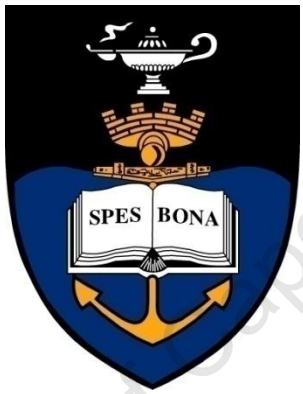
Distributed IP Mobility Management for Hosts and Networks

Petro Pesha Ernest



This thesis is submitted in fulfilment of the academic requirements
for the degree of
Doctor of Philosophy in Electrical Engineering
in the Faculty of Engineering and The Built Environment
University of Cape Town
March 2014

As the candidate's supervisor, I have approved this dissertation for submission.

Name: Professor H. Anthony Chan

Signed by candidate

Signed: Signature Removed

Date: 3 March 2014

Name: Dr Olabisi E. Falowo

Signed: _____

Date: _____

Declaration

I hereby declare that: (1) The above thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance of my supervisor, I have received no assistance apart from that stated below; (2) except as stated below, neither the substance or any part of the thesis, has been submitted in the past, or is being, or is to be submitted for a degree in the University, or any other University.

I am now presenting the thesis for examination for the Degree of PhD in Electrical Engineering. I also grant the University free license to reproduce the above thesis in whole or in part, for the purpose of research.

Petro Pesha ERNEST	March 01, 2014
Name	Date

To my lovely wife (Beatrice Pesha) and sons (Remigius Kabyemela and Patrick Mugisha)

Abstract

The Internet was originally designed for stationary nodes. With the advancement of mobile nodes (such as smartphones and tablets) that have wireless Internet access capability, the original design of the Internet is no longer sufficient. These mobile nodes are capable of communicating while moving and changing their point of attachment in the Internet. To maintain communication session(s) continuity for these mobile nodes, the Internet needs mobility management mechanisms.

The main mobility management protocols standardised by the Internet Engineering Task Force (IETF) are mobile IP (MIPv6 and MIPv4) and their numerous extensions and variants, including proxy MIP (PMIPv6 and PMIPv4). The architectural structures of these protocols employ a centralized mobility anchor to manage the mobility of the mobile nodes in the control and data planes. The mobility anchor manages the mobility binding information and the forwarding of data packets for all mobile nodes registered in the network. However, in the context of the rapid growth in the number of mobile users and the data traffic volume, as well as the trend towards a flat architecture in mobile networks, the centralized mobility management approach provides insufficient mobility support to the mobile nodes.

For example, to manage the demand for increased mobile users, a huge amount of data traffic will be pushed to the centralized mobility anchor. Yet, routing huge volumes of traffic via the centralized mobility anchor can be non-optimal in terms of routing efficiency. Thus, the centralised mobility anchor can be a potential bottleneck, and a single point of failure. Consequently, failure of the mobility anchor may lead to a service outage for a large number of mobile nodes. Ultimately, the centralized mobility management approach does not scale well with the increase in number of mobile users and the data traffic volume. These problems are also costly to resolve within the centralized mobility management approach and its related centralized network architecture.

Distributed mobility management (DMM) is one recent approach that can efficiently address the shortcomings of centralized mobility management. It provides an alternative paradigm for developing IP mobility management – without employing centralized mobility anchors. In this paradigm, either the mobility anchors, or their mobility management functions,

are distributed to different networks/elements. The mobility anchors, or the mobility management functions, are brought to the edge of the networks, which is closer to the mobile nodes. Distributed mobility management also offers dynamic mobility features that allow a mobile node to anchor traffic at different mobility anchors. However, to date, mobility management schemes that have been developed based on the DMM approach are still in the preliminary stages, and there is no current standard in place. These developed DMM schemes are still experiencing problems, such as long routing paths, especially for long-lasting data traffic, a lack of route optimization for ongoing communication, and a lack of synchronization of the mobile nodes' location in different networks. Moreover, the majority of these proposed schemes still need to be analysed, in order to quantify their feasibility.

The thesis proposes three novel network-based distributed mobility management schemes, which are based on the DMM approach. The schemes enhance PMIPv6 to work in a distributed manner, in order to address the problems of centralized mobility management. Furthermore, the schemes address the following issues: (1) the lack of route optimization for ongoing communication; (2) the lack of synchronization of the mobile nodes' location in different networks; and (3) the long end-to-end packet delivery delay problems in recently proposed DMM schemes.

The first scheme, called the network-based distributed mobility management scheme with routing management function at the gateways (DM-RMG), decomposes the logical mobility management functions of the Local Mobility Anchor (LMA) in PMIPv6 into internetwork location management (LM), routing management (RM), and home network prefix allocation (HNP) functions. After the decomposition, the RM function is collocated at the gateways of different networks. In this way, the data-plane routing function of the respective mobile nodes is served by the corresponding local RM function at the network gateway.

The DM-RMG scheme offers distributed mobility management for individual mobile nodes (i.e., mobile hosts) during mobility events. DM-RMG also implements a mechanism to optimize the handover delay. The results obtained from analytical modelling and simulation show that the DM-RMG scheme outperforms the centralized mobility management schemes, as well as currently proposed distributed mobility management schemes in terms of the end-to-end packet delivery delay under different network load conditions. The optimized handover

performance of the DM-RMG scheme, investigated under different traffic patterns and mobile node speeds, shows that the scheme also mitigates the internetwork handover delay and packet loss.

The second proposed scheme, called network-based distributed mobility management for the network mobility (NDM-RMG), uses a similar approach to DM-RMG. However, it proposes a network-based DMM scheme for Network Mobility (NEMO). The main goal of the NDM-RMG scheme is to address the problems of centralized mobility management protocols for NEMO, including the pinball routing problem in nested NEMO. NDM-RMG is compared with centralized mobility management schemes for NEMO, and recently proposed distributed IP mobility management schemes for NEMO by means of analytical modelling and simulation evaluations. NDM-RMG shows better performance in terms of reducing the packet delivery latency, the size of the packet header, and the packet overhead experienced over the wireless link.

The third proposed scheme, called network-based distributed mobility management scheme with RM and HNP allocation functions distributed to the access routers (DM-RMA), distributes the RM and the HNP allocation functions at the access routers with the mobility client function. This brings the mobility-related functions closer to the mobile nodes, that is, to the edge of the network. An analytical model is developed to investigate the mobility cost performance of the scheme, due to signalling, packet delivery, and tunnelling. The analytical results indicate that DM-RMA performs better than the previous DMM schemes in terms of packet delivery, tunnelling and total costs.

Network simulator-2 (ns-2) is used to model the DM-RMA scheme. The simulated scenarios confirm that DM-RMA performs better than other proposed DMM schemes in terms of reducing the location update latency at the location managers, end-to-end packet delivery delay, handover delay, and packet loss.

In addition to the three proposed DMM schemes, this thesis proposes a routing optimization scheme for PMIPv6. The main goal of this scheme is to enable PMIPv6 to offer route optimization to mobile nodes in a PMIPv6 domain. The scheme reduces the route optimization-establishment latency, the packet delivery latency, and the packet loss. Using ns-2 simulations and considering different simulated scenarios, the results show that the scheme

reduces route optimization-establishment latency and delayed packets during the route optimization operation, as compared to previously proposed PMIPv6 route optimization schemes. The results also show that the scheme reduces packet loss when a mobile node undergoes handover in the PMIPv6 domain.

Acknowledgements

First and foremost, I thank GOD for the gift of life and his blessing.

Second, I wish to express my sincere gratitude to my supervisors Prof. H. Anthony Chan and Dr Olabisi E. Falowo for their encouragement, consistent guidance, and support during this research work.

Third, I would like to thank the management team of the Dar es Salaam Institute of Technology (DIT), Tanzania, for allowing me to pursue this PhD study, and for the financial support to accomplish this research project.

Fourth, I am deeply grateful to Mr Neco Ventura and the University of Cape Town, Postgraduate Funding Office, for their financial support to attend academic conferences.

Fifth, I would like to thank Dr Linoh A. Magagula and my colleagues at the Communication Research Group (CRG), University of Cape Town, for the discussion on my work and their friendship. Thanks to Dr Linoh for the constructive comments on my thesis chapters.

Finally, I wish to express my deepest gratitude to my family for the unconditional support and love. I thank my wife and my sons for their valuable prayers and patience during all these years: I love you all. Thanks must also go to my father and mother for their enduring prayers and support. My thanks also extend to my brothers and sisters for their ongoing support throughout the years.

Table of Contents

Distributed IP Mobility Management for Hosts and Networks i

Declaration iii

Abstract v

Acknowledgements ix

Table of Contents x

List of Figures xiv

List of Tables xviii

Publications xix

Glossary xxi

Chapter 1 Introduction 1

1.1 The Evolution of Mobility Management resulting from the Trends in Mobile Data Traffic and Mobile Network Architecture 1

1.2 Problem Statement 5

1.3 Research Objectives and Contributions to Effective Distributed Mobility Management for Host and Network Mobility Support in IP Networks 6

1.3.1 Thesis Objectives 6

1.3.2 Thesis Contributions 7

1.4 Scope of the Research 9

1.5 Thesis Outline 10

Chapter 2 IP Mobility Management Background: Centralized and Distributed Approaches 12

2.1 Overview of IP Mobility Management Protocols and Related Terms 13

2.1.1 Mobility Management: Definition, Types, and Concepts 14

2.1.2 Classification of IP Mobility Management 15

2.2 Host-based IP Mobility Management Protocols 18

2.2.1 Overview 18

2.2.2 Protocol Descriptions 19

2.2.3 MIPv6 Limitations 20

2.2.4 MIPv6 Extension 21

2.3 Network-based Mobility Management Protocol: PMIPv6 22
2.3.1 Overview 22
2.3.2 PMIPv6 Description 23
2.3.3 PMIPv6 Enhancements and the Limitations 26
2.3.4 Comparison of Host-based and Network-based Mobility Management 27
2.4 Distributed Mobility Management Protocols 28
2.4.1 Overview 28
2.4.2 Protocol Descriptions 29
2.4.3 Application Scenario for Distributed Mobility Management 32
2.4.4 Methods of Distributing Mobility Functions 33
2.4.5 Host-based Distributed Mobility Management 34
2.4.6 Network-based Distributed Mobility Management 37
2.5 Comparison of Distributed Mobility Management Schemes 39
2.6 Comparison of Centralized and Distributed Mobility Management Schemes 40
2.7 Summary 41
Chapter 3 Network-based DMM with Routing Management Function at the Gateway Routers: DM-RMG 43
3.1 Introduction 43
3.2 Motivation and Design Approaches for DM-RMG 43
3.3 The DM-RMG Architecture and Operation Mechanism 46
3.3.1 MN Registration Procedure and Data Flow after Registration 48
3.3.2 Handover Procedure to Visited Network 50
3.3.3 Data Flow after Handover and Route Optimization Procedures 51
3.4 DM-RMG Extension with TMAG to Support Seamless Handover 53
3.4.1 Introduction and Motivation 53
3.4.2 Handover Mechanism with TMAG 56
3.5 Simulation Environment and Scenarios 58
3.5.1 Network Simulator Version 2 Overview 58
3.5.2 Simulation Scenarios 59
3.6 Simulation, Analytical, and Performance Evaluations 62
3.6.1 Analytical Evaluation for End-to-end Delay 62
3.6.2 Simulation Results and Performance Analysis for the End-to-end Delay 64
3.6.3 Analytical Performance Evaluation for the Handover Delay 69
3.6.4 Simulation Results and Performance Analysis for Handover Latency and Packet loss 70
3.7 Comparative Qualitative Analysis of DM-RMG with Other DMM Schemes 74

3.8 Summary **75**

Chapter 4 Network-based Distributed Mobility Management for Network Mobility: NDM-RMG **76**

4.1 Introduction **76**

4.2 Motivation and Design Approaches for NDM-RMG **76**

4.3 Related Work **78**

4.4 Architecture of the Proposed NDM-RMG Scheme **82**

4.4.1 Architecture Overview *82*

4.4.2 Mobile Router Registration and Prefix Acquisition Procedures *84*

4.4.3 Handover Mechanism for a Non-Nested NDM-RMG Scenario *85*

4.4.4 Handover Mechanism for a Nested NDM-RMG Scenario *88*

4.5 Analytical Performance Evaluation **89**

4.5.1 Packet Overhead Analysis *90*

4.5.2 End-to-end Packet Delay Analysis *91*

4.5.3 Packet Delivery Cost Analysis *93*

4.5.4 Binding Update Cost Analysis *94*

4.5.5 Numerical Results and Discussion *94*

4.6 Simulation Evaluation in ns-2 ..... **97**

4.6.1 Simulation Scenario in ns-2 ..... *97*

4.6.2 Simulation Results and Analysis *98*

4.7 Summary **100**

Chapter 5 Network-based DMM with Distributed Routing Management at Access Routers: DM-RMA **101**

5.1 Introduction and Motivation **101**

5.2 Related Work **102**

5.3 DM-RMA Overview and Operation Mechanism **104**

5.3.1 Initial MN Registration and Communication Establishment *105*

5.3.2 Handover to another Network *106*

5.3.3 Handovers within the Same Network *108*

5.4 Performance Evaluation **108**

5.4.1 Network Model *109*

5.4.2 Mobility and Traffic Models *110*

5.4.3 Total Cost *111*

5.4.4 Signalling Cost *111*

5.4.5 Packet Delivery Cost *114*

5.4.6 Packet Tunnelling Cost (C_{TC}) *115*
5.5 Numerical Results and Discussion **115**
5.6 Simulation Evaluation in ns-2 ..... **121**
5.6.1 Simulation Scenarios *122*
5.6.2 Simulation Results and Analysis *124*
5.7 Summary **128**
Chapter 6 Route Optimization Mechanism for Proxy MIPv6 using CMAG **129**
6.1 Introduction and Motivation **129**
6.2 Related Work **130**
6.3 Architecture of PMIPv6 with CMAG **131**
6.3.1 Architecture Overview *132*
6.3.2 The Mechanism of Operation of PMIPv6 with CMAG *133*
6.4 Performance Evaluation **136**
6.4.1 Simulation Scenario in ns-2 ..... *136*
6.4.2 Simulation Results and Analysis *138*
6.5 Summary **145**
Chapter 7 Conclusion and Future Work **146**
7.1 Conclusion **146**
7.2 Future Work **148**
References **150**

