th>
<th>HIP</th>
<th>SIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handover delay</td>
<td>$t_{MIPv4} = 2t_d + t_{Mc} + t_{BU_{-MIPv4}}$</td>
<td>$t_{MIPv6} = 2t_d + t_{Mc} + t_{RR} + t_{BU_{-MIPv6}}$ (with RR process) or $t_{MIPv6} = 2t_d + t_{Mc} + t_{BU_{-MIPv6}}$ (without RR process)</td>
<td>$t_{HIP} = 2t_d + t_{Hc} + t_{BU_{-HIP}}$</td>
<td>$T_{SIP} = 2t_d + t_{Sc} + t_{BU_{-SIP}}$</td>
</tr>
</tbody>
</table>
We assume that the IP configuration of Mobile IPv4/v6 is by the DHCP method, which is shown in Figure 6-9. Four packets need to be exchanged. The Mobile IPv4 covered in this analysis is the standard Mobile IPv4 without the extension of packet routing. Table 6-3 gives the equation of the overall handover delay of different single layer protocols.

As the RR process is needed before Mobile IPv6 BU packet if the token has expired, there are two possible delays for Mobile IPv6. One is with RR process, and the other one is not. If an MN moves rapidly among a number of access points in the network, the RR process may not be necessary, since the interval between two handover processes is shorter than the lifetime of token created in RR process. However, this rarely occurs. Table 6-4 gives the overall handover delay of multi-layer mobility management protocols.

Table 6-4  Handover process delay for multi-layer protocol stacks

<table>
<thead>
<tr>
<th>Protocol</th>
<th>M-SIP in IPv4</th>
<th>M-SIP in IPv6</th>
<th>SHIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handover delay</td>
<td>( T_{M-SIPv4} = \min (T_{MIPv4}, T_{SIP}) )</td>
<td>( T_{M-SIPv6} = \min (T_{MIPv6}, T_{SIP}) )</td>
<td>( T_{SHIP} = T_{HIP} )</td>
</tr>
</tbody>
</table>

Figure 6-12 and Figure 6-13 show the overall handover delay against the delay between MN and CN, and the delay between MN and HA respectively. When the delay between MN and HA and the delay between HA and CN are fixed (e.g. 30ms and 25ms respectively), Mobile IPv4 and M-SIPv4 perform best in overall handover delay. This is because Mobile IPv4 and M-SIPv4
are independent from the delay between MN and CN. The Registration Request packets are handled only in FA and HA. Although, Mobile IPv4 provides the shortest handover delay, the drawback is the triangle routing. M-SIPv4 is using the SIP re-INVITE to create a multi-media route directly to overcome this problem. Mobile IPv6 and M-SIPv6 without RR processes are the second best in the overall handover delay. The result is consistent with our previous binding update signalling delay analysis. However, it is very rare to have two handovers continuously before the token has expired, so this result is not useful for real world scenarios. Moreover, when the RR process is carried out, the performance of Mobile IPv6 degrades dynamically and becomes the worst in this analysis. M-SIPv6 is dependent on either Mobile IPv6 or SIP. As SIP handover delay is shorter than that of Mobile IPv6, so the handover delay of M-SIPv6 with RR process is the same as for SIP. Similar to the result of the analysis in Chapter 4, handover delay of HIP is shorter than Mobile IPv6 with the RR process. As SHIP is using the HIP handover mechanism, the handover delay is the same as for HIP. They have shorter handover delays than does SIP.

If we fix MN and CN delay with 20ms but leave the delay between MN and HA arbitrary, Mobile IPv4 and M-SIPv4 are no longer the best candidates. SIP, HIP and SHIP have stable handover delays as they are independent of HA. HIP and SHIP outperform SIP significantly. Mobile IP related solutions are the worse in this situation. Mobile IPv6 with RR
process is the worst. Mobile IPv4 is better than SIP, but when the delay between MN and HA is longer, the overall delay of SIP becomes shorter than that of M-SIPv4.

Generally speaking, the overall handover performance of SHIP is better than that of M-SIP regardless of whether IPv4 or IPv6 are used. The overall handover performance of M-SIP is dependent on the delay between MN, HA and CN, while that of SHIP is dependent only on the delay between MN and CH. The signalling overhead and the load to the network of SHIP are small. Moreover, unlike M-SIPv4, the packeting routing is always optimized with a short handover delay.

6.3.3 Comparison of different SHIP schemes

We have proposed two SHIP schemes in 6.2. One has the HIP-based media session only, and the other one has a HIP-based media session with SIP signalling. In the first scheme, caller and recipient exchange HI/HIT in SIP signalling messages. The time required for the initial call ringing is similar to that of the current SIP or M-SIP scheme. It has a longer time in the second scheme, as HIP Base Exchange must be completed before any SIP signalling. In addition, the second scheme is more difficult adopt in the current VoIP architecture. In the first scheme, all SIP network entries can be reused almost without any modification. The significant changes occur in SIP Registrar, Proxy and Redirect Server, and location server. They are based on HI/HIT instead of the IP address to locate the SIP UA. However, in the second scheme, it is different. Even though it provides a more secure environment to protect the media session, and the SIP signalling between two nodes, it makes it impossible for the other SIP network entries to read the SIP signalling message, as they are going to be protected by IPSec. Many of the current solutions, such as interconnection with PSTN network, no longer work in the second scheme. It requires a lot of modifications or re-designing. The first scheme is more suitable to be applied into the current existing network, while the second scheme is only suitable for the high security direct peer-to-peer communication.
6.4 LIMITATION OF SHIP

Although SHIP can provide a better performance in multi-layer mobility management, it has not yet been adopted in the commercial network and multi-casting and is still not supported in HIP. The functionality and performance can be further enhanced with the future development of HIP.

Another obstacle to the adoption of SHIP in the current network is the current lack of micro-mobility management for HIP. Most of the micro-mobility management protocol is based on Mobile IP. In the M-SIP environment, micro-mobility management protocols which are depended on Mobile IP, such as Cellular IP [6], can be used. Without Mobile IP, those micro-mobility protocols cannot provide the full functionality. We are going to introduce our new proposed HIP-based micro-mobility management protocol called micro-HIP (mHIP) in order to solve this problem in next chapter.

6.5 SUMMARY

In this chapter, we have proposed and analysed a new multi-layer mobility management solution, SHIP, for future VoIP networks. It is a hybrid scheme of SIP and HIP (SHIP). It can provide a complete mobility management for all services.

Compared with M-SIP, the widely discussed multi-layer mobility management solution, SHIP is better in overall handover processing performance and efficiency. SHIP avoids the re-INVITE message in SIP and therefore, its handover signalling overhead is smaller. With the HIP handover mechanism, SHIP provides a shorter handover processing delay than M-SIP. Its performance and functions could be further enhanced with the future version of HIP. By applying SHIP in the HIP aware 3G or NGN network, as discussed in the previous chapter, it can provide a secure VoIP environment. We believe that SHIP can be a good candidate for all-round mobility management in future wireless IP networks.
Chapter 7: HIP and Micro-Mobility Management

Micro-mobility management is a niche topic in mobility management worth considering. An example of micro-mobility is a laptop switching from a wire Ethernet to a wireless Ethernet. Most handovers in the wireless network are intra-domain handovers. The path from the gateway to the CN remains unchanged after intra-domain handover. Using the macro-mobility management protocol to handle intra-domain handover wastes network resources and introduces unnecessary handover delay. A micro-mobility scheme is supposed to handle the intra-domain handover to reduce the handover latency and packet loss.

Even though HIP outperforms Mobile IP in the handover, most of micro-mobility management protocols are based on Mobile IP, as HIP is still very new.

In this chapter, firstly the design objectives of HIP-based micro-mobility management protocols are set. Secondly, existing studies of micro-mobility management related to HIP are reviewed. Finally, an improvement on one of the schemes is proposed.

7.1 MACRO-MOBILITY MANAGEMENT VS MICRO-MOBILITY MANAGEMENT

Mobility management can be divided into two categories: macro-mobility management and micro-mobility management [55, 57, 58]. In macro-mobility management, handover between any two IP addresses is treated equally. The movement of an MN can be either within a domain or across domains. After the handover in the Physical Layer and Data Link Layer are complete, the MN notifies its CN of the change of IP address. The CN conducts the handover signalling and sends the communication packets to the new IP address. This simplifies the handover process in any given situation. Because the handover is processed by the CN, the process needs to implement only in the terminal nodes kernel. No modifications are required for other network
components, unless special servers need to be introduced, such as HA in Mobile IP and RVS in HIP. However, the unnecessary handover delay for the intra-domain handover is obvious.

In the scenario of an MN’s micro movement, the path from the CN to the network gateway remains unchanged. Macro-mobility management can still be used for the intra-domain handover. However, the signalling message to the CN (indicating the intra-domain handover) is unnecessary. The intra-domain handover can be handled internally by the network components and there is no impact on the CN. In another words, the movement of an MN within a domain is hidden from the CN. This intra-domain handover processing can reduce handover latency and keep the movement of an MN within a specific domain confidential. When an MN performs the handover between domains, the macro-mobility protocol will be used instead. This strategy can improve handover performance but the process is more complex. The network components, such as routers, need to be upgraded into a micro-mobility-enabled middle box in order to handle this process. To date, most micro-mobility processes are Mobile IP-based, such as, Cellular IP [6]. HIP outperforms Mobile IP in macro-mobility management [5], but there is no complete solution to HIP micro-mobility management so far. In the next section, the development of HIP micro-mobility management is discussed.

7.2 OBJECTIVES OF HIP-BASED MICRO-MOBILITY MANAGEMENT

Micro-mobility management protocols target the intra-domain handover performance. The main target of micro-mobility management is to reduce the handover latency and the packet loss in the handover process. The handover signalling volume in a micro-mobility management protocol should not be more than that of the macro-mobility management protocol.

Location confidentiality can be achieved by micro-mobility management. As the intra-domain handover signalling is limited within the domain, the change of IP address is mapped by the network components inside the domain, so the CN (and any other external network components) does not know the movement of the MN. The external nodes cannot decipher the structure of the network by intercepting packets from MNs.
Another consideration in developing HIP-based micro-mobility management is the issue of security. HIP is a secure protocol [3, 4] and the HIP-based micro-mobility management should maintain the security level. Moreover, HIP supports different security connection schemes, such as IPSec ESP [66] and SRTP header [68]. The HIP-based micro-mobility management protocol should also be able to support any connections scheme.

Multi-homing is one of the important features of HIP, and HIP-based micro-mobility management should also support this feature natively. The targets of HIP-based micro-mobility management are summarised in the following:

**Objectives**

1. Reduction of Intra-Domain Handover Latency  
2. Reduction of packet loss  
3. Restriction of handover signalling volume  
4. Maintaining the HIP security level  
5. Support of different HIP connection schemes  
6. Multi-homing support

### 7.3 EXISTING WORKS OF HIP RELATED MICRO-MOBILITY MANAGEMENT

Most of the micro-mobility management protocols developed so far are Mobile IP based. And there is only limited discussion about HIP in this area at this moment. Besides our proposed scheme in Chapter 8, there is only one study targeting the HIP-based micro-mobility management. Beside that, HIP is being used to demonstrate the notion of improving the security of micro-mobility management. This section reviews these two existing schemes. The following subsection proposes a new extension to one of these schemes to reduce unnecessary signalling.

#### 7.3.1 Novaczki’s Scheme

Novaczki et al. of Budapest University of Technology and Economics have proposed a solution for HIP-based micro-mobility management [82]. A Local RVS (LRVS), which acts as a Mobile Routing Point (MRP), a micro-mobility management scheme enhanced router, is introduced in their scheme. The concept is similar to the foreign agent in Mobile IPv4. An MN
needs to register itself not only in RVS, but also in the LRVS. When an MN carries out an intra-domain handover, it notifies the LRVS directly about the change of IP address. The LRVS directs all communication streams into its new location. The initialization and intra-domain handover operation under Novaczki’s Scheme are given in the following steps.

1. An MN connects to an access point (AP) of a domain. The AP broadcasts a modified ICMPv6 Router Advertisement message that contains the HIT and IP of LRVS.
2. The MN registers itself in the LRVS.
3. The MN registers itself in the RVS via LRVS.
4. The LRVS modifies the source IP address of MN’s registration packet into a global routable IP address and forwards it to the RVS.
5. RVS registers the MN’s information with the LRVS IP address.
6. A CN sends I1 to the MN’s RVS to initiate the HIP-based connection.
7. The RVS forwards the I1 packet to MN’s LRVS.
8. The LRVS forwards the I1 to MN based on its database record.
9. The MN and the CN process the rest of the HIP-Base Exchange via the LRVS.
10. HIP connection is established. The LRVS learns the flow identifier (destination IP, SPI, ESP) of the established SA pair during the HIP Base Exchange.
11. The MN performs the intra-domain handover.
12. The MN sends an UPDATE packet to the LRVS to update its IP address.
13. The LRVS redirects the MN’s communication stream based on the SA information.

Novaczki’s scheme is highly dependent on the LRVS for micro-mobility management. The LRVS acts as a gateway to map all HIP-based connections to the corresponding MNs. According to the description in [82], Novaczki’s scheme can reduce the intra-domain handover latency of HIP. Moreover, as all the handovers are still using the HIP UPDATE mechanism for the handover process, the handover signalling volume is restricted. However, the paper discusses only HIP mobility with single SA pairing; it does not cover multi-homing scenarios. Furthermore, Novaczki’s scheme uses ESP information as the flow identifier, so it can be applied only to the HIP-ESP based connection. It does not fulfil all the objectives set out in Chapter 7.2.

7.3.2 Ylitalo’s Scheme

Ylitalo et al. from Ericsson Research NomadicLab are studying the security of IP-based micro-mobility [83] and they used HIP to demonstrate their scheme. Although their aim was not
to target HIP-based micro-mobility management, we still consider it as a candidate for HIP-based micro-mobility management.

The Lamport one-way hash chain [84, 85] and secret splitting technique [86, 87] were proposed to implement a secure micro-mobility management and so the middle boxes inside the domain need to be upgraded in order to handle the intra-domain signalling. Ylitalo’s scheme uses the Lamport one-way hash chain with delay authorisation to allow the middle box to authorize signalling packets without gleaning any additional information. The Lamport one-way hash chain can also prevent acquiring passwords / authentication information by eavesdropping.

**Lamport One-Way Hash Chain**

The Lamport One-Way Hash Chain is designed to protect passwords from two possible cracking methods: one is by gaining access to the passwords stored inside a system (i.e. reading the password file); two is by intercepting users’ communication with the system (i.e. eavesdropping) [84]. Lamport used characteristics of the hash function, $y = H(x)$, to overcome this problem. The hash function is easily calculated for the value of $y$ by knowing the value of $x$, while it is quite difficult to obtain the value of $x$ by knowing the value of $y$. The operation of the Lamport One-Way Hash Chain is as follows:

1. The user generates a random number $x$, which is called ‘seed’.
2. The user needs to define the length, $L$, of the Lamport One-way Hash Chain.
3. The generate hash chain is $H^{L-1}(x), H^{L-2}(x), \ldots \ldots H(H(H(H(x)))), H(H(x)), H(x), x$, where $x_i = H^{(L-1)-i}(x)$.
4. The user and system exchange the $x_0$ in the secret channel.
5. The system stores $x_{i-1}$ for the password authentication.
6. The user enters the password $x_i$ for the authentication purpose.
7. The system verifies $x_i$ by checking if $x_{i-1} = H(x_i)$. If it is verifies, the system stores $x_i$. The user needs to input $x_{i+1}$ for next password authentication process.
8. When the hash chain has reached the seed value $x$, the user needs to define a new seed and re-generate the new hash chain, which repeats from step 1.

As the hash function $H$ is a one-way function, it is very difficult to guess $x_i$ by knowing $x_{i+1}$, so even if the hacker knows $x_{i+1}$ by looking at the password file in the system or...
eavesdropping on the network, the hacker still cannot pass the authentication process by lacking $x_i$ information.

**Ylitalo’s scheme Overview**

The general ideal Ylitalo’s scheme uses the Lamport One-Way Hash chain with delay authentication to allow the middle boxes to verify that signalling packets are from the valid nodes. Moreover, it also uses the split secure key technology to allow the closest anchor point to learn the shared secret key to decrypt the next value of the hash chain to generate a valid signalling packet. An anchor point in Ylitalo’s scheme is an upgraded middle box that is used to maintain the states for each MN in its region. Middle boxes do not care who generates the intra-domain handover signalling packet, they only care if signalling packets are generated by a holder who knows the next valid hash chain value.

An MN attaches a current hash value ($H_i$) and an encrypted next hash value ($H_{i+1}$) in the handover signalling message and signs the packet by $H_i$. When the handover signalling packet is processing in the other node (e.g. CN or the old anchor point), it decrypts $H_{i+1}$ and sends it back to the MN in the signalling packet. The middle boxes use $H_{i+1}$ to verify $H_i$ as well as the signature of the previous signalling packet. Similarly, when the MN replies to that feedback signalling packet, it also attaches $H_{i+2}$ to allow the middle box to verify the $H_{i+1}$ in the previous signalling packet which was sent by the other node. With this strategy, the middle box does not need to know who generates the signalling packets; it needs to know if the signalling is generated by the valid hash chain value holder. Middle boxes assume that only the MN or trusted node has the next valid value in the hash chain.

The MN uses the split-key technique to allow an anchor point in the region to learn the secret key. The anchor point acts as a CN in handling the intra-domain handover signalling. The MN uses secret splitting techniques to allow the closest anchor point to learn the common secret for decryption for the next handover signalling packet of the intra-domain handover. We discuss this detailed operation in the next section.

**Ylitalo’s scheme operation**

The Ylitalo’s micro-mobility handover process is described in the following [83]:

---

© 2008 Joseph Y.H. So
1. After the HIP-Base Exchange is complete, a MN requests a macro-mobility handover.
   A. The MN bootstraps a hash chains by revealing $H_i (i=0)$ in a HIP UPDATE signalling packet.
   B. The MN also generates a secret key, splits it into two pieces, $K_1$ and $K_2$, where the secret key is $K_{\text{xor} \ 2}$. The MN sends the $K_1$ and hash of $K_{\text{xor} \ 2} (\text{by SHA1})$ and an encrypted $K_2$ and $H_{i+2}$ with the $H_i$ value in the UPDATE packet to the CN. The packet is signed by the $H_{i+1}$.

2. When the closest anchor point receives this UPDATE packet, it will zero the $K_1$, remember SHA1($K_{\text{xor} \ 2}$) and forward the packet to the CN. All the other middle boxes in the path will authenticate this UPDATE packet later.

3. When the CN receives the UPDATE packet, it decrypts $K_2$ and $H_{i+2}$. The CN has the $H_{i+2}$ value, so it can also generate $H_{i+1}$, as $H_{i+1}$ is equal to hash of $H_{i+2}$, i.e. $H_{i+1} = \text{Hash}(H_{i+2})$. The CN sends Address Checking (AC) packet. $K_2$ and $H_{i+1}$ are attached in the AC packet. This packet is signed by $H_{i+2}$.

4. When middle boxes receive the AC packet, they use the $H_{i+1}$ to verify the previous UPDATE packet sent by the MN. This AC packet will be verified later.

5. The closest anchor point receives this AC packet and zeros $K_2$. It verifies the $K_1$ and $K_2$ by using SHA1($K_{\text{xor} \ 2}$). It forwards the AC packet to the MN.

6. MN sends the Address Check Reply (ACR) packet to the CN. It contains $H_{i+2}$. Middle boxes can use $H_{i+2}$ to verify the previous AC packet.

7. When MN conducts the intra-domain handover, it repeats step 1B to step 2 again. It creates a new $K_1$ and $K_2$. It encrypts the new $K_2$ and $H_{i+2}$ by the old $K_1$ and $K_2$.

8. When the closest middle box in the old path receives this UPDATE packet, it re-directs the UPDATE packet to the old anchor point instead of the CN. The old anchor point is the only node which knows the old $K_1$ and $K_2$, so it can decrypt the $K_2$ and $H_{i+2}$. The old anchor point takes the role of the CN in this intra-domain handover to generate the AC packet (which is signed by $H_{i+1}$) and sends it to the MN via the closest middle box in the old path.

9. The MN replies to the AC with the $H_{i+2}$ embedded in the packet.

10. After the AC has been verified in middle boxes, middle boxes will redirect MN’s packets to the CN’s new location.

11. Once the Lamport One-Way Hash Chains reach the seed number, the MN needs to notify the CN of the new Lamport one-way hash chain. This can be done using a macro-mobility handover process.

   Figure 7-1 depicts the intra-domain handover. The old anchor point acts as CN for this handover. Figure 7-2 depicts the general packet structure for the handover signalling.
Analysis of Ylitalo’s scheme

Ylitalo’s scheme is focused on improving the security level of micro-mobility management. Most of the objectives, which we have listed for the micro-mobility management, are out-of-scope in their design. According to [83], the intra-domain handover performance is better than that of HIP, but micro-mobility management cannot be processed instantly. At least one macro-mobility handover per connection is required in order to allow the closest anchor...
point to learn the shared secret key. The MN still needs to notify the RVS about the change of IP address, so it does not implement the micro-mobility management completely. Furthermore, when the Lamport Hash Chain reaches the seed value, the MN needs to perform a macro handover process. To conclude, Yliatlo’s scheme improves only the security level of intra-domain handover and the handover latency, but it does provide a new perspective in the discussion of HIP-based micro-mobility management.

In summary, Ylitalo’s scheme starts the discussion of secure environment for the micro-mobility management while, Novacckiz’s scheme provides real HIP-based micro-mobility management. Novacczki’s scheme is highly centralised in the LRVS, while Ylitalo’s scheme disrupts the load into different middle boxes. The load of the LRVS is very heavy if there are a lot of rapidly moving MNs in the domain. If the load of the LRVS is too heavy, the queuing time and the processing time will be increased, causing a negative feedback to the handover latency of MN. Both schemes only partially fulfil the objectives we have set for HIP-based micro-mobility management.

### 7.4 IMPROVEMENT OF YLITALO’S SCHEME – ILHC SCHEME

Ylitalo’s scheme limits most of the intra-domain handover signalling inside the domain. One of the problems is that if an MN performs intra-domain handover rapidly, the hash keys will run out quickly and the MN needs to notify the CN of the new Lamport One-Way Hash Chain. In this section, we propose a modification to Ylitalo’s scheme, so as to remove the unnecessary bootstrap of the hash chain.

#### 7.4.1 Infinite Length Hash Chain (ILHC)

The Infinite Length Hash Chain (ILHC) [88] was developed to improve the Lampot One-Way Hash Chain. The disadvantage of the Lampot One-Way Hash Chain is that the user is allowed to authenticate with the system N times only. For the \(N+1^{th}\) authentication, the system needs to restart or bootstrap the other Lampot Hash Chain. This causes an extra signalling packet in Ylitalo’s scheme when the chain reaches the seed value. By adopting the ILHC, the signalling packet for the rapidly moving MN can be reduced.
Following are the definitions of symbols used in the later discussion:

- Let function $A$ be a public-key algorithm, where $d$ is the private key and $e$ is the public key.
- Let $s$ and $c$ be an original data and ciphertext
  
  $A(s, d) = c$  \hspace{1cm} (7.1)  
  
  $A(c, e) = s$  \hspace{1cm} (7.2)  

- Let $A^N(s, d)$ denotes running function $A$ N times recursively to the initial input $s$ using the private key $d$.

ILHC uses the nature of public and private key pairs to create a one-way hash chain. The ILHC is outlined below:

$s, A(s, d), A^2(s, d), A^3(s, d), \ldots, A^{N-1}(s, d), A^N(s, d), \ldots$  \hspace{1cm} (7.3)

The user generates a one-time password by using his private key in the ILHC scheme:

$A(s, d) = P_0$  \hspace{1cm} (7.4)

When the system authenticates the password, the password can be verified by applying a user public key, where

$s = A(P_0, e)$  \hspace{1cm} (7.5)

The general formula for the $i^{th}$ password is:

$P_i = A^i(s, d)$  \hspace{1cm} (7.6)

By knowing password $P_i$ and without knowing the user private key $d$, it is impossible to generate $P_{i+1}$. However, the system can authenticate $P_i$ easily by:

$P_{i+1} = A(P_0, e)$  \hspace{1cm} (7.7)

The overall process is similar to the Lampot One-Way Hash Chain. The system can authenticate the $N^{th}$ one-time password by knowing the $N-1^{th}$ password. Unlike the traditional hash chain, the first item of the ILHC is the seed value, whereas in the traditional hash chain, it is the last item. In theory, the ILHC can be of unlimited length by re-applying the function $A$ recursively.
7.4.2 New ILHC Scheme

The HI, which is used as a node identity in HIP-based networks, is a public key. In the HIP Base Exchange, two nodes exchange their public key in the R1 and I2 packets. In the normal HIP Base Exchange process, the Initiator may encrypt its public key in the I2 packet. As in the ILHC Scheme, the Lampot One-Way Hash Chain is replaced by ILHC, so the HI can re-use the public key for the ILHC process. As middle boxes need to know the public key of the MN, if an MN is the Initiator in the HIP Base Exchange process, it must not encrypt its public key in I2. Middle boxes learn the public key of an MN during the HIP Base Process.

In the macro- and micro-mobility management process, there is no difference between Ylitalo’s Scheme and the ILHC scheme. The only difference lies in its parameters. Figure 7-3 shows the packet structure of the ILHC scheme. All Lampot One-way Hash Chain parameters (Hi) are replaced by ILHC (Pi). The [H0\text{new}] is removed, as it no longer needs to exchange the first value of the hash chain.

| UPDATE: A, K1, SPIs, HMAC (Pi+1, SHA1(A)) |
| where A = EIDs, SHA1(K1\text{xor}2), ENC_{K2,e} (K2||P_{i+2}), ENC_{K1\text{old}, K1\text{xor}2} (K2||P_{i+2}), P_i |
| AC: EIDs, K2, Pi+1, SPIs, HMAC(P_{i+2}, SHA1(EIDs||K2||P_{i+1})) |
| ACR: EIDs, P_{i+2}, HMAC(K_{1\text{xor}2}, EIDs||P_{i+2}), HMAC(K_{2,e}, EIDs||P_{i+2}) |

Figure 7-3 ILHC scheme micro- and macro-mobility exchanges use common packet structure.