List of Figures

Figure 1-1 Cisco forecasts on mobile data traffic growth [4] 2

Figure 2-1 Mobility management in MIPv6 without route optimization 20

Figure 2-2 Mobility management in PMIPv6: an overview 24

Figure 2-3 An example of a partially DMM approach 30

Figure 2-4 An overview of a fully DMM approach 31

Figure 2-5 Distributed mobility management application scenarios 32

Figure 3-1 DM-RMG functional architecture 46

Figure 3-2 MN registration signalling flow of DM-RMG 49

Figure 3-3 Data packet flow when the MN is in home network 50

Figure 3-4 Signalling call-flow diagram when an MN moves to a visited network 51

Figure 3-5 Data flows from a CN to an MN before route optimization and after route optimization 52

Figure 3-6 MN performing subsequent handover from Net2 to Net4 53

Figure 3-7 Signalling call-flow when the MN performs handover from Net2 to Net4 54

Figure 3-8 TMAG in overlapping region shared by GW2/RM2 and GW4/RM4 networks 56

Figure 3-9 Signalling call-flow when the MN performs a handover from Net2 to Net4 networks with TMAG configured in the overlapping region between these networks 57

Figure 3-10 Simulated network topology 60

Figure 3-11 The sub-optimal and the optimal paths that the packets follow from the CN to the MN 63

Figure 3-12 Comparison of end-to-end delay between the optimal and sub-optimal paths with change in the background traffic load, with three intermediate routers on the path to MN's communicating IP address-anchoring network 65

Figure 3-13 The influence of the increased distance between the MN's communication 66

Figure 3-14 The impact of the distance between the MN's communicating IP address anchoring network and MN's visited network on the end-to-end delay of the optimal and sub-optimal paths 67

Figure 3-15 Comparison of packet delivery latency before and after the MN's handovers to a visited network between the optimal path and the sub-optimal path 68

Figure 3-16 Impacts of traffic patterns on the handover delay 71

Figure 3-17 Impacts of traffic patterns on packet loss 72

Figure 3-18 Impacts of MN's speed on the handover delay 73

Figure 3-19 Impacts of MN's speed on packet loss 73

Figure 4-1 NEMO Basic Support Protocol Operation 79

Figure 4-2 Pinball routing in Nested NEMO Basic Support Protocol 80

Figure 4-3NDM-RMG architecture with RM co-located at the gateways for non-nested NEMO 83

Figure 4-4 Mobile router and MNN attachment and registration in the NDM-RMG domain 85

Figure 4-5 Mobile network handover signalling call flow from Net1 (RM1) to Net2 (RM2) for non-nested NDM-RMG 86

Figure 4-6A NDM-RMG framework with RM co-located at the GWs for nested NEMO 88

Figure 4-7 End-to-end delay with different levels of nesting 95

Figure 4-8 Packet delivery cost with different levels of nesting 96

Figure 4-9 Binding update cost with different levels of nesting 96

Figure 4-10 Simulated nested NEMO topology 97

Figure 4-11 The impact of the number of levels nesting on Packet delivery delay 98

Figure 4-12 End-to-end delay, according to the distance between HAs/RMs 99

Figure 4-13 Measured packet size in the wireless link 100

Figure 5-1 DM-RMA architecture 104

Figure 5-2 Initial attachment and session establishment procedures 106

Figure 5-3 Handover signalling call flow and new session setup 107

Figure 5-4 A network model used for cost modelling 109

Figure 5-5 The variation of signalling cost with MN speed 117

Figure 5-6 The impact of domain size on signalling cost 118

Figure 5-7 The effect of average session length on packet delivery cost 118

Figure 5-8 The impact of average session length on packet tunnelling cost 119

Figure 5-9 The impact of the probability that the traffic is hand off traffic on packet delivery cost 120

Figure 5-10 The impact of the probability that the traffic is hand off traffic on packet tunnelling cost 120

Figure 5-11 The impact of SMR on total cost 121

Figure 5-12 Topology used for simulation evaluation 122

Figure 5-13 Packet delivery latency before, during and after handover 126

Figure 5-14 The impact of distance between domains on packet delivery latency 127

Figure 5-15 Inter-domain handover delay and packet loss 128

Figure 6-1 Architecture for PMIPv6 with CMAG 132

Figure 6-2 Signalling flow for MN registration and establishment of the optimized routing path for PMIPv6 with CMAG 134

Figure 6-3 Handover procedures and optimal routing path updates when the MN performs handover signalling flow of the proposed scheme 135

Figure 6-4 Network topology for simulation 137

Figure 6-5 The impact of latency between MAGs and LMA over the route optimization-

establishment latency and the amount of delayed data packets 139

Figure 6-6 The impacts of various delays between MAGs and LMA on average end-to-end delay 140

Figure 6-7 Influence of data traffic transmission rates over average number of delayed data packets during the routing path optimization procedure 141

Figure 6-8 The impacts of various delays between MAGs and LMA on handover delay and packet loss 142

Figure 6-9 The impact of various latencies between MAGs on handover delay and packet loss 144

List of Tables

Table 3-1 Summary of the statistics of the end-to-end delay 67

Table 4-1 Parameter notations and values [75][76] 91

Table 5-1 Parameters used for numerical results 116

Table 5-2 Parameters for simulation evaluation 123

Table 5-3 Summary of intra-domain handover delay and packet loss 124

Table 5-4 Summary of location update latencies 125

Publications

Selected peer-reviewed papers published from this thesis are documented in the following:

1. Petro P. Ernest, Olabisi E. Falowo and H. Anthony Chan, "Enhance Distributed Mobility Management Schemes for NGWNs," Proceedings of Southern African Telecommunication Networks and Applications Conference, East London, South Africa, 4-7 September, 2011.
2. Petro P. Ernest and H. Anthony Chan, " Enhanced Handover Support and Routing Path Optimization with Distributed Mobility Management in Flattened Wireless Networks," Proceedings of the 14th International Symposium on Wireless Personal Multimedia Communications (WPMC) Workshop on Mobility Management for Flat Networks (MMFN 2011) , Brest, France, 3-6 October, 2011.
3. Petro P. Ernest, H. Anthony Chan, and Olabisi E. Falowo, "Distributed Data Path and Mobility Function Scheme for PMIPv6 in Flattened Networks," Proceedings of Southern Africa Telecommunication Networks and Applications Conference, George, South Africa, 2-5 September, 2012.
4. Petro P. Ernest, H. Anthony Chan, and Olabisi E. Falowo, "Distributed Mobility Management Scheme with Mobility Routing Function at the Gateways," Proceedings of IEEE GLOBECOM, Anaheim, California, USA, 3-7 December, 2012, pp. 5476-5481.
5. Petro P. Ernest, Olabisi E. Falowo, H. Anthony Chan, and Linoh A. Magagula, "Fast Route Optimization Considering Mitigating Packet Loss for Proxy MIPv6 with Coordinating MAG," Proceedings of Southern Africa Telecommunication Networks and Applications Conference, Stellenbosch, South Africa, 1-4 September, 2013 (received best paper award).
6. Petro P. Ernest, Olabisi E. Falowo, and H. Anthony Chan, "Network-based Distributed Mobility Management: Design and Analysis," Proceedings of the 9th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Lyon, France, 7-9 October , 2013, pp. 516-523.

7. Petro P. Ernest, H. Anthony Chan, Olabisi E. Falowo, Linoh A. Magagula, and Sandra Céspedes, "Network-based Distributed Mobility Management for Network Mobility," Proceedings of the 11th IEEE CCNC, Las Vegas, Nevada, USA, 10-13 January, 2014, pp. 708-716.

8. Petro P. Ernest, H. Anthony Chan, Jiang Xie, and Olabisi E. Falowo, "Mobility Management with Distributed Mobility Routing Function," Springer Telecommunication Systems, Special Issue on Mobility Management for Flat Networks, Accepted for publication.

9. Petro P. Ernest, Olabisi E. Falowo, H. Anthony Chan, "Design and Performance Evaluation of Distributed Mobility Management Schemes for Network Mobility," Under Review, Elsevier Journal of Network and Computer Applications.

Glossary

AAA	Authentication, Authorization, and Accounting
AP	Access Point
AR	Access Router
BA	Binding Acknowledgement
BRep	Binding Reply, which is the message delivered from CMAG to MAGs
BReq	Binding Request, which is the message delivered from MAGs to CMAG
BU	Binding Update
BuC	Binding Update Cost
BUL	Binding Update List
CBR	Constant Bit Rate
CF	Control Function
CMAG	Coordinating Mobile Access Gateway
CN	Correspondent Node
CoA	Care-of-Address
D-PMIP	Distributed PMIP
DF-PMIP	Fully Distributed PMIP
DHCPv6	Dynamic Host Configuration Protocol for IPv6
DHT	Distributed Hash Table
DMM	Distributed Mobility Management
DM-RMA	Network-based DMM with Routing Management at the ARs with mobility client function
DM-RMG	Network-based DMM with Routing Management at the Gateways
DP-PMIP	Partially Distributed PMIP

DR	Delegating Router
DS-MIPv6	Dual Stack Mobile IP version 6
EPC	Evolved Packet Core
EPS	Evolved Packet System
FMIPv6	Fast Mobile IP version 6
GW	Gateway Router
HA	Home Agent
HAWAII	Handoff-Aware Wireless Access Internet Infrastructure
HIP	Host Identity Protocol
HMIPv6	Hierarchical MIPv6
HNP	Home Network Prefix
HoA	Home Address
ICMD	Inter-domain Central Mobility Database
ID	Identifier
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPv6	Internet Protocol Version 6
LIPA	Local IP Address
LM	Internetwork Location Management
LMA	Local Mobility Anchor
LR-PMIPv6	Localised Routing for PMIPv6
LTE	Long Term Evolution
MAG	Mobile Access Gateway
MAP	Mobile Anchor Point

MIPv6	Mobile IP version 6
MLT	MAG List Table
MN	Mobile Node
MN-HoA	MN Home Address
MN-ID	MN Identifier
MNN	Mobile Network Node
MNP	Mobile Network Prefix
MR	Mobile Router
NIST	National Institute of Standards and Technology
ns-2	Network Simulator Version 2
NDM-RMG	Network-based DMM with Routing Management at the Gateways for NEMO
NEMO	Network Mobility
Net	Network
NIQ	Node Information Query
N-DMM	NEMO-based DMM
OTcl	Object-oriented Tool Control Language
PBA	Proxy BA
PBQ	Proxy Binding Query
PBU	Proxy BU
PDC	Packet Delivery Cost
PMIPv6	Proxy Mobile IP version 6
P-CMAG	PMIPv6 with CMAG
PoA	Point of Attachment

PQA	Proxy Binding Query Acknowledgement
PSTN	Public Switched Telephone Network
RA	Router Advertisement
RM	Routing Management
RO	Route Optimization
RR	Requesting Router
RRP	Return Routability Procedure
SAE	System Architecture Evolution
SCTP	Stream Control Transmission Protocol
SIP	Session Initiation Protocol
SIPTO	Selective IP Traffic Offload
SMR	Session-to-Mobility Ratio
S-PMIP	Signal Driven PMIP
TCP	Transmission Control Protocol
TMAG	Tracking Mobile Access Gateway
UDP	User Datagram Protocol
VNI	Visual Networking Index
VoIP	Voice over Internet Protocol
WiFi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
3G	Third generation mobile communication networks

Chapter 1 Introduction

1.1 The Evolution of Mobility Management resulting from the Trends in Mobile Data Traffic and Mobile Network Architecture

Nowadays, access to the Internet services is no longer bound to a specific geographical or topological location. Instead, users can wirelessly access these services anywhere, and anytime. Motivated by this flexibility, the number of users accessing Internet services while on the move is rapidly increasing. This is producing data traffic volume at an exponential pace. This increase in both the number of users and data traffic volume is further fuelled by other reasons [1][2][3] that include:

(i) The availability of various wireless communication technologies, such as third generation (3G) and wireless local area network (WLAN), which are becoming affordable to most users and offer wide coverage;

(ii) The technological advancement of mobile devices which are 3G and WLAN capable, and the availability and affordability of these devices, (such as laptops, notebooks, smartphones, and tablets);

(iii) The affordability of 3G USB modems;

(iv) The invention and development in Smartphone applications, which can access multiple Internet services concurrently, and for extended duration; and

(v) The development of mobile Internet services, such as video streaming, voice over Internet Protocol (VoIP) and audio streaming, which are accessible via mobile web-browsing.

As a result of increases in both mobile users and data traffic volume, the mobile network is currently heavily loaded with mobile data traffic. And the data growth is continuing; and furthermore, it is expected to grow by 11.2 exabytes per month by 2017: a 13-fold increase over 2012 [4], as shown in Figure 1-1. Unfortunately, the mobile network is usually configured with hierarchical and centralised management architecture [5]. Therefore, it is not able to easily cope

with these increases. This is due to scalability and reliability problems that are costly to resolve within the centralized architecture [1].

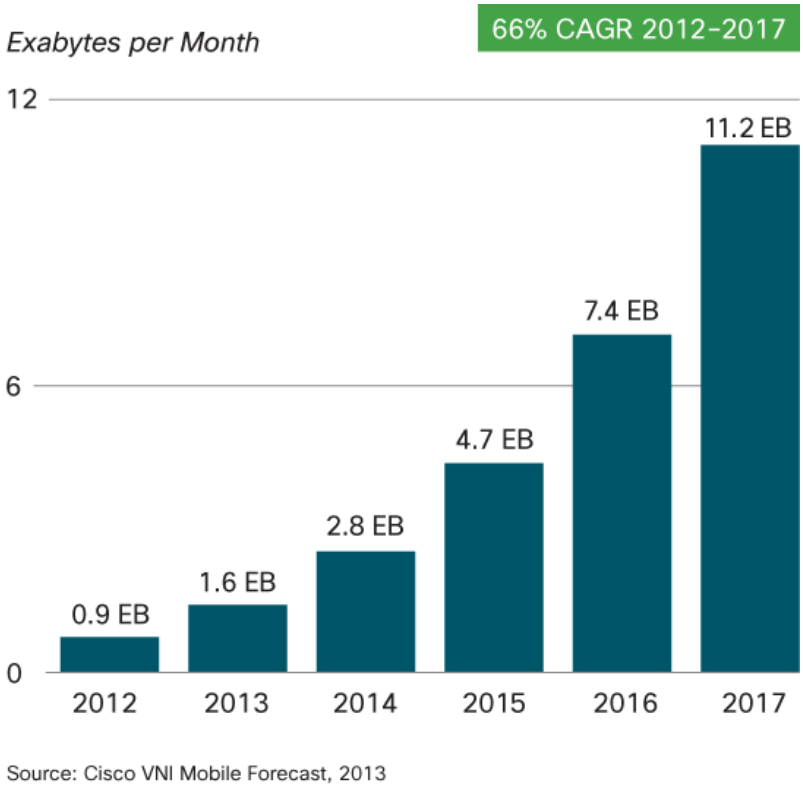


Figure 1-1 Cisco forecasts on mobile data traffic growth [4]

In order to accommodate the data traffic growth, and to relieve the traffic load from the mobile core network, the network operators have currently adopted mechanisms, such as:

(i) Expanding the network capacity: deploying dense cells like femto- and picocells [6][7] and spectrum efficient technologies, i.e., 3GPP Long Term Evolution (LTE) and WiMAX;

(ii) Traffic off loading: adopting the off loading of mobile data traffic through WiFi networks [8][9] and using selective traffic off load mechanisms, such as selective IP traffic off load (SIPTO) and Local IP address (LIPA) [10]; and

(iii) Upgrading the equipment of the core network nodes.