When the MN sends out the UPDATE packet, it needs to include not only its HIT, but also its HI to allow middle boxes, which are not in the old path, to perform the delay authentication. The middle box verifies the UPDATE and AC packets when it receives the AC and ACR packets.

In this ILHC scheme, HI is integrated into part of the delay authentication. As one characteristic of HI is self-certification, it can verify itself by HIT and the signature that is generated by the MN using its private key. The ILHC scheme provides further proof about the owner of the signalling messages and it reduces unnecessary signalling of Ylitalo’s scheme. As the length of ILHC is unlimited, the MN does not need to bootstrap the hash chain again, no matter how many times the MN has processed the handover.
7.5 SUMMARY

In this chapter, six objectives of the HIP-base micro-mobility management protocol are set and two current studies related to HIP-based micro-mobility management protocols are reviewed. Ylitalo’s scheme focuses on security improvement in micro-mobility management; the MN uses the Lamport One-way Hash Chains and splitting key technologies to improve the security level in micro-mobility management. Novaczki’s scheme uses a centralised server, Local RVS (LRVS) to handle all intra-domain handover signalling. However, both schemes only partially fulfil our objectives.

We have proposed a new scheme, Infinite Length Hash China (ILHC) scheme, which is an improvement on Ylitalo’s scheme, in this chapter. We use ILHC to replace the Lamport One-Way Hash Chain to perform delay authentication to reduce unnecessary signalling in Ylitalo’s scheme.

In the next chapter, a new HIP-based micro-mobility management scheme, micro-HIP (mHIP) is proposed.
Chapter 8: Micro-HIP (mHIP)

HIP is a secure wireless IP network mobility management protocol. The extension of HIP to micro-mobility management becomes an interesting topic. In the previous chapter, the requirements of the HIP-based micro-mobility management have been discussed. Two related schemes have been reviewed. Our analysis shows that neither has met all the requirements.

In this chapter, a new HIP-based micro-mobility management called Micro-HIP (mHIP) is proposed. It addresses all the requirements set for a HIP-based micro-mobility management scheme. This chapter is organised as follows:

First a general framework and architecture of mHIP is introduced. Then the security arrangement and the overlay routing in mHIP architecture are presented. The operation process of mHIP is described. Finally, the analysis and discussion on the security and the performance of mHIP are provided.

8.1 MHIP OVERVIEW

Our goal is to design an HIP-based mobility management protocol, which fulfils all the objectives we have set in Chapter 7. Micro-HIP (mHIP) is such a scheme proposed by us. It was developed during the same period that Novackzi’s scheme and Ylitalo’s Scheme were proposed.

In a general micro-mobility management protocol, some middle boxes are introduced to the network. We call all these micro-mobility management-enhanced middle boxes Mobile Routing Points (MRP). The intra-domain handover is handled by these MRPs. Besides, they also re-direct MNs’ packets to the updated location of MNs. In our proposed mHIP architecture, these MRPs are called mHIP Agents, which will be discussed in detail later.

In the standard HIP handover process (macro-mobility management), a CN needs to send an HIP Address Checking (AC) packet to an MN to make sure the new address is reachable. This AC packet is signed by the CN. In HIP-based micro-mobility management, this AC packet is sent by a MRP instead of a CN. In order to maintain HIP’s security level as in a macro
environment, a MN accepts only the AC packet which is from an authorized node. Also, other MRPs in the network need to be able to verify HIP-based intra-domain handover signalling in order to redirect HIP-based data packets to an MN. Lamport One-Way Hash Chain, delay authorization and splitting secret technology are used in the Ylitalo’s scheme, which has been introduced in the Chapter 7. However, a MN needs to generate a new secret key in each handover. Furthermore, MRPs cannot verify the current signalling packet until they receive the next signalling packet. An MRP needs to have a buffer to store the previous signalling packet. Different from Ylitalo’s scheme, Novaczki’s scheme uses a single server as MRP, i.e. Local RVS (LRVS), to handle all intra-domain handover signalling packets. As all handover signalling packets are handled by a LRVS, the load of the LRVS is very heavy. Our mHIP scheme has multiple all mHIP Agents, which can handle HIP intra-domain handover signalling packets. To combat a masquerade attack, mHIP Agents need valid signatures to allow all other mHIP Agents and an MN to verify the AC packet.

The routing of HIP-based data packets is an important issue. An HIP header exists only in a HIP signalling packet, but not in the HIP data packet in practice, even though conceptually a HIP header is supposed to be present in an HIP data packet. Ylitalo’s scheme and Novaczki’s scheme use an SPI field in the IPSec header to identify the HIP data packet for overlay routing. But, this strategy cannot be extended to other non-HIP ESP data packets. We will introduce an overlay routing scheme in mHIP to accommodate any HIP-based communication scheme.

8.2 MHIP AGENT

An mHIP Agent is an mHIP enhanced middle box. mHIP Agents handle intra-domain handover signalling. After an intra-domain handover has been completed, the mHIP Agents in the domain will update their mapping and re-direct packets to the updated location.

All mHIP Agents in the same domain form one virtual centralised server. Apart from the original HI/HIT that mHIP Agents have, they also share one common HI/HIT, which is called “Net HI/HIT”. This shared HI/HIT does not represent a host but a whole network, similar to the Net ID concept in the IP address. All the mHIP Agents can sign messages on behalf of the group.
By adopting a suitable signature scheme, no matter which mHIP Agent signs a packet, a MN can verify the signature without knowing which particular mHIP Agent signed it.

mHIP Agents are classified into two categories: the mHIP Gateway and mHIP Routers. An mHIP Router is a 3.5 layer router, which re-directs HIP packets to a MN’s current location. An mHIP Router also handles the intra-domain handover signalling.

An mHIP Gateway is a special mHIP Router and it is the root of all mHIP agents. It manages all mHIP Routers in its domain. mHIP Routers have to register themselves in the mHIP Gateway before they can handle micro-mobility management.

Figure 8-1 depicts the simplifier mHIP architecture.

![mHIP Architecture Diagram]

**Figure 8-1 mHIP Architecture**

### 8.2.1 mHIP Router

The role of an mHIP Router is similar to that of a Cellular IP Node. A mHIP Router re-directs HIP-based data packets to an MN. The mHIP Router learns the mapping during a HIP Base Exchange or a HIP UPDATE signalling. The mHIP Router gets the MN information (or next mHIP Router information) from the reverse path of the signalling.
When an mHIP Router registers itself in an mHIP Gateway, the mHIP Gateway exchanges information with the router to allow the mHIP Router to sign the packet on behalf of the whole domain. These information parameters are encrypted by the HI of the mHIP Router (public key). mHIP Routers also broadcast a modified IPv6 Router Advertisement to alert the MN of the HIT and IP address of the mHIP Gateway.

mHIP Routers handle HIP-based intra-domain handover signalling. Once the closest mHIP Router in the old path receives an intra-domain handover signalling, it acts as a CN to reply the handover message. After the handover process has been completed, it updates the mapping and redirects the packet to the updated location.

The details operation of mHIP Router registration and handling handover signalling packet will be presented in Chapter 8.5.

8.2.2 mHIP Gateway

An mHIP domain must have one mHIP Gateway for interfacing with external and internal networks. The HIT of an mHIP Gateway acts as a Net HIT to represent all mHIP Agents inside the domain.

When an mHIP Router is configured in the network, it registers itself in the mHIP Gateway. The mHIP Gateway authenticates mHIP Routers based on its local security policy. The mHIP Gateway generates security parameters and sends them to mHIP Routers in the HIP signalling packets.

The mHIP Gateway also acts as an LRVS in the mHIP domain. All MNs are required to register themselves in the mHIP Gateway. The process is similar to the standard HIP Registration process. The mHIP Gateway modifies source IP addresses of out-going HIP-based packets by its own routable IP address.

Detailed operation of mHIP Gateway is given in Chapter 8.5.

8.3 SIGNATURE SCHEMES IN MHIP

We have mentioned that a HIP-based micro-mobility management should maintain the HIP security level in the previous section. Every HIP signalling packet needs to be signed and
verified by the MN or other mHIP Agents. Novaczki’s scheme uses a single LRVS to handle and sign AC packets, while Ylitalo’s scheme uses the Lamport One-Way Hash Chain, delay authorization and splitting secret technology for the verification of AC packets. A new signature scheme is developed in mHIP.

In mHIP, all mHIP Agents form a virtual server. The role of this virtual server is same as a LRVS in Novaczki’s scheme. The Net HI/HIT is the HI/HIT of this virtual server. An MN can use the Net HI/HIT to verify the signature in an AC packet. All mHIP Agents can generate a valid signature on behalf of this virtual server. From the MN point of view, the signalling packets are signed by one virtual server. Following are the requirements set for a signature scheme in the mHIP scheme:

- Compulsory:
  - Unforgeability: only mHIP agents are able to sign messages on behalf of the group.

- Optional, but highly recommended:
  - Anonymity: the identity of the actual signer is hidden from the verifier
  - Unlinkability: it should be hard to tell if two different valid signatures are from a same agent.

Apart from the requirements mentioned above, the mHIP network should also allow a new mHIP Router to join at any time without modifying the current setting of the mHIP network, i.e. changing the HI/HIT. Moreover, the scheme should use HI as the Public-Key in the signature validation algorithm. HI is the public key, so reusing the HI in the signature scheme can reduce the signalling volume.

There are some signature schemes for a group to allow one of the group members to create an electronic signature for a document on behalf of the group or as a proxy signer. These schemes are not designed for network management. The application of these schemes in network management will be proposed. In the following section, several signature candidates and their implementation in mHIP are discussed.
8.3.1 Shared Private Key Scheme

The Shared Private Key Scheme is the easiest scheme to allow a group of people to sign a message on behalf of a group. The group members share a common private key. When a group member wants to sign the packet on behalf of the group, he uses this common private key to sign the message. The receiver uses the corresponding public key to verify the signature. The algorithm is a normal public key signature scheme.

Applying Shared Private Key Scheme in mHIP

The Shared Private Key Scheme is the easiest way to be implemented in mHIP. The HI/HIT of the mHIP Gateway is the Net HI/HIT.

Once an mHIP Gateway authenticates an mHIP Router to join an mHIP domain, the mHIP Gateway sends the corresponding private key of the Net HI/HIT to the mHIP Router, which is encrypted by the public key of the mHIP Router. The mHIP Router knows the private key pair of the HI, so it can sign an AC packet on behalf of the mHIP Gateway. An MN knows only that the AC packet is signed by the mHIP Gateway.

Although this scheme fulfils all the requirements listed above, it breaks the fundamental assumption of public-key algorithm. The private key is supposed to be kept secret and should not be sent to any third party. There is a risk of leaking the private key during the communication if there is any security threat to the system.

8.3.2 Group Signature Scheme

A Group Signature scheme should allow group members to sign messages on behalf of the group anonymously. The concept of the Group Signature scheme was introduced by Chaum and van Heyest in [89]. A receiver can use a group public key to verify the signature, but is unable to know which member of the group signed it. Only a designated group manager can open the group signature and trace the actual signer. As such, an outsider is unable to identity whether the signature is generated by the same signer. A member is also unable to sign a message on behalf of the other members. The properties of a secure group signature scheme are defined as follows:

- Unforgeability,
• Anonymity,
• Unlinkability,
• Exculpability,
• Traceability.

Exculpability means no group member or group manager can sign on behalf of other
group members and traceability means the group manager can always open a valid signature and
identify the actual signer.

Group Signature scheme is not yet widely applied. The reason is that most of the Group
Signature Schemes’ algorithms are inefficient [90] and they are insecure [90-98], particularly
with regards to linkability and not traceability. The creation of an efficient secure group signature
is still a challenge and a hot topic.

However, the Group Signature scheme still meets our basic requirement set for an mHIP
Agents’ signature scheme. Most of the studies in efficient Group Signature algorithms are
classified as insecure because of traceability and linkability [90-98]. Our mHIP scheme focuses
only on unforgeability. Traceability or linkability is not compulsory. Only authorized nodes can
generate a valid signature for the signalling packet, and the mHIP does not require any nodes to
trace the signer. In this section, we choose the Xie-Yu Group Signature algorithm [99] because of
its efficiency in signature generation and verification. Here is a brief description of the Xie-Yu
Group Signature algorithm.

**System Initialization Phase**

The Xie-Yu scheme involves four parties: the trusted centre, the group manager (named
as group authority in [99]), the group members and the verifier (receiver). There are four phases
in the Xie-Yu scheme: the system initialization phase, partial secret key generation phase, group
signature generation and verification phase, and signer identity verification phase.

The system initialization phase is used to create the common parameters for the signature
verification.

1. The trust centre chooses four large primes: \( p, q, p' \) and \( q' \), where \( p = 2p' + 1 \), and
\( q = 2q'1 + 1 \).
2. Let $g$ be a generator of a multiplicative sub-group $\mathbb{Z}_N^*$ with order $v = p'q'$.
3. Chooses $e$ randomly as the $\gcd(e,v) = 1$ and $ed \equiv 1 \pmod{v}$.
   
   Let $h()$ be a one-way collision resistant cryptographic hash function.
4. The trust centre assigns a group secret key $x$.
5. Compute the group public key, $y = g^v \pmod{N}$.
6. $e$, $y$, $N$, $g$ and $h()$ are published, while $p$, $q$, $p'$, $q'$, $d$, $x$ and $v$ are kept as secret.

**Partial Secret Key Generation Phase**

The partial secret key generation phase is the process that generates keys for each group member to sign messages [99]:

Let $A = \{U_1, U_2, \ldots U_n\}$ be the group of $n$ members.

1. The trust centre assigns $x_G$ as the group manager private key.
2. The trust centre assigns a large prime $ID_i$ as $U_i$’s secret identity information, where $U_i \in A$.
3. Compute $U_i$’s partial secret key by $x_i = ID_i x d$.
4. Compute $U_i$’s identity verification parameter $T_i = g^{ID_i} \pmod{N}$.
5. Compute $U_i$’s public key $y_i = T_i^{x_G} \pmod{N}$.
6. Compute group authority’s public key $y_G = g^{x_G} \pmod{N}$.
7. The trust centre sends $\{x, T, ID_i\}$ to each $U_i$ and $\{x_G, T\}$ to the group manager.
8. The trust centre publishes $y_i$ and $y_G$.
9. Group members can verify the parameters by $T_i^{x_i e} = g^x = y \pmod{N}$ and $y_i^{ID_i} = T_i^{x_G ID_i} = y_G \pmod{N}$.

**Group Signature generation and verification phase**

When $U_i$ signs a message $m$ on behalf of the group, it performs the following steps to generate the signature [99]:

1. Randomly chooses a large prime $k$.
2. Computes $z = k ID_i$, and $r = g^k \pmod{N}$.
3. Computes its secret key $s_i = x k$.
4. Computes $c = h(r^e \pmod{N}, z, r, m)$, $s = k - s_i c$, and $A = T_i^c \pmod{N}$.
5. $U_i$ sends the group signature $\{c, s, z, r, A\}$ to the verifier.

When the verifier receives the message and the signature, it verifies the signature with the following step:

1. Computes $R = g^{s e} y^c \pmod{N}$.
2. Verifies
A. \( C = h(R, z, r, m) \)
B. \( R = r^e \mod N \)
C. \( A^i = r^e \mod N \)

If all of the above equations hold, then the group signature is valid.

There is also a ‘signer identity verification phase’, but this is redundant in the mHIP scheme and not used. Our mHIP scheme focuses only on the message generated and signed by a valid group member, and does not subsequently trace the actual signer.

**Applying Group Signature Scheme in mHIP**

By applying the Xie-Yu Group Signature scheme in the mHIP scheme, the group public key, \( y \), is the HI and the hash of \( y \) is HIT. The mHIP Gateway is the trusted centre and the group manager. When a node enters a new mHIP domain, it registers itself in the mHIP Gateway and the mHIP sends \( \{e, y, N, g, h()\} \) in R2. The mHIP Gateway signs the packets with group secret key \( x \). When a new mHIP Router is assigned to the mHIP domain, the mHIP Router registers itself via the normal mHIP registration process. The mHIP Gateway encrypts the \( \{y, x, T, ID\} \) and sends it to the mHIP Router in R2.

When an MN performs intra-domain handovers, it sends an HIP UPDATE packet with its HI, \( H_{\text{MN}} \). The closest mHIP Router in the old path handles the intra-domain handover signalling. The closest mHIP Router in the old path replies to the AC packets with the parameters \( \{c, s, z, r, A\} \) to MN. The \( H_{\text{MN}} \) parameter lets other mHIP Routers, which do not know the HI of MN, to verify the packet from the MN. The MN and other mHIP routers validate the AC by checking the group signature. (If \( c, s, z, r \) and \( A \) are zero, the AC packet is sent by the mHIP gateway; the signature is validated by a normal HIP signature checking process). As the MN knows the group public key \( y \) and other necessary parameters \( (e, y, N, g \) and \( h() \)\), it can verify if the AC is from valid mHIP routers. For the details operation, please refer to Chapter 8.5.

Figure 8-2 depicts the basic format of a micro-handover packet by applying a Xie-Yu Group Signature scheme. The UPDATE and ACR packets are the same as the standard HIP packets, while the HMAC of AC is replaced by the Group Signature parameter.
Even though the Group Signature scheme can be used as a signature scheme for mHIP Agents, some of its features are useless for the mHIP Agent and the computing resources are wasted.

8.3.3 **Ring Signature Scheme**

Ring Signature schemes are a downgraded version of the Group Signature scheme, in which there is no group manager and all nodes form a ring in order to sign the message on behalf of the group. The concept of the Ring Signature was introduced by Rivest et al. in 2001 [100]. A secure Ring Signature scheme needs to be unforgeable, anonymous, unlinkable and untraceable. A signer or a verifier needs to know all public keys of the other members in the ring in order to generate or verify a ring signature.

**Applying Ring Signature Scheme in mHIP**

Although in theory the Ring Signature meets the requirements of the mHIP agent signature scheme, it is recommended for implementation only in small networks, because an MN needs to know all public keys of all mHIP Agents. However, if the network is small, one LRVS / mHIP Gateway is enough to handle all intra-domain handover signalling packets. Considering all these factors, we exclude the Ring Signature from the mHIP signature scheme consideration list.

8.3.4 **Proxy Signature Scheme**

Proxy Signature scheme was introduced by Mambo et al. in 1996 [101]. The original signer is allowed to delegate its signing power to other parties, which are called proxy signers, to sign the packet on behalf of the original signer. The properties of a Proxy Signature are unforgeability and traceability. The original signer generates a proxy private key to allow the proxy signer to sign the message.

If a Proxy Signature is applied to mHIP, all mHIP Routers will be proxy signers of the mHIP Gateway to sign the packet. The mHIP Gateway will pass the necessary information to

---

**Figure 8-2 Group Signature mHIP micro-mobility handover packets**

UPDATE: EIDs, SPIs, HMAC (EIDs, SPIs,
AC: EIDs, SPIs, {c , s, z, r, A},where c=h(r^e mod N, z ,r, (EIDs, SPIs))
ACR: EIDs, HMAC(HI_{MN}, EIDs)
mHIP Router to allow them to sign the packet as a proxy. The HI/HIT of mHIP Gateway is the Net HI/HIT.

However, most Proxy Signature schemes have proven to be insecure, i.e., forgeable [102-104]. The design of an efficient and secure Proxy Signature scheme is still a hot topic. At this stage, existing algorithms of Proxy Signature scheme are not suitable yet for mHIP networks. Therefore, they cannot be considered in our selection.

8.3.5 Summary of Signature Schemes

The four different signature schemes are reviewed and summarized in the following.

The Shared Private Key Scheme fulfils all requirements listed in the Chapter 8.3 and it is easy to implement. No extra parameter needs to be introduced in the HIP signalling packet. However, it breaks the fundamental assumption of Public Key algorithm, i.e. private key should not be shared. There is potential risk that the private key can be leaked.

The Group Signature Scheme allows MNs or other network components to verify the signature efficient. The implementation of the scheme in mHIP is proposed. The only concern is that some of its features are not needed in mHIP; therefore the full version of the scheme would be a waste of resources.

The Ring Signature Scheme is a downgraded version of the Group Signature scheme. The scheme requires that verifiers need to know all signers’ public keys. As it is difficult for MNs to know the entire mHIP Agents’ public keys, the Ring Signature Scheme is not a good candidate for mHIP.

The Proxy Signature Scheme is another good signature scheme in theory, but there is no practical secure Proxy Signature algorithm available.

Although all of the above signature schemes are suitable for mHIP, they each have their own disadvantage. Table 8-1 summarizes the features of different signature schemes discussed. Based on our discussion, the Shared Signature Scheme and the Group Signature Scheme are the acceptable signatures scheme candidates for mHIP.
Table 8-1 Comparison of different signature schemes

<table>
<thead>
<tr>
<th></th>
<th>Shared Private Key Scheme</th>
<th>Group Signature Scheme</th>
<th>Ring Signature Scheme</th>
<th>Proxy Signature Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unforgeability</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (in theory)</td>
</tr>
<tr>
<td>Anonymity</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Unlinkability</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Using Standard HIP signalling packet format</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
</tr>
<tr>
<td>Any other special consideration</td>
<td>Private key not be shared with others</td>
<td>Too Big for mHIP</td>
<td>No</td>
<td>No secure Proxy Signature Algorithm so far</td>
</tr>
</tbody>
</table>

8.4 OVERLAY ROUTING

Overlay Routing is the routing which relies on information presented in both the network layer and the layers above. In HIP, HI/HIT is used to identify a host in upper layer protocols while an IP address is used to identify the location of a node in the Network Layer Standard routing in IP networks is based on the IP address only. In HIP, RVS is used to provide a mapping between the IP address and the HI/HIT. In macro-mobility management, after a connection is established, the MN notifies its CN of the change of IP address. However, all intra-domain handover signalling packets would be limited to the domain in the HIP-based micro-mobility management. The CN and other external network entities are not notified. Therefore, the mapping of the HI/HIT with the updated IP address needs to be conducted within the domain. In theory, all HIP packets come with HIP headers. MRPs should act as a 3.5 layer router to forward HIP-based data packets to the current location of an MN. In theory, the HIP header is next to the IP header in the HIP data packet. When an HIP-based connection is established between an MN and a CN, they would exchange the information related to the connection. The standard HIP header does not provide enough information for the micro-mobility management. In the multi-
homing scenario, an MN may communicate with its peers by different network interfaces. Unlike Mobile IP based micro-mobility management, an MN will not have different HI/HIT for different network interfaces. These network interfaces would share one common HI/HIT. Based on the information in a standard HIP header, it cannot identify to which network interface the HIP data packet should be sent.

Also, an MN should be able to move a connection from one interface to another. To support all these features, we propose a 32-bit field connection number to be added in the HIP header, as shown in Figure 8-3. The 32 bits selection is based on the following consideration. The collision probability of connection number within one MN is very low. Secondly, it fits well in a standard HIP/IP header, whose size is the multiple of 32 bits. This is similar to the consideration of SPI in an IPSec header. This connection number is randomly generated by an MN. A CN is notified of the connection number in the HIP Base Exchange. If the host is in a multi-homing status, and has more than one active connections with the same CN (i.e. SIMA), it needs to assign different connection numbers for different streams.

<table>
<thead>
<tr>
<th>0</th>
<th>8</th>
<th>16</th>
<th>24</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Next Header</td>
<td>Header Length</td>
<td>Packet Type</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Checksum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sender’s Host Identity Tag (HIT)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Receiver’s Host Identity Tag (HIT)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Connection Number</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data Payload</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8-3 Modified HIP Conceptual Header for HIP Data Packet
During the HIP Base Exchange, mHIP Agents in the path learn the connection number for each HIP stream. mHIP Agents remember the destination of the HIT (MN’s HIT) connection number and map it with the IP address of the MN or the next mHIP Agent. mHIP Agents learn the mapping IP from the reverse path of the HIP Base Exchange or HIP UPDATE signalling packets. This is similar to the operation of the Cellular IP Node [6]. Once the intra-domain handover has been completed (refer to 8.5.5 for further details), mHIP Agents re-direct incoming HIP data packets to the MN’s new location.

However, in practice, there is no HIP header in HIP data packets, for example, only the IPSec ESP header in the HIP ESP data packet, as depicted in Figure 8-4(c). Therefore, our consideration is how the connection is implemented in the mHIP overlay routing.

![Figure 8-4 HIP Packet Structure](image)

**8.4.1 mHIP ESP Scheme**

ESP[65] and Secure Real-time Transport Protocol (SRTP)[67] are two security protocols that can be used in HIP [66, 68]. In the HIP ESP scheme, as depicted in Figure 8-4(c), the data packet structure is the same as that of normal IPSec ESP packets. HIP daemon uses the ESP header to identify the connections. The mHIP agents no longer use \(<\text{HIT}_{\text{source}}, \text{HIT}_{\text{destination}}, \text{Connection no}>\) to map to an IP address of an MN. It uses \(<\text{HIT}_{\text{source}}, \text{HIT}_{\text{destination}}, \text{IP}_{\text{source}}, \text{SPI}>\) instead. A source (CN) IP address is needed in the mapping, because SPI is not globally unique. The source IP address is obtained during the handover process. If the CN is also a mobile node, it notifies the MN when the IP address has been changed. The mHIP agents in the path can update the source IP address. SPINAT [105] is one of the overlay routing method schemes on the IPSec SPI, and is presented in the following.
SPINAT is an advanced version of Network Address Translation (NAT). In the basic version of NAT, a global IP address is translated into a local IP address based on an incoming source IP address. The destination IP address in the IP header will be modified into a local IP address, so incoming packets are able to be routed to the destination node. This is one of the solutions to solve the global IP shortage. Network Address Port Translation (NAPT) is the popular advanced version of NAT. The NAPT device conducts the mapping of the external IP address and port number and local IP address and port number in the IP header. The concept of SPINAT is similar to that of NAPT; the address translation at the SPINAT device is based on the SPI value in the IPSec header instead of the port number in the IP header. The translation of SPINAT is depicted in Figure 8-5.