However, although expanding the network capacity increases the throughput of the radio access, it may lead to the development of new usages, and hence fuel the traffic growth. And while the off loading techniques remove the traffic overload from the mobile core network, they

offer limited mobility support within small localized regions [11]. Upgrading the core network is an easy approach, which seems technically and technologically possible. But mobile operators' average revenue per user is getting lower; and such an approach is not an economically feasible solution [12][1]. All these solutions are effective only in some scenarios; and they cannot effectively address the mobility demands in the mobile Internet.

To further address the problems caused by the hierarchical and centralized architecture of the networks, the mobile network is currently evolving to a flat architecture. That is to say, it is adopting fewer levels of network hierarchy. This is intended to reduce operation costs and facilitate traffic off loading [1][13]. In flat networks the network entities may be distributed in such a way that there is no central entity to manage mobility. Hence, mobility management that is suitably adapted to flat and distributed architecture is important [14].

At the same time, the mobile architecture, for example, third Generation Partnership Project (3GPP) Evolved Packet System (EPS) [15] and WiMAX [16], is evolving to an all-IP network for both voice and data communications. This implies the need for the adoption of IP-based mobility management protocols to support mobility for a moving node. For example, EPS has adopted proxy mobile IPv6 (PMIPv6) [17] and dual stack mobile IPv6 (DSMIPv6) [18] to provide mobility support for the moving nodes. These protocols are IP-based mobility management protocols standardized by the Internet Engineering Task Force (IETF) [19].

Unfortunately, current IP-based mobility management protocols rely on a central mobility anchor to manage mobility. Examples of the mobility anchors include Home Agent (HA) in Mobile IPv6 (MIPv6)) [20] and Local Mobility Anchor (LMA) in proxy MIPv6 (PMIPv6) [17].

The mobility anchor is usually deployed in the core network, which for many cases means it is located far away from the mobile nodes – that is to say, far away from the edge of the network. The anchor maintains the binding of mobility context information for all mobile nodes (MNs) registered in the network. So, the data traffic for MNs needs to traverse the anchor, irrespective of the point of attachment of the specific MN. Moreover, with the current growth of mobile users and the increase in data traffic, a huge amount of data traffic will be pushed to the core network – for the sake of mobility management. The routing via the central anchor may be non-optimal; and this can negatively impact the routing efficiency.

Additionally, the anchor manages the mobility signalling for all MNs. Consequently, as the number of mobile nodes increases and the data traffic volume grows, the mobility anchor could find it difficult to efficiently handle such increases – due to scalability and reliability problems related to the centralised route and the mobility context management [14]. Furthermore, the centralized mobility management approaches cannot satisfactorily support the mobility in flat network architecture. This is due to limitations that include a single point of failure, traffic bottleneck, non-optimal routing path, scalability problems and long handover [14].

To effectively support mobile users, centralised IP mobility management need to be redesigned and/or enhanced, in order to meet the current growth in the number of mobile users and the increase in the mobile data traffic volumes, as well as the trends towards flat mobile network architecture. An alternative mobility management design for new mobility management, and/or the enhancement of the existing one, is needed.

In order to overcome the shortcomings caused by the deployment of a centralized mobility anchor in mobility management, and to effectively support mobile users, the IETF has proposed a new paradigm for IP mobility management development – Distributed Mobility Management (DMM) [21]. The main concept behind DMM is to provide mobility support without relying on a centrally deployed mobility anchor. The mobility anchor as a whole, and/or mobility management functions, are distributed to different networks/elements, resulting in the mobility anchor or the mobility management functions being brought closer to the MNs, for example, closer to the edge of the network.

Moreover, the traffic is anchored at different mobility anchors, so that the scalability and routing inefficiency issues are reduced. Furthermore, a single point of failure is eliminated. Also, the mobility support may be activated or deactivated, depending on whether the MN's session needs mobility support, or not [22]. This also reduces the amount of state information that must be maintained in various mobility agents of the mobile network.

DMM is a current paradigm for developing IP mobility management protocols, within both the research community and the IETF. It has gained significant attention because of the fact that existing IP mobility management protocols cannot efficiently withstand the trends described above (i.e., growth in the number of mobile users and increase in the data traffic volume, as well as the evolution of mobile network to flat and an all-IP network). As the way forward, the IETF

has formed the DMM working group [21] to look at the possibility of extending the existing IP mobility protocols to work in DMM scenarios that fulfil the DMM requirements [14].

To date, there is no standardized protocol for DMM, but various proposals are available from the IETF and the research community. These proposals will be discussed later in this thesis. Most of the proposed protocols still have some limitations; and only a few of the proposed protocols have been analysed to determine their feasibility. These limitations include: long routing path resulting from the MN's communication remaining anchored at the MN's communicating IP address anchor point, the lack of route optimization for ongoing communication, the lack of synchronization of the MN location information in different networks, and long end-to-end packet delivery delay – especially for long duration traffic.

1.2 Problem Statement

Existing IP mobility management protocols (e.g., MIPv6, PMIPv6, and NEMO Basic Support) offer centralized mobility support for mobile nodes or mobile routers. This centralized approach introduces a central mobility anchor, which is in charge of routing management, location management, and home address (or home network prefix) allocation functions for each registered mobile node (or mobile router). However, this centralized approach, having all of the logical functions bundled in a single network element, and considering the current increase in the number of mobile users and the increased mobile data traffic volume, has performance drawbacks, such as low scalability, traffic bottleneck, and single point of failure.

Furthermore, it requires routing of the data traffic from/to the mobile nodes through the mobility anchor, which causes non-optimal routing and creates a significant delay on packets, especially when the two communicating nodes are close to each other, but are far from the mobility anchor. The major problem here is how to overcome the drawbacks of centralized mobility management.

Mobility management schemes developed based on the distributed mobility management (DMM) approach can overcome these limitations. The distributed mobility approach distributes the mobility anchors or the mobility management functions to different networks/entities, such that the mobile node is served by the nearest mobility anchor or function. DMM can be achieved by using either the partially distributed approach, in which only the data-plane is distributed, or

by the fully distributed approach, in which both the data- and the control-planes are distributed.

A number of mobility management schemes that extend the centralised IP mobility management schemes to operate in a DMM manner have recently been proposed in the literature. However, most of the proposed schemes still incur long routing path and long end-to-end packet delivery delay – especially for long duration traffic. Moreover, in fully distributed DMM schemes, where the control plane is distributed, there is still the challenge of how the new network could discover the old network, from which the MN has been detached.

Thus, there is a need for further research, in order to enhance the centralised IP mobility management protocols to operate in a distributed manner by incorporating additional mobility functionalities, in order to address some of the problems present in the existing centralized mobility management schemes and the recently proposed DMM schemes.

Therefore, this thesis addresses some of the problems mentioned above by developing new IP-based distributed mobility management schemes for mobile hosts and mobile networks.

In addition, the thesis covers the route optimization problem in standard PMIPv6, and develops a new route optimization scheme that extends PMIPv6 protocol to perform route optimization for an MN that is roaming in a PMIPv6 domain.

1.3 Research Objectives and Contributions to Effective Distributed Mobility Management for Host and Network Mobility Support in IP Networks

1.3.1 Thesis Objectives

Distributed mobility management is a new paradigm for developing mobility management without relying on a central mobility anchor for flat and distributed IP networks. Consequently, the main objective of this thesis is to develop IP mobility management schemes to provide distributed mobility management for mobile nodes (mobile hosts) and mobile networks, following the concept of DMM; and then to examine the developed schemes and study their performance through analytical evaluation and simulations. More specifically, the objectives are summarized as follows:

a) To review the existing IP mobility management schemes in terms of mobility function deployment;
b) To critically analyze the most referenced distributed mobility management schemes;
c) To employ PMIPv6 to develop network-based IP mobility management schemes, in order to provide distributed mobility support for moving hosts and networks. The schemes will address the shortcomings in the existing centralized and distributed mobility management schemes, which include reducing end-to-end packet delivery latency, triangular routing problem, packet overhead over the wireless link, and packet delivery cost;
d) To evaluate the performance of the developed mobility management schemes, and to compare the performance improvement with various distributed and centralized mobility management schemes – by using analytical evaluation and simulation methods;
e) To develop a route optimization mechanism for PMIPv6 that mitigates triangular routing, route optimization-establishment latency, and packet loss.

1.3.2 Thesis Contributions

The following summarises the major contributions of this thesis. Some of these contributions are already documented in the author's peer-reviewed publications (as listed under publications).

1) A review of the centralised IP mobility management protocols, both standardized host-based and network-based, in terms of the deployment of centralised mobility anchors for mobility management, is presented. This includes a discussion of the limitations of these protocols regarding their ability to cope with the rapid growth in both the number of mobile users and the increase in the data traffic volume. In addition, a comprehensive review of the distributed mobility management (DMM) approaches is given. The approaches are classified and qualitatively compared.

2) A network-based DMM scheme that releases the load burden of the central mobility anchor (in centralised IP mobility management schemes) by splitting the logical mobility management functions of the LMA in PMIPv6, and co-locating the routing management (RM) function with the gateway routers in different networks is proposed. This scheme optimizes the data routing path of the MN that has moved from its communicating IP

address-anchoring network to a visited network. The proposed scheme is named network-based DMM with RM function at the gateways (DM-RMG) [23]. The DM-RMG scheme is further extended by introducing an entity named tracking mobile access gateway (TMAG), in order to reduce the handover delay and packet loss when the mobile node moves between visited networks. The packet delivery latency, the handover delay, and the packet loss performances of the DM-RMG scheme are evaluated by means of analytical modelling and simulations conducted in ns-2. Furthermore, a qualitative study of this scheme – in comparison with the most referenced DMM schemes – is conducted.

3) A network-based DMM scheme for non-nested and nested NEMO scenarios, named network-based DMM for network mobility (NDM-RMG), is developed. The NDM-RMG scheme is built on the concept of decomposing the LMA logical functions in PMIPv6 and co-locating the routing management (RM) function with the gateway routers in different networks to enable distributed mobility support for mobile networks. The NDM-RMG scheme mitigates the pinball routing problem and the high packet overhead in standard NEMO [24], particularly in the nested NEMO scenario. The analytical functions are developed for performance evaluation of the NDM-RMG scheme in comparison with the NEMO Basic Support protocol and other related distributed NEMO schemes, in terms of packet overhead, end-to-end delay, packet delivery cost, and binding update cost. Simulations are conducted, using ns-2, in order to evaluate the performance of the NDM-RMG scheme in terms of packet delivery latency and packet overhead.

4) A network-based DMM scheme that splits the logical mobility management functions of the LMA in PMIPv6, and co-locates the RM and home network prefix (HNP) allocation functions to distributed access routers with a mobility client function is developed. The developed scheme provides an optimal path for both old and newly established sessions of the MN after handover, and releases the load burden from the central mobility anchor, such as LMA. Moreover, the scheme alleviates the need for tunnelling between RMs and access routers of the DM-RMG scheme. Additional mobility functions, such as tracking, updating, and node information query are introduced at the access routers to allow them to track the MN's movement and learn the MN preconfigured mobility information. The proposed scheme is named network-based DMM with RM and HNP allocation functions distributed to the access routers (DM-RMA). An analytical model is developed for the

performance evaluation of the DM-RMA scheme in comparison with other similar distributed mobility schemes in the literature. The analytical model is used to evaluate the impact of network topology, session length, session-to-mobility ratio, and the probability of hand off traffic on packet delivery cost, tunnelling cost, signalling cost, and total cost. Furthermore, simulations are carried out in ns-2, in order to evaluate the impact of decomposing and distributing the logical mobility management functions on intra- and internetwork handover delay, packet loss, location update latency, and packet delivery latency.

5) An additional mechanism, which extends the PMIPv6, to support route optimization in the PMIPv6 domain is developed, and named PMIPv6 with coordinating MAG (PMIPv6 with CMAG) [25]. PMIPv6 with CMAG reduces the long routing path caused by triangular routing, route optimization-establishment latency, and packet loss in the PMIPv6 domain. The mechanism strategically utilizes the MAG, which is located in the shortest path to other MAGs in the PMIPv6 domain, in order to offer support for route optimization queries. Simulations are conducted in ns-2, in order to evaluate the performance of the scheme. The simulations study the influence of the CMAG location with respect to neighbouring MAGs as well as LMA on delayed packets, end-to-end delay, route optimization-establishment latency, and packet loss.

1.4 Scope of the Research

This research focuses on enhancing the centralized IP mobility management protocols to operate in a distributed manner, in particular, network-based mobility management protocol, PMIPv6. The research proposes schemes to release the load burden of the mobility anchor in centralized IP mobility management protocols. Moreover, the research addresses the non-optimal path and the long packet delivery, particularly the long duration traffic of ongoing communication during MN handovers, present in the recently proposed DMM schemes. The performances of the proposed schemes are evaluated by means of analytical modelling and discrete event simulations against centralized IP mobility management schemes and other related IP mobility management schemes developed by using the distributed mobility management concept.

The scope of the work considered in the research is network-layer mobility management

solutions.

1.5 Thesis Outline

This thesis is structured as follows. Chapter 2 provides an overview of IP mobility management, focusing on the mobility definition, types, and classifications. The layer 3 mobility management protocols standardized at the IETF are presented for both host-based and network-based mobility protocols. Following an overview of DMM, focusing on the basic principles of DMM (application and classification), a wide range of the most referenced DMM protocols are presented and classified. A qualitative comparison of the discussed mobility approaches is then provided. The chapter concludes with a motivation to enhance the existing IP mobility management protocol, in order to provide DMM support.

Chapter 3 describes the development of the newly distributed mobility management protocol, DM-RMG. This chapter presents the motivation for the design of this protocol, design considerations and operational mechanisms, as well as the protocol enhancements to provide a seamless handover. The simulation studies and the implementation in network simulator version 2 (ns-2) are presented, along with analytical and qualitative analyses in comparison with other DMM schemes, are given in this chapter.

Chapter 4 presents the development of a new distributed mobility management protocol, NDM-RMG. The chapter discusses the motivation for the design of this protocol, design considerations and operation mechanisms (both nested and non-nested NEMO), as well as the protocol enhancement to mitigate the pinball routing in NEMO basic support. The developed analytical functions to evaluate this protocol are also provided. NDM-RMG is compared with NEMO Basic Support protocol, and one other DMM scheme developed for NEMO. Additionally, simulation evaluation using ns-2 is discussed in this chapter.

Chapter 5 discusses the development of a new distributed mobility management protocol, DM-RMA. The chapter presents the motivation for the design of this protocol, design considerations and operational mechanisms. A mathematical model developed to evaluate the protocol is also presented in this chapter. DM-RMA is then evaluated and compared with another related DMM scheme, using the mathematical model. The simulation implementation, results and discussion are also provided in this chapter.

Chapter 6 describes the development of a new mobility management protocol that extends PMIPv6 to support route optimization, PMIPv6 with CMAG. The chapter presents the motivation of the protocol design, design considerations and operational mechanisms. PMIPv6 with CMAG is evaluated through simulation, and compared with PMIPv6 and two other well-known route optimization mechanisms developed based on PMIPv6. The simulation results are also presented.

Chapter 7 concludes the thesis by summarizing the results obtained and discussing some proposals for future work.

Chapter 2 IP Mobility Management Background: Centralized and Distributed Approaches

The original motivation for Internet design considered facilitating communication between stationary nodes [26]. However, with the development of wireless access technologies and the invention of portable devices, such as laptops, smartphones, and tablets, nodes are becoming mobile and capable of accessing Internet services while on the move, i.e., changing locations in the Internet. Thus, mobility management is essential for enabling mobile nodes to continue accessing Internet services while on the move.

However, it is very challenging to support mobility in the Internet – due to the fact that nodes use an Internet Protocol (IP) address for communication. The IP address is used to identify a node, as well as the location where the node is currently attached to the network. Such an approach basically simplifies the Internet routing mechanism. That is, if the identifier of the node is known, its location can automatically be derived to allow the routing mechanism to reach the node. However, in a scenario where the node changes its network location, its IP address also changes; so that the address reflects the current location of the network to which the node is attached. This also causes a change to the node identifier, since it is coupled to this IP address. Consequently, any active data session(s) on the mobile node breaks, because the node's identifier is among the quadruple parameters that the higher layers (i.e., transport and application) use to identify a session.

In order to maintain active session continuity, the IP address needs to be preserved, irrespective of change mobile node location in the network. Thus, it becomes very challenging to maintain active session(s) of the mobile node using the fundamental concepts of the Internet design. This is because the routing mechanism cannot route packets to the new location of the mobile node if the IP address is kept unchanged. Mobility management protocols provide a primary means to address this problem, while maintaining session continuity independent of the node's location changes (i.e., change of access router) in the Internet and/or change of the access media technology.

This chapter discusses the concept of mobility management and mobility management protocols in IP networks standardized by the IETF. It discusses the deployment limitations of the

existing IP mobility management protocols in coping with the rapid increase in both the number of mobile users, and the increase in data traffic volumes. Moreover, the chapter examines distributed mobility management as an alternative approach to develop IP mobility management protocols in overcoming these limitations. Efforts in developing distributed mobility management protocols in the research community and the IETF, along with the protocols' strengths and weaknesses are also discussed. Finally, a qualitative comparison of centralized mobility managements and distributed mobility managements is also provided.

2.1 Overview of IP Mobility Management Protocols and Related Terms

As discussed at the beginning of this chapter, IP mobility management protocols are important mechanisms to ensure session continuity for the node on the move. The protocols comprise three key elements [20]: (i) The Mobile Node (MN), which is an end-user mobile device for which mobility support is provided; (ii) a mobility anchor point, which is the entity in the network responsible for providing the mobility management service to the MN; and (iii) the correspondent node (CN), which is a node that is involved in active communication with the MN. Moreover, the protocols use the mechanism of separating the two roles of the IP address – a locator and an identifier of a node – in order to provide mobility support for the MN. They achieve this by equipping the MN with two IP addresses, namely: a home address and a care-of address [20].

The home address (HoA) serves as the identifier of an MN and does not change irrespective of the MN movement. The care-of address (CoA) serves as the locator to indicate the actual point of attachment of the MN to the Internet. It is the one used by mobility agents to route packets to the MN. The mobility anchor is the special router, which resides in the MN's home network. It provides to an MN the HoA(s); and it stores the binding information between the HoA and the CoA of the MN. This binding information is what allows the reachability of the MN, while away from its home network.

The MN uses the HoA to communicate with CN(s). Consequently, the packets from the CN(s) are first routed to the MN's mobility anchor. The mobility anchor gets the MN's binding information and uses the CoA to tunnel packets directly to the MN, or to the mobility entity that

is able to deliver them to the MN. So, the MN communications are preserved, regardless of its location change, i.e., CoA change.

2.1.1 Mobility Management: Definition, Types, and Concepts

Mobility in wireless networks refers to movement of the mobile node, which involves a change of point of attachment to the Internet while maintaining the active communication of the mobile node. This type of mobility is referred to as node mobility (or host mobility) [27]. Mobility can also be defined in terms of personal mobility and session mobility [28], as well as network mobility [24]. Personal mobility refers to the mobility that allows a person to maintain his/her reachability, irrespective of change of point of attachments, or of devices used in the network. The person continues accessing his/her subscribed network services using a single identifier, i.e., the same name or address.

Session mobility covers mobility that involves a network session migration from one networking node to another, without interrupting the ongoing session. For instance, users having more than one session on a particular node can move one of these sessions to another node. Network mobility allows a moving network/subnet to change the point of attachment to the networks – with uninterrupted communication of its communicating nodes. These mobility types serve different types of mobility requirements; and each type has different properties to mobility management.

This thesis focuses on mobility management for the node and the network, particularly at the network layer. In the following section, node mobility is discussed. A discussion on network mobility is presented in Chapter 4.

As described in the above paragraph, mobility involves nodes changing their point of attachment in the network; this is referred to as handover or handoff [20]. The handover includes terminating the existing connectivity and getting a new one, which involves either a change of access point at the link layer level only, or a change of IP point of attachment, subnetwork. The former is referred to as link layer mobility; while the latter is called IP layer mobility, also known as network layer mobility. During the handover duration, the MN is unable to receive from or send packets to its CN(s). The duration hereof is known as handover latency. The handover disrupts the ongoing communication, because the packets sent to the mobile nodes

during handover get lost, unless mechanisms to prevent packets from being lost are employed. This negatively affects the quality of communications. Mobility management protocols handle the mobility of the moving node – with the aim of reducing handover latency and packet loss.

Mobility management is a mechanism that maintains communication continuity for a mobile node; while the mobile node changes points of attachment in the network. It is comprising two main components: location management and handover management [26][29]. Location management plays the role of locating and tracking the actual location of the mobile node in the network, thereby guaranteeing the reachability of the mobile node, independent of its movements.

The location management information is periodically updated by the mobile node itself, or by a mobility agent, whenever the mobile node changes location in the network. This allows for up-to-date mobile node location information. Handover management is responsible for maintaining active sessions; while the mobile node moves and changes its point of attachment in the network.

2.1.2 Classification of IP Mobility Management

This subsection classifies IP mobility management; and it then provides a brief explanation of each category. IP mobility management can be classified in different ways. In this thesis, the mobility management is classified based on the layers of Transmission Control Protocol/Internet Protocol (TCP/IP) stack reference model implementing the mobility support [30], the network element implementing mobility support, the scope of the operation of the mobility protocol, and the mobility management –based on the implementation architecture.

a) Mobility management based on layers

The IP mobility management protocols, under the TCP/IP protocol stack reference model, are classified as the link layer, the network layer, the transport layer, and the application layer mobility support, respectively.

Mobility management at the link layer provides mobility support for mobile nodes that change access points only within a range of access router/subnet. It facilitates setting up of the wireless radio connection, when the mobile nodes change the access point. The mobility support

provided at this layer is specific to wireless access technology (e.g., 3G, WLAN, LTE, etc.), that is, a mobile node employing a particular link layer technology cannot be supported when it moves across different technologies. Moreover, this mobility support is limited within the scope of the access router. As a result, when the mobile node movements spans access points that are under different access routers, the mobility support from the higher layers (e.g., network, transport, and application) are involved in facilitating this handover.

Network layer mobility provides mobility support to mobile nodes that change their IP point of attachment/subnet in the network. It is built on the principles of separating the roles of the IP address of being node identifier and locator. Mobile nodes configure two IP addresses: one acting as an identifier, and the other as a locator, routing address. A mobility anchor is employed to maintain the mapping between these two addresses, and to intercept packets destined to the mobile node identifier. Thus, this mobility management approach is also referred to as anchor-based mobility management. Mobile IPv6 [20] and its variants are examples of the anchor-based network layer mobility management protocols. The network layer is considered suitable for implementing mobile node mobility, because mobility involves a change of the node's IP address. So, mobility is seen as an address translation problem [31], which is suitably tackled by using the network layer mobility support approaches. Besides, this layer exists in all Internet nodes and the mobility support at this layer is transparent to the higher layer protocols [12].

Mobility at the transport layer uses end-to-end techniques for mobility, where the end nodes take care of their mobility without involving any third party. To facilitate this, both communicating end nodes need to have mobility support capability. Nonetheless, the mobility support at this layer is transport-specific. Examples of transport layer mobility protocols include Stream Control Transmission Protocol (SCTP) [32] and MSOCKS [33], just to mention a few examples.

Mobility management at the application layer is supported by specific applications, using end-to-end signalling. It does not need any modification to the IP stack of the MNs for mobility support purposes. An example of such an approach is Session Initiation Protocol (SIP) [34]. The approach is supported by a SIP server, where the MN registers its new location, i.e., a new IP address, when it moves to a new point of attachment. Similarly, the CNs are notified about the new IP address. To benefit from this mobility support, applications need to support SIP.

The above discussion suggests that mobility at different layers has different properties and imposes different mobility management requirements. This thesis, therefore, focuses on mobility management at the network layer, which is one of the important solutions for addressing Internet mobility. In particular, in this we consider anchor-based mobility solutions at the network layer. This includes Mobile IPv6 and its extensions and variants, such as Proxy MIPv6 [17]. There are also routing-based mobility management protocols at the network layer, such as HAWAII [35], Cellular IP [36], but these will not be discussed in this thesis.

Similarly, other mobility management protocols, such as Host Identity Protocol (HIP) [37] are not discussed. In the following discussion only the IETF anchor-based network layer mobility management approaches are considered.

b) Mobility management based on element implementing mobility support

Mobility management functions at the network layer can reside either in both the network and the mobile node or in the network only. The former is referred to as host-based; while the latter is called network-based mobility management. In host-based mobility management, the mobile node implements mobility functions and plays a role in handling the signalling related to the mobility management operation; whereas in network-based mobility management, the network entities take care of the mobility management without involving the mobile node. Sections 2.2 and 2.3 discuss in detail these classes of mobility management.

c) Mobility management based on the operation scope

Mobility management at the network layer can be considered based on the movement scope of the mobile node [29]. The mobility management that handles the mobility of the mobile node across network domains is called macro-mobility management, in other words, mobility that spans a large area. The mobility management which takes care of the movement of the mobile node between access routers under the same network domain is referred to as micro-mobility management – mobility over a small area.

d) Mobility management based on implementation architecture

Mobility management at the network layer can also be categorized based on how the mobility management logical functions of the mobility anchor are handled in the network. In a network architecture where all the mobility management logical functions are kept on a single

network entity, the mobility management approach is referred to as centralized mobility management. Centralized mobility management employs a mobility anchor, which is centralized and maintains the MNs' mobility context, intercepts packets for the MNs, and allocates HoA or HNP to the MNs.

On the other hand, the mobility anchors and mobility management logical functions may be distributed to multiple locations of wireless and mobile networks. With such an approach, the mobility management is called distributed mobility management (DMM) [12][21]. The mobility management functions are then available in multiple network entities, and the mobile nodes are served by the closest network entity.

These two categories are employed either as host-based or network-based mobility managements. We discuss in detail these categories of mobility management in Section 2.2, Section 2.3 and Section 2.4, respectively.

2.2 Host-based IP Mobility Management Protocols

2.2.1 Overview

The IP host-based mobility management protocols enable mobile nodes to remain reachable, while roaming in the Internet, with the mobile nodes being involved in managing their mobility. The mobile nodes are, therefore, aware of their mobility; and they need to perform mobility management operations, in order to maintain their reachability. Because of their participation in mobility management, these protocols require a modification of the mobile nodes' IP stack, so that they can handle mobility-related signalling. Examples of the host-based IP mobility management protocols include Mobile IPv6 and its extensions, such as Fast MIPv6 [38], Hierarchical MIPv6 [39], Dual stack MIPv6 [18].

The basic principles of operations of the host-based mobility management protocols in IPv6 networks are derived from MIPv6. In the following subsection, therefore, these principles and the mobility challenges of these protocols, in regard to deploying a centralized mobility anchor to handle mobility for all mobile nodes, are discussed, in relation to MIPv6. The goal is to review these mobility management protocols, and to show their limitations in dealing with the current ever increasing number of mobile nodes and the increased data traffic volume. The efforts toward enhancing MIPv6 to overcome such limitations are presented, and the extension of

MIPv6 with distributed mobility management to solve these limitations is discussed in Section 2.4.