![Figure 8-5 SPINAT](image)

IPSec control signalling, i.e. HIP Base Exchange or IKE, is used to control the state of a connection and to register triggers and communication contexts in a SPINAT device. A trigger consists of an identifier and an IP address of an end-point, and a communication context stores the information to implement the translation, i.e. SPI values [105]. The trigger can be registered by the SA establishment signalling, (i.e. HIP Base Exchange), or SA update signalling, (i.e. HIP UPDATE signalling). The communication context can be created or updated when the receiver replies to the initial trigger signalling message [105].

Stripped End-to-End Tunnel (SEET) mode[105], which is based on IPSec Bound End-to-End Tunnel (BEET) for ESP [106], was proposed for SPINAT to allow the devices to modify the SPI value so that SPI collision is avoided. Therefore, only SPI is needed in order to identity connections. The SEET mode does not change the IPSec ESP header structure, but excludes it in the header integrity protection computation [105].
New mHIP SPINAT

Although SPINAT with SEET mode provides an overlay routing without SPI collision, the SPI value is excluded in the header integrity protection computation, so if a hacker performs a man-in-the-middle attack to change the SPI value, an MN or an MRP cannot detect it. In this section, a new mHIP SPINAT is proposed. It prevents the SPI collision without excluding the SPI value in the header integrity protection computation.

An mHIP Agent modifies the destination IP address to let the normal router route the HIP data packet to the MN’s current location. The mHIP scheme tries to avoid the change of the SPI. The mHIP agent maps the destination IP address to the current IP address of MN based on the source of the IP address and the SPI value. In order to avoid SPI collision in the SPINAT, we present two different ways to modify SPINAT for the mHIP ESP scheme without modifying the incoming SPI value.

In our first modification, the concern is to maintain the current IP header structure. The source IP address is also used in the overlay routing. SPINAT daemons of mHIP agents learn \(<\text{HIT}_{\text{source}}, \text{HIT}_{\text{destination}}, \text{SPI}_{\text{source}}\>\) and corresponding MN IP address in HIP Base Exchange. With incoming HIP data packets, the mHIP agent looks up the mapping to find out the destination IP address based not only on SPI values, but also on the source IP addresses. As the same source would not use the same SPI value in different active connections at any given time, collision will not occur. For nodes which share the same global routable IP address (within a same NAT/mHIP domain), it is possible to have SPI collision, although the probability is very small. We propose to let mHIP Gateway notify the second node to take a different SPI. Moreover, when a CN sends an HIP UPDATE packet to an MN, mHIP Agents also need to update the mapping, because the source IP address has changed.

In our second attempt, we propose a modification of the ESP header. A 128-bit destination HIT field is added into the ESP header. This HIT field is included in the header integrity protection computation. The mHIP learns the \(<\text{HIT}_{\text{source}}, \text{HIT}_{\text{destination}}, \text{SPI}>\) and the MN’s IP address in the HIP Base Exchange. As the destination HIT is included in this header, an mHIP agent does not need to learn other identities, such as source IP addresses, for the mapping.
The SPI takes on the role of the connection number in the HIP conceptual header. In this way, the implementation of mHIP Agents is simpler than the first modification. mHIP Agents do not need to modify the mapping if a CN send out a HIP UPDATE packet. Moreover, even if different connections from the same global routable IP address are using the same SPI number, i.e. under the same NAT / mHIP domain, no extra notification is needed, as their destination HITs are different.

8.4.2 mHIP SRTP Scheme

Apart from the ESP scheme, Secure Real-time Transport Protocol (SRTP) [67] is another protocol being discussed for HIP [68]. It is based on Real-Time Transport Protocol (RTP) [107] with an added security mechanism. A SRTP connection re-uses the RTP header to identify a connection. SRTP is commonly used in SIP protocol to secure the VoIP. However, all the HIP-based micro-mobility management protocols have not yet been concerned with the HIP SRTP connection. In the following, an HIP SRTP based mHIP is proposed.
SRTPNAT

In SRTP, the sender and receiver are required to maintain a cryptographic state information, which is called ‘cryptographic context,’ for each SRTP stream. According to RFC 3555, using the local network address is not sufficient for the identifier, so a Synchronization Source (SSRC) is used. SSRC is “globally unique”, as the probability that two sources independently pick the same SSRC among 1000 sources is $10^{-4}$ [108]. Moreover, according to RFC 3711, a cryptographic context is identified by the triplet context identifier [67]:

$$\text{context id} = \langle \text{SSRC}, \text{destination network address}, \text{destination transport port number} \rangle.$$  

In our proposed SRTPNAT, the SSRC and the source IP address are used as an index to map to the local IP address. mHIP Agents provide a mapping between $\langle \text{HIT}_{\text{source}}, \text{HIT}_{\text{destination}}, \text{SSRC}, \text{IP}_{\text{source}} \rangle$ and corresponding MN IP address.

mHIP Agents learn the SSRC in the HIP Base Exchange, in which the SSRC is included in the General Parameter (SRTP_PARA) of R1 and I2. When mHIP Agents receive new incoming HIP SRTP data packets, they redirect the packets into the current location of the MN based on the SSRC and the source IP address. Similar to SPINAT, if two HIP SRTP data streams from the same global IP source use the same SSRC, the mHIP Gateway notifies the second CN to modify its SSRC. Moreover, when a CN sends an HIP UPDATE packet to an MN, mHIP Agents also need to update the mapping, because of the source IP address being changed.

8.4.3 Overlay Routing Table Storage

Our discussion so far has not touched on the issue of routing or mapping information storage in mHIP Agents. mHIP Agents act as one virtual server in a domain. Routers and other middle boxes between mHIP Agents and APs are also considered as part of this virtual server. In theory, this virtual server connects to all MNs via the APs directly and it is the interface to external networks. In this section, we discuss how to distribute the overlay routing information inside the virtual server.

Distributed Hash Table

Distributed Hash Table (DHT) is a decentralised distributed system that provides a look-up service similar to hash tables such as Chord [109], Pastry [110], Symphony [111], Tapestry
and Content-Addressable Networks (CAN) [113]. It was our first consideration when selecting an algorithm for storing the overlay routing table in mHIP. DHTs were originally designed for peer-to-peer (P2P) systems in order to provide a decentralised, scalable and fault tolerant environment.

The first generation P2P network was unstructured. There were two ways to locate a desired piece of data in P2P network: centralised database (Napster) or flooding of queries (Gnutella). In the centralised environment, the routing table size is \( O(n) \) and the number of hops of look-up is \( O(1) \). However, the load of the centralised server is very heavy. In the flooding of queries method, the table size is reduced to \( O(1) \), but the number of hops required is \( O(n) \). DHTs provide a load balancing solution giving an average performance of \( O(\log N) \) in the data look-up, and the size of the routing table ranges from \( O(\log N) \) to \( O(1) \) [109-113].

The micro-mobility environment is very similar to the file sharing in a P2P network. A client in a P2P network requests a file, which may be attached to any host. A client looks up the location of the file from the database (centralised, flooding or DHT) and requests the host to send the file. In the micro-mobility environment, a CN or a gateway asks for the updated location of an MN and then sends the packet to the MN. The role of the gateway or the CN is similar to a client in the P2P network, while the role of APs is similar to the file hosting nodes. The MN can be projected as the file in the P2P network. The only difference is the direction of the transfer. In the P2P network, the file host sends the file to the client, while it is reversed in the micro-mobility environment. The approach of Novaczki’s scheme is similar to that of Napster in that it uses a centralised server as the controlling entity. In theory, DHTs can be applied to the micro-mobility environment.

However, we discovered that applying DHTs to mHIP is not realistic. Although DHTs can provide the load balancing in the network and reduce the routing table hosting to one node, it takes extra time in the data look-up. Unlike the file searching application, micro-mobility requests fast mapping to the current IP address. Moreover, an MN may perform the handover after the mHIP Gateway has forwarded the packets. mHIP Routers in the path also perform the mapping look-up. The adoption of DHT for all mHIP Agents as overlay routing tables introduces
unnecessary signalling for the domain. Furthermore, the look-up performance also depends on
the traffic in the network: if the network is busy, the lookup performance of DHTs will be
affected. To conclude, DHT implementation for the overlay routing table is not suitable.

**Cellular IP style routing table**

The Cellular IP style routing table is another consideration. Cellular IP [6] is a Mobile IP
based micro-mobility management protocol designed by Columbia University and Ericsson.
mHIP is somewhat similar to Cellular IP in the network structure. In Cellular IP networks, there is
a Cellular IP Gateway and Cellular IP Node to route the packets to the updated location of an
MN. The Cellular IP Node is the special router for a Cellular IP network. The Cellular IP
Gateways and Cellular IP Nodes learn the address mapping from the reverse path. The MN sends
the signalling message to the Cellular IP Gateway and then the Cellular IP Gateway and Cellular
IP Nodes in the path learn the next node information. The Cellular IP Gateway and Cellular IP
store the next node IP, or MN IP for the packets routing purpose.

To apply the Cellular IP style routing table to mHIP Agents, each mHIP Agent needs
to remember the next hope of an mHIP Agent or the MN for each connection. The mHIP Agents
modify the destination IP address in the IP header if it is not matched with its latest record. mHIP
Agent learns the reverse routing path for each connection from HIP Base Exchanges or HIP
UPDATE packets. Because mHIP Routers collectively remember this information, it is not
necessary to send the intra-domain handover signalling to the mHIP Gateway. Therefore, there is
no overloading concern for the mHIP Gateway.

### 8.4.4 Summary of Overlay Routing

In this section, the overlay routing mechanism in mHIP is discussed. Standard
conceptual HIP header structure has been modified by adding a 32-bits connection number to
support micro-mobility management and multi-homing in mHIP. A new mHIP SPINAT and new
SRPTNAT are two proposed overlay routing schemes for mHIP.

Overlay routing table storage in mHIP is also discussed. A distributed storage solution,
Cellular IP style routing table, is found to be effective. More operation details will be presented in
the next section.
8.5 **MHIP PROTOCOL OPERATION**

After our presentation and discussion on mHIP structure, signature and overlay routing schemes, our focus will turn to mHIP operation in this section. As the HIP SRTP scheme is still not mature yet, the presentation is the HIP/mHIP ESP scheme.

8.5.1 **mHIP Router Registration**

After an mHIP Router is configured in the network, it receives a modified ICMP Router Advertisement which contains the IP address and the HIT of a mHIP Gateway in a domain. The mHIP Router registers itself in the mHIP Gateway through the standard HIP Base Exchange and HIP Registration processes [69]. We propose to add two new Register Types (Reg_Type) to the HIP Base Registration process for the purpose of registration.

<table>
<thead>
<tr>
<th>Number</th>
<th>Registration Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>mHIP Gateway (mHIP_Gateway)</td>
</tr>
<tr>
<td>202</td>
<td>mHIP Router Registration (mHIP_Router_Reg)</td>
</tr>
</tbody>
</table>

The operation steps of mHIP Router registration are listed in the followings (Figure 8-7):

1. An mHIP Router starts an HIP Base Exchange by sending I1 packet to the mHIP Gateway.
2. The mHIP Gateway sends R1 packet with a Registration Information (Reg_Info) parameter to the mHIP Router. The Reg_Info contains the Reg_Type value of the mHIP Gateway and the mHIP_Router_Reg.
3. The mHIP Router replies by I2 packet with the Registration Request (Reg_Req) parameters and sets the value of the Reg_Type as mHIP_Router_Reg.
4. The mHIP Gateway replies by R2 packet with Registration Respond (Reg_Res) or Registration Fail (Reg_Fail). If the Group Signature scheme is used, \{e, y, N, g, h()\} are also sent in R2 packet.
5. If the registration is successful, the mHIP Gateway sends the parameters (for example a private key in Shared Private Key Scheme, or \{x, T, ID,\} in the Group Signature key scheme) in the R2. The parameter(s) are encrypted by the mHIP Router’s own public key.
8.5.2 MN Registration in mHIP Gateway

Every MN needs to register itself in the mHIP Gateway in order to benefit from the mHIP.

1. An MN attaches to an AP in the mHIP domain.
2. The AP broadcasts a modified ICMP Router Advertisement which contains the HIT and the IP address of a mHIP Gateway.
3. The MN registers itself in the mHIP Gateway in the HIP Base Exchange process. It sends I1 to the mHIP Gateway.
4. Similar to the mHIP Router Registration process, the mHIP Gateway sends R1 to the MN with the Reg_info parameter.
5. The MN sends I2 to the mHIP Gateway. I2 has the mHIP_Gateway as the value of the Reg_Type in the Reg_Req parameter.
6. The mHIP sends R2 with Reg_Resp to indicate the successful registration. \( e, y, N, g, h() \) are also exchanged in R2 if the mHIP domain uses the Group Signature scheme.
7. The MN registers itself in its RVS in the standard HIP RVS Registration process [64].
8. The mHIP Gateway adds the FROM parameter with its HIT [64] into the I2, which is sent by MN to the RVS, and re-write the source IP address as its own routable IP address.
9. The RVS registers the MN’s information by using the mHIP Gateway’s HIT and global routable IP address.

8.5.3 mHIP Idle Handover

An idle handover is a handover process when there is no active connection to an MN. The mHIP idle handover process updates the primary IP address map of an MN’s HIT to the new location. Instead of HIP Base Exchange, a newly proposed extension of HIP is used to allow MN to update its record in mHIP Gateway by HIP UPDATE packet.

1. After under-layer handover processes are completed, the MN checks the modified ICMP Router Advertisement to see if the HIT of the mHIP Gateway remains unchanged. If yes, it means the MN is still under the same domain.
2. The MN sends an HIP UPDATE packet with Reg_Req value of “mHIP Gateway” in the ESP_INFO; the SPI value is set to NULL.

3. The mHIP Gateway requests the Address Checking with the Reg_Resp value.

4. If the Address Checking process is passed, the mHIP Gateway updates its primary IP map by the MN old address with its new IP address.

![Figure 8-8 mHIP idle handover](image)

8.5.4 mHIP Connection Establishment and Tear Down

The connection establishment in mHIP is similar to that in HIP. If a CN is an initiator, it resolves an MN’s RVS IP address in a DNS. The CN sends I1 packet to the MN via the MN’s RVS. When the mHIP Gateway receives the I1 packet, which is forwarded by the RVS, it forwards this I1 packet to the MN’s current location. The MN uses the standard HIP process to handle the I1 packet. The MN replies with a R1 packet to the CN. All mHIP Agents in the path remember HIT pairs and the IP address of the MN (the MN may use a different network interface to send the R1). They create a state of “Connecting” for the connection. When the MN sends a R2 packet with a SPI value, mHIP Agents in the path change the state to “Connected”, and include the SPI in the mapping. The remainder of the process is the same as for the standard HIP Base Exchange.

When a connection establishment is requested by an MN, the process is similar. As the destination HIT of I1 can be NULL, mHIP Agents do not create any state for this connection at this moment. The mHIP Gateway forwards R1 packet to the MN based on the default mapping. When the MN sends an I2 packet with an SPI value, mHIP Agents store the mapping and set the state as “Connecting”. The mHIP Agents in the path change the state to “Connected”, once the CN replies the MN by R2 packet.
Once the MN or the CN sends a HIP CLOSE packet, mHIP Agents in the path change the connection state to “Closing”. When the other node replies with a HIP CLOSE_ACK packet, mHIP Agents change the connection state to “closed” and then remove the connection mapping.

If there is no active connection for a period of time, mHIP Agents in the path change the state of connection to “timeout” (e.g. the default value of HIP ESP scheme is 15 minutes [66]) and remove the connection mapping.

8.5.5 mHIP intra-domain handover process and multi-homing support

When an MN initiates an intra-domain handover, a mHIP Gateway will update MN’s primary IP address, and the mHIP Agents in the path also need to update this active connection map. However, the MN may be in a multi-homing scenario, so not every connection of the MN needs to perform an update.

1. A MN performs the under-layer handover if necessary.
2. The MN sends an UPDATE packet with LOCATOR parameter to CN.
3. mHIP Agents, which are not in the old path, receive the UPDATE packet and create the connection mapping – the state is “connecting”.
4. When the UPDATE packet reaches the closest mHIP Agent in the old path (mHIP Edge Agent):
   A. If the MN does not change the SPI value (using same SA scenario – Figure 8-10)
i. The mHIP Edge Agent acts as a CN to reply the packet. It changes the connection state to “handover” and keeps the old and new IP address of the MN.

ii. The mHIP agent signs an AC packet. It also attaches the HI of MN in the AC packet.

iii. The MN replies to the AC packet. mHIP agents in the path verify the packet, update the map and change the state to “connected”.

B. If the MN changes the SPI value (a new SA), which means a new connection to CN needed to be established:

i. The updated packet is forwarded to the CN. All the mHIP agents in the path remember the new SPI and the new IP address and change the state to “handover”.

ii. When the CN requests the UPDATE ECHO_REQUEST, mHIP agents forward the packet to the new IP address.

iii. Once the handover has been completed, mHIP Agents update the connection map with the new SPI and IP address. The connection state changes to “connected”.

5. The mHIP Edge Agent may inform the mHIP Agents that are no longer in the path to remove the mapping via the HIP NOTIFY packet. If the mHIP Edge Agent does not inform other mHIP Agents, the active connections in the map are removed when they expire.

Figure 8-10 mHIP intra-domain handover with the same SA
The handover mechanism of HIP is based on connections. In order to minimise the change to the kernel, mHIP uses the same strategy. However, this introduces some unnecessary signalling load. In intra-domain handovers, for a handover between two APs, all the UPDATE packets, sent out by the same interface are handled by the same Bridge mHIP Agent. In order to make the handover more efficient, an optional extension is proposed in the following.

The standard HIP Update packet allows only one ESP_INFO to be included. In our proposal, multiple ESP_INFOs in one HIP UPDATE can be accommodated. In an mHIP handover, if an MN has multiple connections for one interface, the MN sends a HIP UPDATE packet to mHIP Gateway directly instead of CN. This HIP UPDATE packet includes all the SPIs that are related to this handover. By this optional extension, mHIP Agents can handle the handover of connections of the same interface in a batch. The flow of the HIP UPDATE packet is the same as normal mHIP intra-domain handover signalling. If some unknown SPI exists in the HIP UPDATE packet, the mHIP Edge Agent forwards it to the mHIP Gateway.

An MN can perform a handover from one interface to another; the process is the same as above. The MN can also notify its CN about additional interface via the normal HIP process, which creates new SA pairs, and mHIP agents store the new SPI value into the map. If the MN wants to move some connections from one interface to another, the standard mHIP handover mechanism is applied. In the SIMA scenario, different connections to the same CN are treated independently.

### 8.6 ANALYSIS AND DISCUSSION

In this section, we discuss security and performance.

#### 8.6.1 Propagation Delay Analysis

The handover latency is mainly affected by the propagation delay, which is measured by Round Trip Time (RTT). Our analysis in this section is focused on RTT. In order to simplify the discussion, we have made the following assumptions:

- mHIP Agents use a Shared Private Key scheme;
• The computation power of all nodes in the network, such as gateway and MN, is very powerful, so the delay caused by computation (e.g. the processing delay) is zero;

• The latency from Point A to Point B is the same as from Point B to Point A.

**Pre-session Handover Delay**

When an MN moves to a new location, it needs to notify its location look-up server (i.e. LRVS, mHIP Gateway) about its new location. Pre-session handover performance is related to the update latency of the primary IP address mapping in the location lookup server. We first access Ylitalo’s scheme and Novaczki’s scheme and then move on to mHIP.

Ylitalo’s scheme does not target the performance of micro-mobility management, so it does not cover pre-session mobility management. An MN needs to notify its RVS about the change of its IP address, so the pre-session mobility performance of Ylitalo’s scheme is the same as that of HIP. According to the Internet Drafts [64, 69], the MN can register itself only via the HIP Base Exchange, so the MN needs to go through the HIP Base Exchange (I1, R1, I2 and R2). The total RTT is 2 RTT\textsubscript{MN, RVS} where RTT\textsubscript{MN, RVS} = RTT\textsubscript{MN, Gateway} + RTT\textsubscript{Gateway, RVS}.

Novaczki’s scheme uses an LRVS, which is the RVS server for the local domain only. An MN’s RVS forwards all the I1 to MN’s LRVS for the further look-up of the MN’s current location. The functionality of the LRVS is similar to that of a standard RVS. The MN can update its primary IP address in the map by using the HIP UPDATE packet instead of HIP Base Exchange[82]. The LRVS carries the address, checking to make sure that the new address is reachable. Once the MN has received the AC packet, it replies to the LRVS. We can see that the update latency of Novaczki’s scheme is 1.5 RTT\textsubscript{MN, LRVS}. The LRVS is located as a gateway to the network, so the latency can be generalized as 1.5 RTT\textsubscript{MN, Gateway}.

Pre-session mobility handling in our mHIP scheme is similar to that of Novackiz’s scheme. A mHIP Gateway is used as an LRVS. An MN uses a HIP UPDATE packet to notify the mHIP Gateway about the change of its IP address. As the mHIP Gateway and the LRVS are located in the same position and the mechanism is the same, so the update latency is also the same, i.e. 1.5 RTT\textsubscript{MN, Gateway}.
Therefore, the mHIP scheme and Novaczki’s scheme outperform Ylitalos’s scheme in terms of pre-session handover.

**Mid-session Handover Delay**

A handover occurring while an MN has active connections is referred to as a ‘mid-session handover’. Mid-session handover performance is the most important factor for assessing a mobility management protocol. In macro-mobility protocols, the MN notifies its CN directly about the change of its IP address. In the micro-mobility protocol, network components in the local network handle intra-domain handover signalling packets and re-direct all data packets into MN’s new location. To make it simple, we assume that the network is in a balanced tree structure, as depicted in Figure 8-11. All middle boxes in the local domain are upgraded to be MRPs (such as mHIP agents, LRVS) if possible. We also assume that all the links in the local network are identical. Again we are going to access Ylitalo’s scheme and Novaczki’s scheme and then move on to mHIP.

The Novackiz scheme approach to mid-session handover is simple. All intra-domain handover signalling packets are handled by an LRVS. When an MN moves into a new location, it sends a HIP UPDATE packet to the LRVS to redirect data packets to the new location. Similar to the pre-session mobility management, the LRVS will request the address checking. This approach is highly dependent on the single LRVS server. The handover latency is the same as the pre-session mobility that is 1.5 \( RTT_{MN,Gateway} \).

Compared with Novackiz’s scheme, Ylitalo’s scheme distributes the load of handling of intra-domain signalling packets over a local network. Every MRP can co-operate with the anchor point (an anchor point is also a MRP) to handle the intra-domain handover signalling. When an MN performs a handover, it sends out an HIP UPDATE packet. The nearest MRP in the old path (NROP) forwards this UPDATE packet to the previous anchor point (PAP). The PAP acts as a CN to generate the HIP AC packet to the MN via the NROP. The MN sends the HIP ADDRESS CHECKING REPLY (ACR) packet to the PAP via the NROP and then the handover is complete. So, the handover latency is 1.5 \( RTT_{MN,NROP} + RTT_{NROP,PAP} = 1.5 RTT_{MN,PAP via NROP} \). In the worst case, the NROP is the gateway. If the domain network is modelled on a tree structure and
all the APs are leaf nodes at the same level, such as the balanced binary tree in Figure 8-11, the
distance between PAP and NROP is the same as that between MN and NROP. If all the links are
identical, then $\text{RTT}_{\text{MN}, \text{NROP}} = \text{RTT}_{\text{NROP}, \text{PAP}}$, so the handover latency in general is 3 $\text{RTT}_{\text{MN}, \text{NROP}}$;
and the worst case is 3 $\text{RTT}_{\text{MN}, \text{GATEWAY}}$.

![Figure 8-11 Balanced Binary Tree Structure with three-level depth](image)

In HIP, all MRPs (mHIP Agents) in the domain are treated by all MNs as one virtual
server. All intra-domain handover signalling packets are handled by this virtual server. When an
MN sends an intra-domain handover signalling packet, this packet is handled by the nearest
mHIP Agent in the old path, which is an NROP.

An MN sends an intra-handover signalling message to its CN and the NROP acts as a
CN to reply to this HIP UPDATE packet. Similar to other schemes, Address Checking is
required, so the handover latency is 1.5 $\text{RTT}_{\text{MN}, \text{NROP}}$. In the worst case, the NROP is the mHIP
Gateway, so the handover latency is 1.5 $\text{RTT}_{\text{MN}, \text{NROP}} \leq 1.5 \text{RTT}_{\text{MN}, \text{Gateway}}$.

In the balanced tree network structure, the cross-over distance of the HIP UPDATE
packet in the mHIP scheme is 50% less than Ylitalo’s scheme, because the HIP UPDATE packet
is processed in the NROP in the mHIP scheme. Figure 8-12 depicts the result of the simulation of
the average cross-over distance of the HIP UPDATE packet in mid-session handover for
different schemes, with different levels of balanced binary tree network structure. We found that
the average crossover distance of the HIP packet in Novaczki’s scheme is shorter than Ylitalo’s
scheme, and mHIP requires the least distance.
Table 8-2 shows the summary of the handover performance of different HIP-based micro-mobility management.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Pre-session Handover</th>
<th>Mid-session Handover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ylitalo’s</td>
<td>2 RTT&lt;sub&gt;MN, RVS&lt;/sub&gt; same as HIP</td>
<td>1.5RTT&lt;sub&gt;MN, PAP via NROP&lt;/sub&gt; = 1.5(RTT&lt;sub&gt;MN, NROP&lt;/sub&gt; + RTT&lt;sub&gt;NROP, PAP&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Novaczki’s</td>
<td>1.5RTT&lt;sub&gt;MN, Gateway&lt;/sub&gt;</td>
<td>1.5RTT&lt;sub&gt;MN, Gateway&lt;/sub&gt;</td>
</tr>
<tr>
<td>mHIP scheme</td>
<td>1.5RTT&lt;sub&gt;MN, Gateway&lt;/sub&gt;</td>
<td>1.5RTT&lt;sub&gt;MN, NROP&lt;/sub&gt; (&lt;=1.5RTT&lt;sub&gt;MN, Gateway&lt;/sub&gt;)</td>
</tr>
</tbody>
</table>

8.6.2 Signalling Cost Analysis

Signalling cost is another measurement of performance of a mobility management protocol. Signalling cost is the cost of location management in a data network, and it includes the location update cost, binding update cost and look-up cost\[114, 115\]. Before we discuss this in detail, some notions are defined in the following:

- \( N_{MN} \) – Total number of MNs in the network.
- \( N_{CN} \) – Average number of CNs / data stream communicating with a MN (Assume each CN will have only one data stream with the same MN)
- \( \omega \) – Wireless proportionality constant
- \( T_R \) – Average MN residence time in a subnet
- $\lambda_s$ – Average session arrival rate
- $l_{mr}$ – Average number of hops between MN and RVS
- $l_{mc}$ – Average number of hops between MN and CN
- $l_{mg}$ – The number of hop between MN and the Gateway (i.e. LRVS and mHIP Gateway) (Novaczki’s scheme and mHIP scheme)
- $l_{mo}$ – Average number of hops between MN and the nearest MRP in the old path
- $l_{oa}$ – Average number of hops between the nearest MRP in the old path and the previous anchor point (Ylitalo’s scheme)
- $\gamma_m$ -- Registration message, location update message or binding update message processing cost at MN
- $\gamma_{rvs}$ – Registration message, location update message processing cost at RVS
- $\gamma_{cn}$ – Binding update message processing cost at CN
- $\gamma_g$ – Registration message, location update message or binding update message processing cost at Gateway (Novaczki’s scheme and mHIP scheme)
- $\gamma_{mpr}$ -- Binding update message processing cost at MRP which is not the Gateway (i.e mHIP Router) (Ylitalo’s scheme and mHIP scheme)
- $c_{lu}$ – per hop location update message transmission cost
- $c_{b}$ – per hop binding update message transmission cost

We are going to assess HIP first, and then Novaekiz’s scheme and Ylitalo’s scheme, and finally we move on to mHIP.