2.2.2 Protocol Descriptions

Mobile IPv6 provides mobility support to mobile nodes, which is transparent to the higher layers (i.e. transport and application layers), while the mobile node is roaming across the IPv6 networks. The protocol uses the principle of distinguishing the two roles of IP address (identifier and locator) to provide mobility support for the mobile node. As discussed previously, the mobile node is, therefore, supplied with two IP addresses: the HoA and CoA. The HoA serves as a stable end point identifier to identify a mobile node; and it is obtained from a mobility anchor in the mobile node's home network.

As an identifier, the HoA remains unchanged, irrespective of where the mobile node roams in the Internet. The CoA is the temporary address the MN obtains from the network that is not its home network; and this is known as the visited network. This address serves as the locator to indicate the actual point of attachment of the MN to the Internet. So, it changes each time the MN moves to a different network. Moreover, the mobility agents use the CoA to route packets to the MN's current location.

The protocol introduces a special router in the MN's home network; this is called a Home Agent (HA). The HA allocates the HoA to MNs, and maintains the binding information between the HoA and the CoA of the MNs, allowing for the reachability of the MNs while away from their home networks. To achieve this, the binding information needs to be up-to-date. So, each time an MN changes a visited network, the MN needs to exchange binding update (BU) and binding acknowledgement (BA) messages, with its HA to update its binding information.

The binding update messages enable the HA to know the current location of the MN. The HA updates/creates the binding entry for the MN in a special table known as the binding cache; and it establishes an IP bidirectional tunnel between itself and the MN. Thereafter, the packets from CN(s) to the MN are intercepted by this HA, which then uses its binding cache entry to locate the MN (i.e., MN's CoA) and then encapsulates the packets in an IPv6-in-IPv6 tunnel with the destination address of the MN's CoA. Figure 2-1 summarises the basic principles of operation of MIPv6.

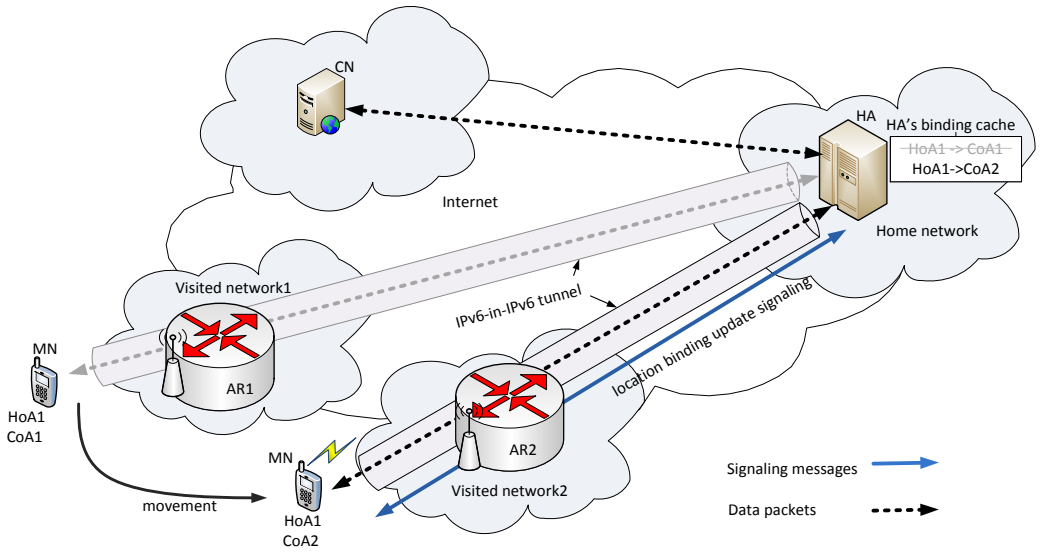


Figure 2-1 Mobility management in MIPv6 without route optimization

As shown in Figure 2-1, the HA maintains the binding information about the MN's current location in the binding cache; it intercepts all packets sent from/to the CN, and encapsulates/decapsulates them in IPv6-in-IPv6 tunnel to/from the MN, while the MN is moving around in the visited networks. The MN configures a new CoA, when it changes the visited network; and it informs these changes to its HA through binding updates signalling. The HA then updates the MN's entry in the binding cache. The CN communicates with the MN, using the HoA, regardless of the visited network that the MN is currently visiting. This hides the MN movement from the CN. All the MNs' data use a tunnel between the MN and the HA.

2.2.3 MIPv6 Limitations

In MIPv6, the MN communicates with its CNs by using its HoA, which is topologically anchored to its HA. As a result, all the packets sent to MN are firstly sent to HA, and then tunnelled to the MN. The packets, therefore, encounter a sub-optimal path, called triangular routing [40]. The triangular routing causes a substantial communication delay, especially when the MN moves topologically far away from its home network; and this results in the inefficient use of the resources in the backhaul of the communication network [40][22][14].

In order to overcome the triangular routing problems, the basic specification of MIPv6 protocol specifies the route optimization; and this is called the Return Routability Procedure (RRP) [20]. The RRP allows the MN to inform its CN(s) of its current location; a direct routing

path between the MN and the CN(s) is then established; and the data packets are then exchanged through this path, thereby bypassing the HA. This reduces the communication delay and the data traffic load on the HA. However, the MN reveals its temporary identifier, CoA, to the CN(s), which infringes the location privacy of the MN [41]. Furthermore, to make the route optimization feasible, the CN(s) must support the RRP; but it is impossible to have all the IPv6 nodes employing RRP. Consequently, legacy IPv6 nodes cannot utilize the benefit of route optimization; and all the data traffic for these nodes will traverse the HA, irrespective of the route optimization.

Consequently, the specified route optimization is limited only to scenarios, where the CN(s) supports the RRP. The RRP, however, still requires the involvement of HA for security reasons; and that further makes the HA a single point of failure. Moreover, it causes a significant signalling load, especially when a huge number of MNs are communicating with various CN(s), and opt to perform route optimization.

The MNs binding information in MIPv6 are stored on the HA, but considering the anticipated growth in the number of mobile nodes and the traffic volume [4], the HA cannot scale well with such increase; as it needs to maintain mobility context, and it manages the tunnel for large numbers of MNs. Moreover, the HA is a single point of failure. This is because if the HA is unavailable, the communication for all MNs is interrupted. Furthermore, if the route optimization is favoured, the RRP cannot be accomplished.

Besides these limitations, the MIPv6 requires a modification of the MN's IP stack, in order for the MN to support its own mobility, which increases complexity of the MN. Moreover, the signalling exchanges and the data tunnelling over the wireless link do not promise well in terms of spectrum usage, given the current rapid growth in the number of MNs. This is because the wireless spectrum is limited, and cannot be increased [3]; and hence, it needs to be used with great care.

2.2.4 MIPv6 Extension

The host-based mobility management protocol defined in IETF has been Mobile IPv6. Numerous extensions have been standardized in IETF, in order to enhance the performance of MIPv6 in solving some of the mentioned limitations. To enhance the performance of MIPv6 in

terms of signalling overhead to the HA, Hierarchical MIPv6 (HMIPv6) [39] was developed. HMIPv6 introduces a Mobility Anchor Point (MAP), a local HA, which limits the signalling exchange between the MAP and the MN, while the MN moves within the MAP domain. This approach brings the anchor (MAP) close to the MN, which reduces binding update delay, signalling overhead to the HA, and the handover latency. Nevertheless, HMIPv6 still requires modification of the MN IP stack, and the packets for legacy IPv6 nodes still encounter triangular routing. Moreover, HMIPv6 still inherits the HA centralization problems that include a single point of failure.

Another improvement to MIPv6 is the Mobile IPv6 Fast handover (FMIPv6) [38] designed to mitigate the handover delay and packet loss due to MIPv6 handover procedures. FMIPv6 uses predictive and reactive handover mechanisms. The predictive mechanism enables the MN to set up IP connectivity prior to actual handover and establishment of the tunnel between the previous and new access routers of the MN. This approach reduces/removes the time needed for IP CoA configuration; hence, it reduces the handover delay, and the tunnel created further reduces the packet loss. Yet, the MN IP stack modification is still needed. Moreover, the HA remains a traffic anchoring point and a manager for all MNs' binding information.

Both approaches, HMIPv6 and FMIPv6, require support from the MNs in mobility management operation. As a result, these extensions still make the MNs complex and difficult to deploy, because all the existing IPv6 nodes need change in their IP stack. Therefore, these protocols still inherit most of the limitations of MIPv6.

2.3 Network-based Mobility Management Protocol: PMIPv6

2.3.1 Overview

The need for protocol stack modification to all the existing mobile nodes in host-based IP mobility management protocols has been a barrier to deploying these protocols in the communication markets. The network-based IP mobility management protocol tackles this problem, in addition to reducing handover delay and packet loss. The protocol removes the MIPv6 functions residing in the MNs, and puts them within the network entities. The network entities handle mobility signalling on behalf of the MN. This alleviates the involvement of the

MN in exchanging mobility-related signalling with the HA, in order to facilitate the handover procedure. Accordingly, the need for MN's IP stack change or for implementing special IP stack software is removed; and the MNs' complexity and power consumptions are thereby reduced. This also enables regular IPv6 nodes, which are MIPv6-incapable to benefit from mobility support.

The current network-based IP mobility management standardized by IETF is PMIPv6 [17]; and it is discussed in detail in the following sub-sections.

2.3.2 PMIPv6 Description

Proxy MIPv6 is built from MIPv6 to provide mobility support for MNs – without requiring their participation in mobility management and IP signalling. It provides mobility support to all the MNs, irrespective of whether the mobile node implements MIPv6 functionality, or not. The protocol introduces network functional elements, namely: Local Mobility Anchor (LMA) and Mobile Access Gateways (MAGs) [17], which handle the mobility of the MN. The LMA is an HA in the PMIPv6 domain; and it runs on a domain gateway – whereas the MAG is the logical function running in the access router of the MN.

The MAGs are responsible for tracking the MNs movements, (i.e., their attachment and detachment events) on the access link, and for performing the signalling operation on their behalf.

In PMIPv6, the LMA assigns a unique HNP to each MN, which registers in the PMIPv6 domain. The MN uses the assigned HNP to configure an IP address, (an HoA), which remains the same, while it roams within the domain [42], because the MN continues receiving the router advertisement (RA) from MAGs containing the same HNP. So, the MN does not need to configure a new IP address when it roams in a PMIPv6 domain. This reduces the time needed to perform IP configuration, and hence reduces the handover delay. The HoA is used by an MN in all of its communications – as the session identifier.

When the MN moves from the initially attached link to a new link, the MAG located in the new link authenticates the MN through an Authentication, Authorization and Accounting (AAA) server. Next, the MAG registers the MN to the LMA on its behalf – by sending the proxy

binding update (PBU) message. Hence, the MNs are not involved in mobility management operations. The binding update message informs the LMA about the current location to which the MN is attached. Such information helps the LMA to generate or update the binding entry for the MN in the binding cache. The binding cache entry associates the MN with the current serving MAG information that enables it to reach a particular MN.

The LMA then replies to MAG with a proxy binding acknowledgement (PBA) message, implying successful update or creation of the MN's binding entry, which also includes the HNP initially assigned to the MN. Thereafter, the LMA and the MAG establish a bidirectional tunnel, and update the routing entry to allow the MN to use the address configured from the HNP. Moreover, the MAG, upon receiving the PBA message, sends a router advisement message to MN containing the same HNP previously assigned to MN; that is, the MAG emulates the home network of the MN at the access link [43]. This makes the MN unaware of its movement, because it cannot detect any IP layer change.

Figure 2-2 illustrates the basic protocol operation of the PMIPv6.

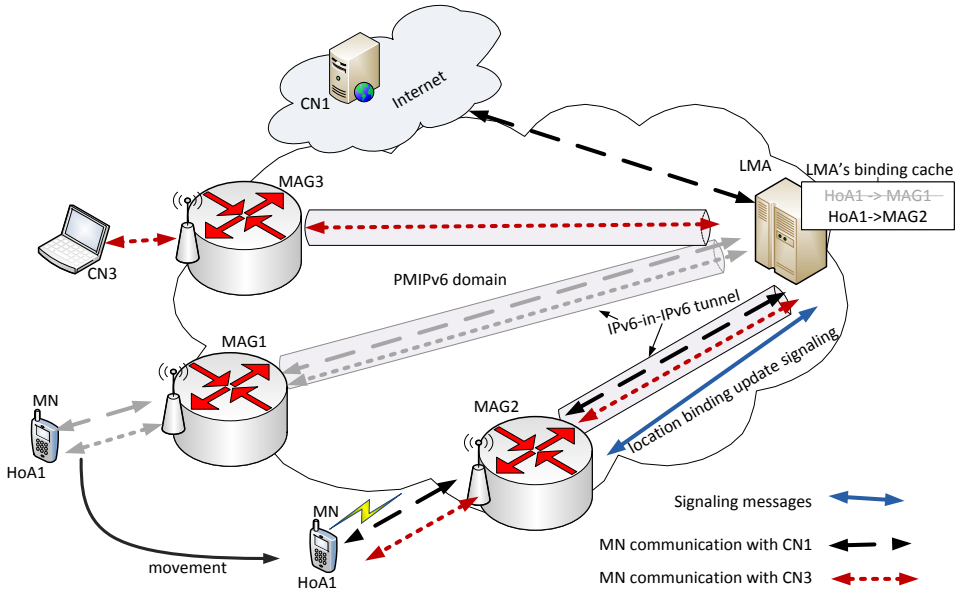


Figure 2-2 Mobility management in PMIPv6: an overview

As shown in Figure 2-2, the MN initially attaches to MAG1; and it establishes a communication with the correspondent nodes, CN1 and CN3; and it then undergoes handover to MAG2. As the MN attaches to the MAG2 network, MAG2 detects the attachment of the MN.

Then, it exchanges the proxy registration signalling, PBU and PBA messages, with the LMA to register the current location of the MN, MAG2 IP address called proxy CoA. Thereafter, the LMA updates the MN's location in the binding caches with a new entry pointing to MAG2; and a bidirectional tunnel is established between LMA and MAG2. Now, all the MN's traffic is received by the LMA, which uses the binding cache to locate MAG2 (which is currently serving the MN); and it then encapsulates packets with the MAG2 IP address as the destination address over the tunnel between LMA and MAG2.