**Location Update Cost**

When an MN moves to a new subnet, it needs to update its location details in its location server. An RVS is a location server in the standard HIP architecture. The MN uses HIP Base Exchange to update its information in its RVS. The location update cost includes the cost of this four packets transmission between MN and RVS, and the HIP Base Exchange procession cost in the MN and the RVS. The location update cost per unit time for HIP ($C_{hip}^{fu}$) is

$$C_{hip}^{fu} = \frac{[4(l_{mr} - 1 + \omega)c_{lu} + \gamma_{rvs} + \gamma_m]N_{mn}}{T_B}$$  \hspace{1cm} (8.1)
where \((l_{mr} - 1)\) is the average number of wired hops between AP and RVS.

Ylitalo’s scheme does not introduce any new method to handle the location update, so

\[
C_{lu}^{ys} = C_{lu}^{hip} = \frac{[4(l_{mr} - 1 + \omega) c_{lu} + y_{rs} + y_m]N_{mn}}{T_R} \tag{8.2}
\]

The mHIP scheme and Novaczki’s scheme introduce a local location server (i.e. mHIP Gateway and LRVS) as a network gateway to store the local location of MNs. Instead of updating the record in RVS, an MN updates its local location server. The MN uses HIP UPDATE packets to update its record in the local location server, which means that Address Checking is required, so the location update cost includes the transmission cost of three location update packets and the processing cost at the MN and the local location server. The location update cost per unit time is

\[
C_{lu}^{mhip} = C_{lu}^{ns} = \frac{[3(l_{mg} - 1 + \omega) c_{lu} + y_{g} + y_m]N_{mn}}{T_R} \tag{8.3}
\]

where \((l_{mg} - 1)\) is the average number of wired hops between AP and gateway.

**Binding Update Cost**

When an MN performs a handover in the standard HIP scheme, it uses HIP UPDATE packets to notify its CNs about the change of IP address. The CN requests MN to perform an address checking for the handover. For each MN, the binding update cost includes the transmission cost of three binding update packets and the processing time at MN and CN. Assuming that each CN will have only one connection with MN, the binding update cost per connection is

\[
C_{bu-hip} = 3(l_{mc} - 1 + \omega) c_{bu} + y_c + y_m \tag{8.4}
\]

where \((l_{mc} - 1)\) is the average number of wired hops between AP and CN. The binding update cost per unit time for HIP is

\[
C_{bu-hip}^{hip} = N_{mn} N_{cn} \frac{C_{bu-hip}}{T_R} \tag{8.5}
\]

In the Novaczki’s scheme, the LRVS takes the role of CN, and the binding update cost per connection is

\[
C_{bu-ns} = 3(l_{mg} - 1 + \omega) c_{bu} + y_{mpr} + y_m \tag{8.6}
\]
where \((l_{mg} - 1)\) is the average number of wired hops between AP and LRVS. The binding update cost per unit time for Novaczki’s scheme is

\[
C_{bu}^{nc} = N_{mn} N_{cn} \frac{c_{bu-ns}}{T_R} \tag{8.7}
\]

In Ylialo’s scheme, the nearest MRP in the old path redirects the packet to the previous anchor point, the old anchor point takes the role of CN, and the binding update cost per connection is

\[
C_{bu-ys} = 3(l_{mo} + l_{oa} - 1 + \omega)c_{bu} + \gamma_{mpr} + \gamma_{m} \tag{8.8}
\]

where \((l_{mo} + l_{oa} - 1)\) is the average number of wired hops between AP and the previous anchor point. The binding update cost per unit time for Ylialo’s scheme is

\[
C_{ys}^{bu} = N_{mn} N_{cn} \frac{c_{bu-ys}}{T_R} \tag{8.9}
\]

In the mHIP scheme, the nearest mHIP Agent in the old path takes the role of CN. The binding update cost per connection is

\[
C_{bu-mhip} = 3(l_{mo} - 1 + \omega)c_{bu} + \gamma_{mpr} + \gamma_{m} \tag{8.10}
\]

where \((l_{mo} - 1)\) is the average number of wired hops between AP and the nearest mHIP Agent in the old path. The binding update cost per unit time for mHIP is

\[
C_{mhip}^{bu} = N_{mn} N_{cn} \frac{c_{bu-mhip}}{T_R} \tag{8.11}
\]

**Lookup Cost**

When a CN wants to set up a connection with an MN, it needs to send the I1 to MN’s RVS, and the RVS performs a data look-up in its database for the session. Let \(V_r\) denote RVS lookup cost per second for each communication and \(\psi\) denotes the linear coefficient of the number of MNs to look-up cost. If the number of MNs is linearly related to location database search cost, then

\[
V_r = \psi N_{mn} \tag{8.12}
\]

So, the look-up cost of HIP is

\[
C_{hip}^L = N_{mn} N_{cn} \lambda_s (\psi N_{mn}) \tag{8.13} [115]
\]
As the pre-session mobility of Ylialo’s scheme is the same as that of HIP, so the look-up cost of Ylialo’s scheme is the same as that of HIP

\[ C_{ys}^l = C_{hip}^l = N_{mn} N_{cn} \lambda_s (\psi N_{mn}) \] (8.14)

In the mHIP scheme and Novaczki’s scheme, the RVS forwards the I1 to the local location server to perform the further look-up. We assume that all the MNs in the RVS are also in the same local network, so the look-up cost of the mHIP scheme and Novaczki’s scheme are

\[ C_{mhip}^l = C_{ns}^l = N_{mn} N_{cn} \lambda_s (\psi N_{mn}) + N_{mn} N_{cn} \lambda_s (\psi N_{mn}) = 2 N_{mn} N_{cn} \lambda_s (\psi N_{mn}) \] (8.15)

**Total Signalling Cost**

Based on the equations discussed, the total signalling cost is the sum of location update cost, binding update cost and lookup cost. Summing up Equation 8.1 to Equation 8.15, the total signalling cost of HIP, mHIP, Novaczki’s scheme and Ylitalo’s scheme can be found, which are:

\[ C_{hip}^u = C_{hip}^{bu} + C_{hip}^l \] (8.16)

\[ C_{mhip}^u = C_{mhip}^{bu} + C_{mhip}^l \] (8.17)

\[ C_{ns}^u = C_{ns}^{bu} + C_{ns}^l \] (8.18)

\[ C_{ys}^u = C_{ys}^{bu} + C_{ys}^l \] (8.19)

Assuming that the local network is a balanced tree structure with 10-level depth and using a similar parameter setting as in [114, 115], we are going to compare the performances of different schemes in the following [115]: \( \omega = 10, \lambda_s = 0.001, \psi = 10, c_{bu} = c_{lu} = 0.1, \gamma_m = \gamma_{cn} = 1, \gamma_{rvs} = 30, \gamma_y = 10, \gamma_{src} = 6, l_{ur} = l_{nc} = 40, l_{mg} = 10 \) and \( l_{mo} = l_{oa} = 5.46. \)

We look at the impact of the number of MNs on the total signalling cost. Let \( N_{CN} = 2, T_R = 30s, 60s \) and \( 120s \) and \( N_{MN} \) varies from 10 to 100. Figure 8-13 shows the total signalling cost against the number of MNs in the network. More update signalling packets are required when more MNs exist in the network, so the number of MNs and total signalling cost increase proportionally. Novackzi’s scheme, Ylitalo’s scheme and mHIP all have lower signalling costs. Under the condition of \( T_R = 30s \) and \( N_{mn} = 100 \), their total signalling costs are 38.31%, 8.89% and 51.10% lower compared with HIP, representatively. Figure 8-14 shows the cost improvement...
over HIP of different schemes for various $N_{MN}$ with different $T_R$. As shown in the figure, mHIP is always better than Novackzi’s scheme and Ylitalo’s scheme.

![Figure 8-13 Total signalling cost vs. number of MNs with different $T_R$ value](image)

![Figure 8-14 Signalling cost improvement over HIP vs. number of MNs in different schemes with different $T_R$](image)

Now, we check the impact of average numbers of connections per MN on the total signalling cost. Let $N_{MN} = 80$, $T_R = 30s$, 60s and 120s and $N_{CN}$ varies from 1 to 10. Figure 8-15 shows the result of total signalling against the average number of connections per MN. Similar to the first analysis, more update signalling packets are required when MN communicates with
more CNs, so the average number of connections per MN and total signalling cost increase proportionally. Under the condition of $T_R = 30s$ and $N_{MN} = 5$, Novackzi’s scheme, Ylitalo’s scheme and mHIP’s total signalling cost 22.25%, 14.14% and 42.60% lower compared with HIP, respectively. Figure 8-16 shows the cost improvement over HIP of different schemes for various $N_{CN}$ with different $T_R$. Generally, mHIP is the best. Novackzi’s scheme outperforms Ylitalo’s scheme if the $N_{CN}$ is small. When $N_{CN}$ increases, the difference in the improvement between Novackzi’s scheme and Ylitalo’s scheme is smaller. Under the condition of $T_R = 30s$ and $N_{MN} = 5$, Ylitalo’s scheme provides a better performance than Novackzi’s scheme when $N_{CN}$ is larger than 6.

![Figure 8-15 Total signalling cost vs. Average number of CNs per MN with different $T_R$](image)

![Figure 8-16 Signalling Cost Improvement over HIP vs. Average Number of CNs per MN in different schemes with different $T_R$](image)
The average Residence Time for MN in subnets gives the frequency of the MNs’ carry handover. It also has an impact on the total signalling cost. The lower the average residence time, the more frequent handovers do the MNs have. Let $N_{CN} = 5$ and $N_{MN} = 30$, and 50, the $T_R$ varies from 30s to 300s. Figure 8-17 show the total signalling cost against the residence time. When MNs are highly mobile, the average residence time is small and the corresponding total signalling cost is high. When the average residence time increases, the total signalling cost is decreased. The total signalling costs of Novackzi’s scheme, Ylitalo’s scheme and mHIP tend to be at the same low level. Figure 8-18 shows signalling cost reduction of different schemes over HIP with different $N_{MN}$. Under $N_{MN} = 30$, Ylitalo’s scheme’s steady cost improvement over HIP is around 15% for any $T_R$. The improvement of Novackzi’s scheme and mHIP scheme over HIP decrease as $T_R$ increases. However, mHIP still outperforms other two schemes significantly.

To conclude, mHIP provides shorter propagation delay in pre-session handover and mid-session handover. The signalling cost improvement of mHIP over HIP is also significant. The mHIP is demonstrated as a good HIP-based micro-mobility management protocol.
8.6.3 Security Analysis

As HIP is a secure protocol, the HIP-based micro-mobility management protocol should not introduce any additional security holes to the HIP mechanism. According to [3], the HIP signalling packets (i.e. HIP UPDATE) are protected by HMAC and HIP_SIGNAUTRE. By these cryptographic verifications, forging or replaying a standard HIP UPDATE packet is very difficult unless the HIP security mechanism is broken.

**mHIP Agent Security Analysis**

mHIP Agents need to choose a secure signature scheme which is unforgeable, so that hackers are unable to act as an mHIP Agent to perform a Man-in-Middle (MinM) attack.

An mHIP Router uses the standard HIP Registration process to register itself in an mHIP Gateway. The mHIP Router reuses the HIP digital signature scheme to protect the registration[69].