Upon receiving the packets, MAG2 then strips off the additional IP header, created by encapsulation, from the received packets, and forwards them to the MN. The packets coming from the MN to the CN1 are tunnelled – by serving MAG, MAG2, in a bidirectional tunnel towards LMA. LMA then decapsulates and forwards them to CN1, the final destination. Similarly, the packets sent by the MN to the CN3 are encapsulated by MAG2, and decapsulated by LMA, which then encapsulates them to MAG3. MAG3 then delivers the packets to CN, after decapsulation.

Although PMIPv6 solves the host-based mobility issues of MIPv6, the MN's HoA is topologically anchored at the LMA. Consequently, all communication of the MN traverses the LMA, irrespective of the point of attachment of the MN and the CN(s), as shown in Figure 2-2. The routing path through the LMA may not be optimal [44], which can increase the communication delay. For example, both MN and CN may be located at the same MAG, or located to MAGs that are closer to each other in the same PMIPv6 domain; yet, the packets traverse the LMA. Furthermore, the binding states for all registered MNs are stored in LMA; and all the MAGs forward the MN's packets to LMA, which makes the LMA a single point of failure ; while to scale up (as the number of MNs and the data traffic volume grow) is difficult.

In principle, the LMA is a centralized anchor point in the PMIPv6 domain, in that the mobility management approach used in PMIPv6 is centralized in both control and data planes [12].

In order to overcome the limitations of PMIPv6 in terms of inefficient routing, various enhancement efforts are happening in the Network-Based Mobility Extensions working group [45] in IETF; and these efforts are briefly introduced in the following sub-section. There have also been recent efforts in the IETF aimed at addressing the centralized mobility management

issues in IP mobility management – instead of looking only at the inefficiency in routing. A new working group, called distributed mobility management (DMM) [21], has been formed to enhance the centralized IP mobility management with a distributed mobility management approach, whereby PMIPv6 is also included. The detail of DMM approach is discussed in Section 2.4.

2.3.3 PMIPv6 Enhancements and the Limitations

As discussed in the previous sub-section, the mobility signalling and the data forwarding in PMIPv6 are managed by LMA, which causes performance issues, such as a longer packet path, and traffic congestion at LMA. To address the longer data path due to packet forwarding via the LMA, the IETF has proposed various route optimization draft proposals [46] that indirectly reduce the traffic load on the LMA. The common goal of these proposals is to enable nodes attached at the same or different MAGs to use a routing path that is directly between MAGs, instead of the path via the LMA. Each proposal achieves this goal by introducing different mobility signalling. Recently, the localised routing for PMIPv6 [47] has been proposed by IETF to be a route optimization standard for the PMIPv6, to solve the non-optimal routing in particular scenarios.

In the proposed standard, the LMA exchanges localized routing initiation and localized routing acknowledgment messages to set up and maintain the optimal path, which is the direct path – between MAGs of the communicating nodes – that bypasses the LMA. Besides the IETF efforts, the research community has proposed various solutions. These are reviewed in [48], and the goal for most of these solutions is to facilitate the tunnelling between MAGs of the communicating nodes.

However, in the route optimization solutions, the LMA still maintains the MN binding information for all MNs; it is involved in the initial data routing before the route is optimized in most solutions; and it manages the binding update and the route optimization signalling. However, the issues of single point of failure may still arise in the route optimization mechanism. Moreover, the basic roles of LMA in mobility management remain unaltered, such that the centralized mobility management of PMIPv6 is inherited. Also, when an MN is communicating with a CN situated outside the domain, a single LMA still manages both traffic

forwarding and signalling messages. Yet, the route optimization is an option in basic PMIPv6 [17][44].

2.3.4 Comparison of Host-based and Network-based Mobility Management

This sub-section compares the mobility management approaches for host-based IP mobility management protocols, i.e., MIPv6, and network-based IP mobility management protocols, i.e., PMIPv6, in terms of the deployment of their mobility management functions. The goal is to examine similarities in their mobility management deployment approach, as well as common challenges they encounter because of this approach. It is these challenges we are seeking to address in this thesis.

The mobility management functions of the existing IP mobility management, both host-based and network-based, include three basic logical functions [49][23]:

(a) The allocation of HNP, or HoA, to an MN;

(b) The internetwork location management (LM): managing and keeping track of the internetwork location of an MN, which includes a mapping of the HNP (or HoA) to an address where the MN is reachable, i.e., the routing address; and

(c) The routing management (RM): intercepting packets to/from an MN's HoA and forwarding the packets, based on the internetwork location information from the LM – either to the destination, or to some other network element that knows how to forward them to the destination.

Currently, most of these IP mobility management protocols (host-based and network-based mobility management protocols) are derived from MIPv6 concepts. The protocols provide mobility support for the MN by relying on a centralized mobility anchor located in the MN's home network. Examples of this anchor are: an HA (in MIPv6), and an LMA (in PMIPv6). The mobility anchor provides to the MN all three mobility management functions: LM, RM, and the allocation of HoA. As a result, the data traffic for all MNs needs to traverse the mobility anchor for the sake of mobility service, irrespective of the location of the communicating nodes. Furthermore, the mobility anchor maintains the mobility context for all mobile nodes registered in the network; and it manages the mobility signalling used to update these contexts. In principle,

these protocols employ the mobility management approach that is centralized in both the control plane and the data plane to support mobility [12][50][3].

Considering the current trends of increase in the numbers of MNs and the mobile data traffic volume [4], these protocols experience problems that include:

(i) Non-optimal routes, when routing via a centralized anchor, which may lead to large packet delivery latency. Such large latency is not desirable for real-time communications, i.e., VoIP;

(ii) Incompatibility with the trend of cellular networks towards a more flat and distributed IP network [5][1];

(iii) Low scalability of centralized routes, and the network overhead, to maintain these routes for large numbers of mobile nodes;

(iv) Vulnerability to a single point of failure and attack;

(v) Wasting of network resources by providing mobility support to all mobile nodes; while many of them do not need such support [12][5].

Addressing these problems in centralized mobility management architecture is costly and complex [1]. In effect, these mobility management protocols cannot cope well with growth in MNs and traffic volume, as well as the evolution of mobile network towards a flat architecture. An alternative mobility management approach (which does not rely on centralized mobility management) could be an appropriate solution to address these problems. Distributed mobility management (DMM) [21][14] is one of the approaches that does not rely on a centralized mobility anchor; and this is currently under discussion in IETF. It provides mobility support with either partially distributed or fully distributed mobility management functions. These approaches are discussed in details in the next section.

2.4 Distributed Mobility Management Protocols

2.4.1 Overview

Based on the above discussion, the centralized mobility management approaches, i.e., MIPv6 and PMIPv6, maintain the three basic functions of the mobility management on the

central mobility anchor, for example, HA and LMA. To overcome the discussed limitations of the centralized mobility management, distributed mobility management (DMM) is a recent alternative approach for mobility management design. And it is the current focus of the IETF distributed mobility management (DMM) working group [21]. The approach idea is to develop an IP mobility support that allows for the distribution of mobility anchors or mobility management functions to different locations in the network. With such a distribution, an MN is served by a close mobility anchor or mobility function. That is, the mobility support is no longer relying on a centralized mobility anchor, in contrast to centralized mobility management.

This section, therefore, discusses the concepts of DMM, the classification of DMM, and the current DMM approaches in the IETF and the research community. It reviews the most referenced DMM approaches and points out their strengths and weaknesses.

2.4.2 Protocol Descriptions

The basic concept of DMM is to provide mobility support to the mobile nodes without relying on a centralized mobility anchor. Therefore, mobility management functions that commonly reside in a centralized mobility anchor are distributed in either both control and data planes, or the control and data planes are decoupled, and the data plane is distributed. The distribution is made in a way such that these functions are available to different networks and closer to the mobile user, for example, at the access router level. Thus, the data traffic and the control signalling are not centrally managed; and thereby, one can overcome many of the limitations and problems of mobility management with a centralized anchor.

With DMM, the three basic mobility management logical functions of the central mobility anchor can be separated and distributed to different network entities. For example, the RM and HoA allocation functions may be decoupled from the LM function; and these two functions are then distributed at the access part of the network, such as to the access routers [51]. In this approach, the LM function is kept centralized; that is, only the data plane of the mobility anchor is distributed. This is called a partially distributed mobility management [12][11]. Thus, the access routers will anchor the traffic with the IP address they assign to MNs. In another scenario, the LM function also can be distributed to the access router level of the network, which results in all three functions of the mobility anchor being distributed among the access routers.

In such a case, the mobility becomes distributed in both the control plane and the data plane; and this is known as the fully DMM approach [12][11]. Figure 2-3 and Figure 2-4 respectively illustrate these two approaches.

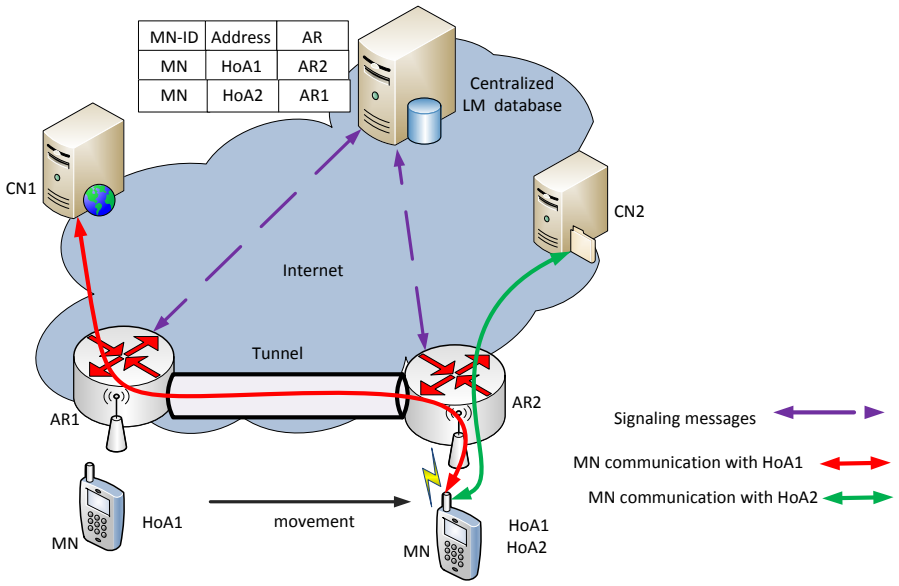


Figure 2-3 An example of a partially DMM approach

In either distribution, partially or fully distribution, the MN can either manage its mobility or the network entities can take care of the MN's mobility without its involvement. So, DMM can be categorized as either host-based DMM, or network-based DMM. Based on these two categories, the DMM mobility management schemes are currently developed by making either the existing host-based mobility management schemes, for example, MIPv6, or network-based mobility management scheme, i.e., PMIPv6, to work in a distributed manner [14].

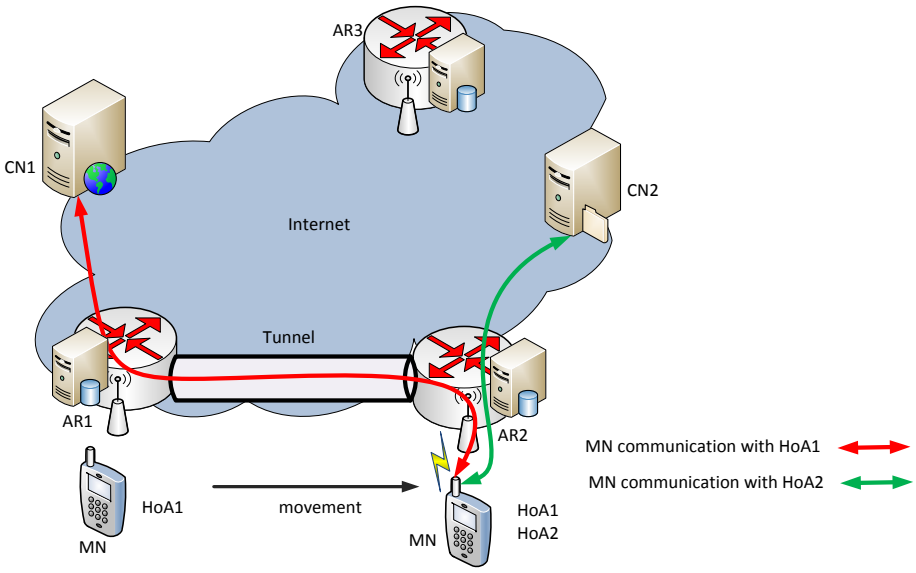


Figure 2-4 An overview of a fully DMM approach

Since the HoA allocation and RM functions may be co-located with the access routers, the MN configures different IP addresses from the different access routers it visits. Consequently, the MN handles multiple IP addresses, thanks to the IPv6 that enables an MN to configure, and to use several IP addresses on its interface(s). And the MN's traffic may be anchored to different access routers that are assigned different IP address(es) used by the MN to establish communication. This is shown in Figure 2-3 and Figure 2-4. Thus, the MN is no longer bound to establish a new communication by using an address that is topologically invalid to its current access router when it performs handover. Instead, the MN uses the new configured IP address that is topologically correct to its current access router to establish new communication(s) [52].

Nevertheless, when the MN performs a handover with active communication, it needs to maintain its old IP address(s) to serve this communication (i.e., the communication[s] established prior to handover).

To maintain communication continuity while the MN roams, the MN (i.e., in host-based DMM) or the RM in the current access router of the MN (i.e., in network-based DMM) needs to exchange mobility signalling with the old RM(s) anchoring the MN communication(s). Subsequently, a bidirectional tunnel is established: either between the old RM(s) and the MN, or

between the old RM(s) and the current RM (Figure 2-3) to serve ongoing communications. So, both host-based and network-based DMM approaches employ tunnelling mechanisms to maintain ongoing communication continuity when the MN undergoes handover. On the other hand, the newly established communication is routed optimally (i.e., using legacy IP routing) without any need for special treatment, such as tunnelling, as shown in Figure 2-3 and Figure 2-4. However, when an MN performs handover before the end of a communication, the communication may be subjected to tunnelling.

2.4.3 Application Scenario for Distributed Mobility Management

As presented in the above sub-section, the DMM can be achieved through the distribution of the mobility anchor nodes, or the mobility management functions in multiple locations of wireless and mobile systems. For these reasons, an MN located in any of these locations is served by the closest mobility entity. The question is: At what network level can such distribution be applied, and how? To answer this question, the network can be considered to constitute three levels, namely: core-network, access network and host [12][2], as illustrated in Figure 2-5. The following paragraphs briefly describe how the DMM can be achieved at these levels; and more details are given in [12][2].