The mHIP Gateway exchanges parameters, such as the private key in Shared Private Key scheme or Group Signature parameters in Group Signature Scheme, with mHIP Routers. The mHIP Gateway uses the mHIP Routers’ own HI (public key) to encrypt these parameters. HI is a self-certificated public key, unless the hacker knows the corresponding private key, the parameters exchange process is protected by the public key encryption scheme. The hacker is unable to gain the information to act as a mHIP Router to perform a MinM attack.
Security Analysis of Group Signature scheme in mHIP

Unlike Shared Private Key scheme, Group Signature scheme has modified the HIP UPDATE signalling packet. An MN and mHIP Agents in the domain use the group public key to verify HIP signalling packets. The mHIP Gateway only assigns parameters, \( \{y_i, x_i, T_i, ID_i\} \), to a valid mHIP Router based on policies implemented in system. As the early discussion in Chapter 8.3.2, the mHIP Gateway generate \( \{y_i, x_i, T_i, ID_i\} \) based on group private key \( x \) and the value of \( d \). It is difficult to get the group secret key \( x \) from \( y = g^x \mod N \), or find \( d \) with the known \( e \) due to the difficulties of the discrete logarithm problem, or the factoring problem [99]. Without knowing the secure parameter, it is impossible to generate the quarte \( \{y_i, x_i, T_i, ID_i\} \) which are used to generate a valid signature. After assigning \( \{y_i, x_i, T_i, ID_i\} \) parameters, the mHIP Gateway encrypts the Group Signature parameters by the mHIP Router’s own public key (HI). These parameters are encrypted and sent in the HIP Base Exchange.

Moreover, it is unfeasible to obtain the partial secret key \((k, s_i)\) from valid signature parameter \( s \), where \( s = k - s_i c \), because of the intractability of solving the bivariate simple equation [99]. Therefore, if the parameters are exchanged securely, it is impossible for a hacker to generate a valid group signature.

Impersonation attacks

Impersonation attack is attack which misleads the connection from a node to a false peer. MinM attacks can also be considered as an example of impersonation attack[56]. A hacker uses a fake HIP UPDATE packet to mislead mHIP Agents redirecting data packets to a fake location, or carries out a MinM attack between an mHIP Agent and an MN. As mentioned in the previous section, the HIP UPDATE packet is protected by signature; it is difficult to generate a fake HIP UPDATE packet (or any other HIP packet) without knowledge of the secret key associated with HI. Moreover, mHIP Agents request an Address Checking to ensure the update process is requested by the MN and the new location is reachable. Even if an attacker uses a bug in the implementation or security hole in some protocols or systems to redirect a HIP connection, the MN can always reclaim their connection by proving the ownership of the private key associated with their public HI [56]. Furthermore, in our proposed mHIP SPINAT scheme, unlike SEET
mode, the SPI value is also included in the header integrity protection. The hacker cannot change the SPI value to mislead an MN. Based on this discussion, we can say that mHIP offers a secure HIP-based micro-mobility management solution.

8.7 APPLICATION OF MHIP AND FUTURE WORK

We have discussed applying of the HIP in various applications in earlier chapters. Now we will explore how mHIP can be applied to these applications.

8.7.1 mHIP enhanced SHIP

In the traditional SIP-based VoIP environment, IPSec or SRTP-based connection are used to provide a secure VoIP connection. In Chapter 6, a SHIP-based a secure VoIP environment is introduced. SHIP integrates HIP and SIP to provide a secure VoIP environment. As all VoIP media channels are HIP-based in SHIP, so mHIP can be applied to the SHIP network to improve mid-session handover performance. The mHIP Gateway should be integrated, or connected directly with the SHIP Registrar, Proxy and Redirect Server to perform the initial mapping for incoming calls set-up. Moreover, as the SHIP connection uses only an HIP UPDATE packet for the mid-session handover, mHIP Agents can handle the intra-domain handover signalling to reduce the handover delay.

8.7.2 mHIP integrated 3G network

Like HIP, mHIP can be implemented in 3G networks to provide micro-mobility management. With mHIP Agents integrated in the network structure, the wireless network benefits from the improvement offered by mHIP.

In the 3GPP UMTS network architecture, the mHIP Gateway can be integrated with the GGSN. The GGSN with mHIP Gateway is the gateway for UMTS and mHIP networks. The mHIP Router is implemented with the SGSN and RNC in the UMTS network to improve mid-session handover performance. The SGSN and RNC nodes conduct handovers between different Base Stations (Figure 8-19).
Because, in the 3GPP2 CDMA2000 network architecture, most of servers in CDMA2000 are connected to the Router behind the Firewall, the mHIP Gateway can be integrated in that Router. The BSC in the CDMA2000 network should be implemented with mHIP Router to improve the handover performance. Figure 8-20 depicts the mHIP-enhanced CDMA2000 network architecture.

In the heterogeneous wireless network, it recommends applying mHIP in the tight coupling network structure. The mHIP Gateway should be the only interface connecting to the external network of the heterogeneous wireless network. The gateway, e.g. GGSN in the UMTS / WLAN interconnected network, is implemented with the mHIP Gateway. Other network components, which handle the handover in the wireless network, should integrate with mHIP Routers.
8.7.3 Future Work

This newly proposed mHIP scheme also has some areas for future improvement. One is the design of optimal signature schemes for mHIP. An mHIP scheme requires a signature scheme to allow all mHIP Agents to generate a valid signature on behalf of the domain. Although several possible signature schemes were reviewed in this chapter, most of them were not originally designed for the networking management. The Shared Secret Key scheme breaks the fundamental assumption of a public key signature algorithm. Group Signature comes with some extra features which are useless in mHIP. It will produce waste of computing resources. The Ring Signature scheme requests the MN to know the entire signers’ public key, which is not practical. An efficient secure degraded Group Signature scheme can improve the mHIP scheme in order to reduce some unnecessary parameters and computing resources. A Ring Signature with the Group Signature style operation will be ideal for the mHIP. The new signature scheme should
provide unforgeable, anonymous, unlinkable and untraceable security features. Receivers should be able to verify the signature by a common public key and a group manager is needed in the new scheme.

Currently, there are no network simulation packages for simulating HIP or mHIP operations. The development of an HIP/mHIP simulation program is a large project in itself. The implementation of mHIP relies on a mature HIP platform and also requires the development of the HIP/mHIP kernel for different operation systems. This can lead to an interesting open source project.

8.8 SUMMARY

A novel HIP-based micro-mobility management protocol, mHIP is proposed in this chapter. mHIP improves intra-domain handover performance significantly while still maintaining the security level of HIP. It also introduces many new ideas for micro-mobility management.

mHIP Agents are introduced as mHIP enhanced middle boxes to handle HIP-based intra-domain handover signalling. They also redirect HIP-based data packets to the update locations of MNs. They share a common “Net HI/HIT” which is used to represent the whole domain. All mHIP Agents can generate a valid signature on behalf of others. MNs or all other network components can use the Net HI to verify signatures generated by mHIP Agents. Different signature schemes are reviewed in this chapter. The means of implementing these signature schemes into mHIP are proposed in this chapter, even though they were not originally designed for networking management.

Our mHIP scheme also considers HIP data packet routing and an overlay routing of mHIP is introduced. A new modified conceptual HIP header for HIP data packet is proposed. Moreover, we have also proposed a new mHIP SPINAT scheme, which is a modification of the current SPINAT scheme, for the HIP ESP scheme. Additionally, the HIP SRTP scheme is also considered and this has not been discussed in the literature. A new SRTPNAT is proposed.
It has been shown that mHIP performance is better than other studies of HIP-related micro-mobility management in both propagation delay and signalling cost. The applications of mHIP for SHIP and 3G wireless communication are also outlined in this chapter.

Generally speaking, our mHIP scheme offers a new solution that provides a better intra-domain handover performance for future HIP-based network.
Chapter 9: Conclusions

9.1 CONTRIBUTION

In this dissertation, the mobility issue in traditional IP networks has been addressed. The big problem is due to the overloading of IP addresses as the network locator and the endpoint identifier. The public-key based IP network (HIP) tries to solve this problem by introducing an extra layer in the OSI 7 Layer model. However, because of the introduction of a new layer and a new namespace, the public-key based IP network cannot be adopted easily as all network requirements would need updating.

In the evolution from the traditional IP networks to the Public-key based IP networks, the kernel of end nodes need to be re-developed. Application programming interfaces (API) and the kernel of the operating system need to be updated for HIP support, especially those IPv4 API. In order to simplify the evolution, we have proposed a step-stone solution – Secure Mobile IP with HIP style Handshaking and Re-addressing, which merges some HIP features into Mobile IP. This proposed scheme has replaced the Mobile IP RR process and binding update using the HIP style re-addressing method. As it still uses an IP address as the endpoint identifier, the Home Address is still an endpoint identifier in the networks. This can improve the security level of Mobile IP and improve the handover performance with a minor update to the operating system kernel and API. This step-stone solution allows the network to change easily from the traditional IP networks into the public-key based IP networks. Results can be also found in [J2, C3].

One of the challenges in adopting HIP-based networks is in wireless communication networks. Even though the handover performance of HIP is better than Mobile IP, the current commercial wireless networks lack support for HIP. Moreover, the HIP protocol needs to co-operate with other protocols in order to achieve a better performance. In Chapter 5, four different HIP enhanced wireless communication network architectures have been proposed. They are the HIP-based GPRS network architecture, the HIP-based UMTS Release 4 architecture, the HIP-
based UMTS Release 5 network architecture and the HIP-based CDMA2000 network architecture. Compared with the homogeneous wireless networks, heterogeneous wireless networks are going to be popular in the future. We have proposed a new, HIP-based heterogeneous wireless network architecture in Chapter 5. A tight coupling between UMTS and Wireless LAN network is used to demonstrate the HIP-based handover in the heterogeneous wireless networks. The structure can be further expanded into different types of wireless network interconnections. In Chapter 6, we have also proposed a new hybrid mobility scheme for the VoIP network. SIP protocol plays an important role in the VoIP network signalling; however, the performance of mobility management of SIP is not good as it is a plain text application layer protocol. In order to improve the performance, a hybrid SIP-HIP protocol – SHIP – is proposed in Chapter 6. SHIP can provide the overall mobility management that SIP cannot provide. SHIP also enhances the security in VoIP naturally. The handover performance of SHIP is better than SIP in many different scenarios. These results are also reported in [J1, C2, C4]

Micro-mobility management focuses the mobility within a domain. Most micro-mobility management protocols are Mobile IP based. A HIP-based micro-mobility management is necessary if HIP is to be chosen as a macro-mobility management protocol. In Chapter 7, we have established the requirements for the HIP-based micro-mobility management protocol. We have also presented the analysis of two proposals from other authors. An improved version of one of these schemes has been suggested in that chapter. In Chapter 8, we have proposed our own scheme – mHIP, which is based on the set requirements. mHIP is the first complete HIP-based micro-mobility management protocol. In mIP, newly introduced middle boxes, mHIP Agents, form a virtual server. All the mHIP Agents can generate a valid signature on behalf of the domain. The mHIP framework allows different signature schemes to be applied. In mHIP, the load of signalling handling is distributed in the whole local network. An mHIP Agent is a 3.5 layer router. We have proposed a 3.5 layer overlay routing, which can be applied to different types of HIP connections, such as HIP IPSec or HIP SRTP. Compared with the other schemes discussed in Chapter 7, mHIP has better intra-domain handover performance. It outperforms
other schemes in terms of propagation delay and signalling cost. These results are also found in [C1].

The research and the result presented in this dissertation have contributed to the application and implementation of HIP protocol in mobility management for wireless IP networks. Our work has covered HIP-based mobility management at both the macro and micro levels. We are considering the proposal of a possible IETF Internet Draft based on our result.

9.2 FUTURE WORK

The work reported in this dissertation also opens up some issues for future research. One is that the study of the HIP-based heterogeneous wireless networks can be further extended to cover the Data Link Layer signalling mechanism to provide a fast handover in HIP. The second one is a further study on the impact of SHIP/HIP on the VoIP networks. The SHIP proposal has not yet been tried on commercial VoIP networks.

Another issue is that of the signature schemes for mHIP. Those suggested in the dissertation all come with some disadvantages. For example, the Share Private Key scheme breaks the fundamental rule of the public and private key scheme; the Group Signature scheme is too big for the mHIP, and the Ring Signature is not efficient in mHIP. An interesting topic would be an efficient signature scheme requiring a group of entities to sign without leaking their own identities. The new signature scheme should be unforgeable, anonymous, unlinkable and untraceable. The receivers can verify the signature by a common public key. A downgraded version of Group Signature might be able to provide the solution.

The development of an HIP simulation tool is another interesting project as no simulation program on HIP networks exists at this moment. The tool would be very useful in future study of the behaviour of HIP and mHIP networks.

The development of the mHIP kernel is an important step for mHIP implementation, requiring a lot of man power, and could lead to a valuable open source project.
Bibliography

[8] H. Soliman, C. Castelluccia, K. E. Malki and L. Bellier, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)," IETF RFC 4140, August 2005
[14] 3GPP, "TS 22.060: General Packet Radio Service (GPRS), Service Description, Stage 2."


[23] 3GPP, "TS 23.002: Network Architecture (Release 5)."


[25] 3GPP, "Internet Protocol (IP) multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3."


[49] 3GPP2, "All-IP Core Network Multimedia Domain: IP Multimedia Call Control Protocol Based on SIP and SDP Stage 3".


[73] T. Henderson, "Generalizing the HIP base protocol," draft-henderson-hip-generalize-00 (work in process), Internet Draft, IETF, 13 February 2005

[74] 3GPP, "TS 23.002: Network Architecture (Release 1999)."


Appendices

Appendix A: Publication lists

Journal


Conference


Other Publication unrelated to this dissertation