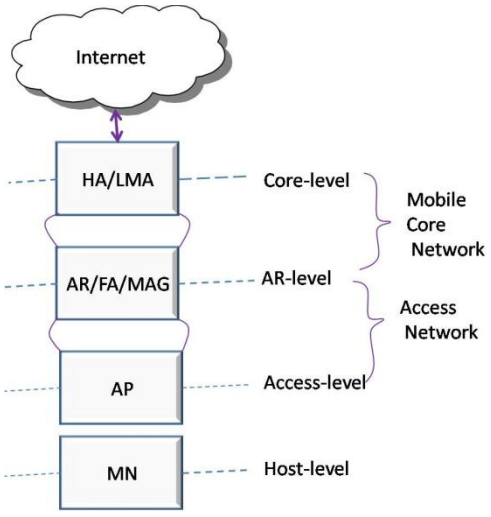


Figure 2-5 Distributed mobility management application scenarios

The core-level distribution topologically distributes mobility anchors as a whole to a specific geographical area, without decoupling the mobility management functions of the

mobility anchor. Hence, the MNs' contexts and the data traffic are managed by decentralized means at different mobility anchors. One of the examples of the core-level distribution is the Global HA to HA scheme [53], which distributes the HAs in the Internet topology, so that the MNs are served by the closest HA.

On the other hand, with access level distribution, the mobility anchors are distributed to access part of the network – such as to the access routers. Every access router, apart from providing access router services to MN, also provides to MN the mobility anchor services, i.e., LM, RM and HoA allocation services. The access Gateway (AGW) distribution [54] uses this distribution approach, where the HA and the foreign agent functionalities are collocated to the AGWs in the access network. Likewise, the solutions in [55][56] also use this distribution approach.

The host-level distribution is a peer-to-peer approach. It principally focuses on facilitating the direct exchange of data packets among the communication peers – once the correspondent peer has been found. To achieve this distribution, it may need to deploy a special information server (i.e., LM server) either in distributed or centralized fashion, in order to facilitate the search for the corresponding peer. The route optimization for mobile MIPv6 [20][40] is one of the examples of this distribution, where the MN and the CN communicate directly in the data plane after the route optimization. However, the HA is involved in the control plane to facilitate such communication. Also the end-to-end mobility management approach, such as MSOCKS [33] and SIP [34] employ the peer-to-peer communication method.

2.4.4 Methods of Distributing Mobility Functions

Mobility management functions residing in the centralized mobility anchor can be either partially distributed or fully distributed [12][2]. The partially distributed approach distributes only the data plane; whereas the fully distributed approach distributes both the data and the control planes. In the following, these approaches are explained.

2.4.4.1 Partially DMM

In the partially distributed approach, the RM and the HoA address allocation functions are separated from the LM function; and they become distributed at the access network level, such as the access routers. The LM function is kept centralized, for example in the centralized

server [57], to manage the MNs' binding information. That is, the data plane is only distributed for route optimization, without removing the signalling anchor. To facilitate the communication, when the RM receives packets for MN, it retrieves the MN location information from the centralized LM, and then forwards the packets, based on this location information.

2.4.4.2 Fully DMM

The fully distributed mobility management also removes the centralized LM; that is, it distributes the mobility functions in both the data plane and the control plane. The access routers may implement the LM, RM and HoA allocation functions in the case of host-based DMM, or they may implement these three functions in addition to the MAG function [17] in the case of the network-based DMM. Thus, each access router offers HA or LMA functions to traffic with the IP address anchored to its network, and regular access router services, or MAG services, to the traffic anchored to another access router.

It is very challenging to locate the RM serving a particular MN for packet delivery in fully distributed mobility support, since the LM function is also distributed. Mechanisms, such as multicasting/broadcasting the data, or multicasting/broadcasting the location query to all the mobility anchors in a given domain [58][57] can be employed. However, these mechanisms may increase the unnecessary traffic and waste network resources. Alternatively, mechanisms, such as anycast [53] and Distributed Hash Table [59] communication may also be employed to serve the same purpose.

2.4.5 Host-based Distributed Mobility Management

In this sub-section, efforts for developing the DMM protocols undertaken by the research community are reviewed, along with the limitations of each protocol. The protocols presented here are those that use the approach of making MIPv6 work in a distributed way, or any approach that involves the MN in managing its mobility, hence called the host-based DMM schemes.

Considering the separation of the mobility management logical functions of the mobility anchor to achieve the DMM, A. Nascimento *et al*. [60] decouple the mobility management function of the HA in MIPv6, and distribute the routing management function, RM, at access routers, while leaving the LM centralized. If the MN undergoes handover, it gets a new CoA

from the new access router to serve the newly initiated communication. Then, it registers the CoA with the location management entity, i.e., the LM. The LM then triggers the creation of the tunnel between the old and the new RMs, to enable session continuity for ongoing communication. The simulation results in [60] show that the proposed scheme reduces the packet loss and the time the MN remains unreachable during binding update procedures, when compared to MIPv6. However, the scheme performs poorly in terms of the handover delay, when compared to MIPv6. In addition, the scheme proposes the MN to manually configure an IP address of its serving location management node. This may limit the deployment of the scheme, given the current increase in the number of MNs. Moreover, the scheme still requires modification to the MN's protocol stack.

P. Bertin *et al*. [52] propose a distributed and dynamic mobility management mechanism designed for a flat architecture. The mechanism allows dynamic distribution of all the mobility logical functions to the access routers: that is, LM, RM and HoA allocation functions are performed by each access router. So, every access router assigns IP address(es) to an MN that attaches to its network. The MN uses the IP address valid to the current access router for a new data traffic flow, so as to support dynamic mobility management. Whenever the MN moves from the old access router to new access with ongoing communication, a tunnel is built between the old RM and the new RM to route the ongoing communication.

This mechanism relies on the MN uplink traffic to learn the MN's preconfigured information, so as to set up the tunnel. However, when the MN has no packets to send, this may result in a long interruption for the ongoing communication. The implementation results [22] under non-loaded network conditions have shown that the mechanism has slightly better handover delay and less end-to-end delay compared to MIPv6. Nevertheless, the mechanism does not consider the route optimization for ongoing communication. As a result, when an MN with long-lasting traffic gets far away from the traffic-anchoring access router, the resulting routing path becomes longer, and the end-to-end delay becomes higher.

F. Giust *et al*. [55] and J. Lee *et al*. [56] also present a distributed mobility management schemes, based on MIPv6 principles. The schemes distributes the three mobility logical functions of the HA in MIPv6 to each access router, which are then called the distributed anchor router (DAR) [55] and the access mobility anchor (AMA) [56].The MN obtains IP addresses

from the different access routers it visits. When the MN performs handover, in [55] the MN needs to exchange MIPv6 signalling with its old DAR(s), so as to establish the tunnel between the old DAR(s) and the MN. While in [56], the MN needs to sends its configured IP address(es) (i.e., the previous configured address and the new configured address) to its current serving AMA, which will then establish the tunnel between itself and the previous AMA(s). The MN's ongoing communication continuity is served by this tunnel; whereas the newly established communication is routed using standard IP routing. Analytically, [55] shows comparable packet overhead and high signalling load with MIPv6; while [56] outperforms MIPv6 in terms of handover delay and throughput performances. The downside of the schemes is the use of the static traffic anchoring, where the traffic remains anchored at the anchor point of its initial establishment, irrespective of where the MN has moved. Consequently, the long route may encounter the long-lasting traffic of the MN when it gets far away from the traffic-anchoring point (i.e., DAR or AMA). Moreover, the end-to-end delay becomes significant (e.g., long-range movement that covers across multiple anchors). In addition, the MN requires implementing mobility client to support MIPv6 signalling.

S. Yan *et al*. [59] also develop a MIPv6 based distributed mobility management architecture. The architecture is organized in network domains; and each domain consists of a distributed mobility agent (DMA) and access routers. The DMA forms an overlay structure over the network domain; and it performs the LM and RM functions of the HA in MIPv6. The access routers inherit the HoA allocation function of the HA; and it employs a distributed hash table (DHT) mechanism to store and retrieve the MN mobility-related information from the DMA. The architecture solves the global reachability of the MN – by extending the DNS server to map the HoA and application layer identifier, so that the CN can use this identifier to resolve the MN's HoA. But the DNS server is not designed for mobility management, because of its security and responsiveness in handling dynamic updates. Moreover, using DHT may lead to a long search time if the number of DMA increases.

R. Wakikawa *et al*. [53] present the distribution of the HAs as a whole in the Internet topology. Each HA advertises the same IPv6 prefix using anycasting. The traffic from the mobile node or the corresponding node is served by the closest topological HA, which reduces the communication delay. However, the signalling involved in synchronizing the mobile node binding location can become very large, as the numbers of the MNs and the HAs increase.

M. Fischer *et al*. [61] propose a distributed IP mobility approach for 3G SAE. The approach distributes the HA functionality in MIPv6 to several network entities, named mobility agents (MAs), close to the edges of the operator network. These include the serving gateways and the access routers of all trusted networks. The approach uses the DHT data structure to manage the binding cache of the distributed MAs. The MA close to the CN intercepts the packets destined for the MN's HoA; and it uses the DHT method to locate the MA currently serving the MN. Thereafter, it tunnels the packets to the current MA, which will then deliver the packets to the MN. When the MN performs a handover, the old MA (after learning the current MA serving the MN) notifies all the corresponding MA(s) to tunnel the packets directly to the serving MA, which optimizes the routing path. Nevertheless, the handover and newly established traffic after handover undergo tunnelling between MAs. This adds unnecessary overheads to the newly established traffic.

H. Ali-Ahmad *et al*. [62] also discuss the DMM scheme, based on the MIPv6 protocol. The scheme is similar to most host-based DMM schemes that distribute all the mobility logical functions of the HA to the access routers. In contrast, the scheme proposes additional mechanisms to support an MN moving from the access routers with HA functionality to an access router without HA functionality. To achieve this, the MN uses the MIPv6 mobility signalling to register its CoA (obtained from the classic access router) to the topologically close access router with HA functionality, which will also anchor its traffic. So, the MN does not use this new CoA to initiate a new communication. Instead, this CoA is used for routing, like the MIPv6. Consequently, the new communication established at the classical access router undergoes tunnelling. Nevertheless, the scheme does not optimize routes for ongoing sessions, and this may lead to long end-to-end delay for long-lasting communication, if the MN moves far away from the communication anchoring access router(s), when the topology spans a large area.

2.4.6 Network-based Distributed Mobility Management

This subsection discusses DMM approaches aimed to extend PMIPv6 protocol to work in a distributed way. Hence, the approaches are categorized as the network-based DMM schemes. These schemes may use either partially distributed or fully distributed mechanisms. The following paragraphs review these schemes developed by the research community and the IETF.

The reviews consider the pros and cons of the schemes.

L.Yi *et al*. [63] split the LMA in PMIPv6 into data plane and control plane, named Data plane LMA (DLMA) and Control Plane LMA (CLMA), respectively. There are several DLMAs which play the role of data forwarding, whereas a single CLMA manages mobility signalling and allocates HNP and DLMA to the MNs. In the control plane, each MAG exchanges PMIPv6 signalling with CLMA to register the MN. Next, the CLMA registers the MN to the chosen DLMA which will then anchor the data traffic for the MN. The data packets are encapsulated between the MAGs and a specific DLMA serving the MNs. Each DLMA needs to buffer the MN data packets during the handover. This may introduce a load burden to the DLMAs, especially when a huge number of MNs attached to a particular DLMA perform handover simultaneously. Nonetheless, the solution does not state where the proposed entities fit in the deployment, i.e., at what level of the network the two entities can be deployed.

F. Gius *et al*. [57] and R. Costa *et al*. [64] present fully distributed and partially distributed schemes based on PMIPv6. The fully distributed schemes co-locate all LMA mobility functionality to the MAGs, whereas the partially distributed schemes move only the RM and HNP allocation functions to the MAGs. In partially distributed schemes, [57] employs a centralised database to enable a new network to discover the previous network and to allow the establishment of the tunnel between them so as to serve the MN's ongoing traffic. On the other hand, [64] utilizes DHCPv6 server to achieve the same purpose. To allow a new network to learn the previous network in fully distributed schemes, [57] utilises IEEE 802.21 services whereas [64] uses the node information queries mechanism. However, these schemes leave the traffic anchored at the anchor point where the traffic was initially established regardless of where the MN has moved to, even if the MN has moved closer to the CN. So, long lasting traffic may experience a long route especially for frequent handover MNs. In addition, [57] and [64] do not evaluate the performance of the proposed schemes to gain the performance benefits of the schemes. Only [57] gives a simple signalling analysis. Moreover, the fully distributed scheme [64] involves the MN in mobility management that violates the network-based features of PMIPv6.

J. Kim *et al*. [58] propose three DMM approaches, partially distributed mobility control (PDMC), data-driven distributed mobility control (DDMC) and signal-driven distributed

mobility control (SDMC). Both DDMC and SDMC use fully distributed architecture whereas PDMC uses partially distributed architecture. The numerical analysis revealed that the approaches outperform PMIP in terms of binding update and packet delivery costs. However, the use of multicasting in DDMC and SDMC pushes unnecessary traffic into the network, which may waste network resources and the link bandwidth.

2.5 Comparison of Distributed Mobility Management Schemes

The existing IP DMM schemes are developed through extending either MIPv6 or PMIPv6 schemes [14] to work in a distributed way. Hence, the schemes provide either host- or network-based mobility support. In both schemes, the MNs need to manage multiple IP addresses [14][52], which require an intelligence capability be deployed to MNs. This is to enable the MNs to make the correct decision in selecting an IP address to use, maintain, and terminate for each communication session. Additionally, the MNs still need to employ mobility client function to support host-based DMM schemes; hence, the signalling overhead over the wireless link and high power consumption for the MNs cannot be avoided in host-based DMM schemes.

Both host-based and network-based DMM schemes mitigate the packets overhead for the MN's newly established communication sessions due to the use of standard IP routing [22][12]. But these sessions may undergo IP tunnelling when an MN performs handover before completion of these sessions. This causes an overhead of 40 bytes due to packet encapsulation. Moreover, most of the host-based DMM schemes still extend the tunnelling effects over the wireless link. Nevertheless, the short duration traffic of a session may not suffer from the tunnelling effect. In contrast, the long lasting traffic of a session (apart from tunnelling overhead) may encounter routing overhead when the MNs moves far away from traffic anchoring node(s) due to the static traffic anchoring used in most of the schemes.

To maintain the network-based mobility feature for the network-based DMM schemes is very challenging, particularly for a fully distributed approach. The network may involve the MNs in discovering their preconfigured information [64], e.g., IP addresses and traffic anchoring node(s), which may lead to violation of the network-based mobility feature. On the other hand, in host-based DMM schemes the MNs are well informed about such information, so they easily manage their mobility.

In both partially or fully distributed, for both host-based DMM and network-based DMM schemes, there is either a direct or indirect signalling exchange between the MN's packet anchoring nodes to either register or update MNs' location information. The amount of signalling is a function of the IP addresses an MN intends to preserve upon handover, specifically IP address(es) of the running session(s). So if the MN performs handover with several running sessions the resulting signalling will be large.

2.6 Comparison of Centralized and Distributed Mobility Management Schemes

The centralized mobility management schemes benefit from deploying a centralized mobility anchor, which is static and well known, so that updating and locating the MN's binding information becomes easier. However, this anchor keeps a huge number of MNs' contexts; and yet, it still manages the data path for these MNs. So the issues related to scalability and reliability cannot be escaped. Moreover, the location update may take longer due to the fact that the central anchor may be located far away from the MNs [44].

In contrast, the distributed mobility schemes mitigate such effects to a local level, such as at access routers. The MN contexts and the data path management are managed in a distributed way at the access network level that removes the static anchor and brings the mobility anchor closer to the MNs [14]. Nevertheless, updating and retrieving the MN's binding information becomes very challenging, especially with fully distributed schemes, because the MN's traffic may be anchored to different anchors. And it may introduce signalling overhead.

The tunnels always exist with centralized schemes, regardless of whether the MN is in motion or not. And the packets are tunnelled, irrespective of whether the packets belong to the handover session, or to the newly established session. On the contrary, the distributed schemes build the tunnel between traffic anchoring nodes on demand, for example, when an MN undergoes handover with traffic to maintain [22]. So, the newly established traffic and the traffic that does not undergo handover, do not encounter the tunnelling overhead, because they are routed using standard IP routing.

The centralized schemes do not need the MNs to implement additional intelligence, such as to manage multiple IP address, whereas the distributed schemes require intelligence to the MN

to manage and use several IP addresses simultaneously, and to select the correct one for a particular usage.

The use of a single identifier, an HoA, in centralized schemes, makes it easier for the MN to be reachable, irrespective of its point of attachment in the network. Yet, all the MN communications need to traverse the mobility anchor located at the home network of the MN [20]. This increases the load burden to the mobility anchor, and causes long transmission delay [44]. In contrast, locating an MN in distributed schemes is not an easy task (specifically in the fully distributed schemes). This is due to the fact that the MN may be managed by different mobility anchors, from which it has established communications. But the MN's communications are anchored at different mobility anchors, which are topologically closer to the MN (such that the effects of traffic load at single mobility anchor are mitigated, as well as the traffic transmission latency).

The centralized schemes are matured schemes, which have undergone several modifications, in order to reach the current standard [20][17]. The distributed schemes are still under proposal; and it is not easy to anticipate how the standard will look in the future.

2.7 Summary

This chapter has presented the concept of mobility management in the Internet for both wireless and mobile systems, and definition, types and classification of mobility management. IP mobility management, as the solution to Internet mobility support, has been discussed in terms of the operation, enhancement and the deployment of mobility management functions. It should be noted that the existing IP mobility support utilizes the centralized mobility management approach. Hence, it becomes difficult for these mobility supports to meet the current demand for the mobile Internet, which is fuelled by both the rapid growth in the number of mobile users, and the mobile data traffic volumes, along with trends towards flat mobile network architecture, which is becoming an all-IP.

An alternative mobility management approach to solve the centralized mobility limitations – with a distributed mobility management approach – has been presented. Additionally, the established concepts, classifications and efforts for developing DMM (from both the IETF and the research community) have been described in this chapter. Moreover, a

comparison of centralized mobility management schemes, and a comparison distributed mobility management schemes, have been presented. Additionally, a detailed comparison of the centralized mobility management schemes and the distributed mobility management mobility has been presented.

The review presented in this chapter has shown that previous mechanisms have attempted to provide DMM through extending MIPv6 and PMIPv6 to work in a distributed way. However, the mechanisms still incur some drawbacks, which include a long routing path, resulting from the MN's session(s) remaining anchored at the MN's communicating IP address-anchor point; a lack of route optimization for ongoing communication; synchronization problem of the MN location in different networks; and long end-to-end packet delivery delay, especially for long-duration traffic.

Moreover, signalling and tunnelling overhead over the wireless link still exist, especially in the host-based mobility management schemes. Therefore, a new DMM mechanism is required to address these issues. The research presented in this thesis, therefore, investigates the distributed mobility management approach; and it develops new IP-based distributed mobility management schemes that overcome some these limitations in the existing distributed mobility management schemes.

The following chapters develop new network-based DMM schemes, which enhance the PMIPv6 to support DMM, while seeking to address some of the above problems.

Chapter 3 Network-based DMM with Routing Management Function at the Gateway Routers: DM-RMG

3.1 Introduction

This chapter presents a new network-based distributed IP mobility management scheme, called Network-based DMM with Routing Management Function at the Gateway Routers (DM-RMG). The discussion in the previous chapter focused on the need to develop new DMM schemes, in order to address the limitations of the centralised IP mobility management schemes and that of the recently developed DMM schemes. This chapter focuses on the proposed DM-RMG scheme, and its performance evaluation. It provides a detailed description of DM-RMG, and its operation mechanisms that are demonstrated by signalling flow diagrams. The chapter also describes routing path and handover delay optimization of the proposed scheme. Furthermore, the chapter discusses the performance evaluation of the DM-RMG scheme, and the simulation environment, which is the network simulator version 2(ns-2).

The performance metrics used in the simulations include packet delivery latency, handover delay, and packet loss. The simulation results are presented in this chapter, in order to show the effectiveness of the scheme when compared with other DMM schemes. A qualitative analysis of the proposed scheme in comparison with most referenced DMM schemes is also discussed in this chapter.

3.2 Motivation and Design Approaches for DM-RMG

IP mobility management schemes defined by IETF use the principles of MIPv6, as discussed in Chapter 2. These mobility management schemes work by intercepting the MN's packets in the home network, and then tunnelling them to the MN in the visited network. This process requires that the network has the capability for managing the location information of MN by binding the MN's HoA and CoA. It also requires modification to the existing routing function of the Internet, which routes packets using a routing table. Such modification to the existing routing procedure include intercepting IP packets and encapsulating them inside other IP packets, with IP headers consisting of the destination addresses of the relevant CoAs. This modified

routing function is referred to as routing management (RM) function (or mobility routing function) in this thesis, as already defined in Chapter 2. This function does not need to exist in all the routers, so that the fixed routes are not affected. IP mobility management schemes (e.g., MIPv6 and PMIPv6) place this routing management function at the mobility anchor in the home network, through which the internetwork route of the packets using HoA address passes. This sometimes results in sub-optimal routes, especially when the MN and the CN are close to each other, but are far away from the home network. The additional problems caused by this approach have already been discussed in Chapter 2. Furthermore, the approach overload the mobility anchor with routing management, allocation of HoA to an MN, and location management functions for all MNs registered in the network.

The problems stated above can be addressed by distributing the mobility management functions and locating the RM at a location through which the packets sent from the CN(s) to an MN always pass, instead of keeping this function centralised at the home network. This location may be the gateway router or the access router. Using the option of locating RM at the access routers, DMM schemes, such as [52][55][57][60] have been developed, as presented in Chapter 2. These schemes, however, still experience limitations, such as:

i. The long routing path that results from the MN's session(s) remaining anchored at the MN's communicating IP address-anchoring network. In this case, when the MN continues roaming, the distance between the communication anchoring network and the MN's visited network may increase, resulting in increases of packet end-to-end delay (especially for long duration traffic that requires consistency for the IP address);

ii. The lack of route optimization for ongoing communication; and

iii. Synchronization of the MN location in different networks, specifically in fully DMM approaches.

The above mentioned problems are the motivation for developing the mobility management scheme, as is proposed in this chapter. Moreover, most of the previously proposed schemes are at a preliminary stage, and still need to be analysed, in order to determine their feasibility. Furthermore, no clear mechanism has been established for the process of enabling a new network to learn about the old network, from which the MN has been detached –

particularly in fully DMM approaches.

In [50], a proposal for locating the RM function at the gateway routers to achieve DMM has been discussed. However, the discussion covers only the preliminary concept for such an approach. It lacks details of the actual functioning of the proposed idea, and no analysis of the proposed idea has been given.

In this chapter, a network-based DMM scheme that co-locates RM at the gateway routers has been investigated, with detailed implementation and preliminary results presented in [23]. The proposed DM-RMG scheme presented in this chapter is developed based on the DMM concepts. The DM-RMG scheme re-uses the existing mobility management protocols by enhancing PMIPv6 to support DMM, while overcoming some aforementioned limitations in the existing DMM schemes. PMIPv6 is selected for enhancement because of its network-based features, which have attracted many network operators to consider PMIPv6 for deployment. For instance, EPC [13] and WiMAX [16] have started considering PMIPv6, in order to offer network-based mobility support for mobile users. For these reasons, it is worth enhancing the PMIPv6 scheme, so that it can support mobility in flat and distributed IP networks.

The proposed DM-RMG scheme decomposes the logical function inherent in the LMA to the following three functions: internetwork location management (LM), routing management (RM) (also referred to as mobility routing), and home network prefix (HNP) allocation, in order to achieve DMM. After decomposition of the function, the DM-RMG scheme co-locates the RM function with the gateway routers of different networks/subnets, which in flat network architecture may coincide with the access router (AR). In this way, the data-plane routing function for the MNs is served by the local RM at the network gateway. In addition, the DM-RMG scheme incorporates some mechanisms to optimize the data path of the MN that moves from its traffic anchoring network to the visited network. This optimization further reduces the packet end-to-end delay; and it removes the static anchoring of the MN traffic. Moreover, DM-RMG maintains the MN location information in a distributed database, consisting of multiple LM servers in multiple networks/subnets, each in charge of a range of HNP/HoA.

The distributed location information database does the following: (1) It enables an MN in a visited network to be reachable at the IP address that an MN configures from a different network; (2) it allows for the optimization of the packet routing path; (3) it enables a new

network, to which the MN has been attached to learn the network(s) from which the MN has been detached; and (4) it alleviates the MN location information synchronization problem.

The proposed DM-RMG scheme is further enhanced by introducing an entity named tracking MAG (TMAG) in the overlapping region shared by the different networks. The mechanism improves the handover performance by achieving seamless handover when the MN roams between visited networks.

3.3 The DM-RMG Architecture and Operation Mechanism

DM-RMG functional architecture comprises a large domain. The domain is divided into several networks, such as Net1, Net2, and Net3, as shown in Figure 3-1. Each network has mobility management functions distributed as follows: (1) The LM function, which forms a distributed database with multiple LM servers; (2) the routing management function, which is distributed to multiple locations at the RMs, with the RMs co-located with the gateway routers (GW), and (3) multiple access routers with mobility client function. The multiple access routers are equivalent to the MAGs in PMIPv6. Each network owns a unique set of HNPs, from which it delegates HNP(s) to MN(s) registered to its network.

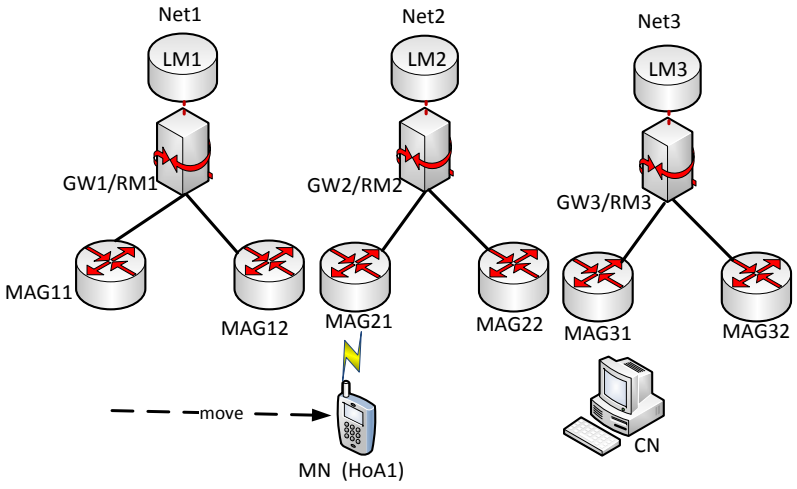


Figure 3-1 DM-RMG functional architecture

Each location management server (LM1, LM2, and LM3) in each of the networks is collocated with the HNP allocation function, so that it assigns HNP(s) to MN(s) registered to its networks. The servers maintain the mapping between the corresponding HNP(s) that they allocate to the MN(s) and the IP address of the RM co-located at the gateway of the network

where the MN(s) is located. This mapping information facilitates the internetworking mobility routing when the MN moves away from its HNP anchoring network. For example, in Figure 3-1, an MN was originally attached to Net1; and it had been allocated an HNP (i.e., HNP1), from which it had configured HoA1. When the MN performs a handover to Net2 and needs to preserve its ongoing session using HoA1, LM1 keeps the mapping between HNP1 and the address of RM2. But if the MN remains in Net1, LM1 does not need to keep such information, which reduces the mobility context load kept in the LM.

The LM1 information in the control plane is kept in a hierarchical form, for example, HNP1 to RM2, followed by HNP1 to the mobility client at the AR kept in RM2, when the MN is attached to Net2. Such mapping avoids frequent signalling updates to LM1, when the MN undergoes intra-network handover within Net2.

The overall LM function for all the networks consists of a distributed database of LMs in the domain; and they can be virtually or physically deployed.

The RM keeps the binding information between MN's HNP and the address of the mobility client at the AR (i.e., MAG) currently serving the MN. To perform internetworking mobility routing, each RM interacts with the LM from its network to retrieve up-to-date location information of the MN. For example, if the MN in Figure 3-1 undergoes handover to the visited network (i.e., Net2), RM1 gets the location information from the LM1, in order to perform internetworking mobility routing. The data plane routing function for the MNs is performed by the RM co-located at the gateway in the visited network (e.g., RM2 at GW2 in Net2).

The MAGs behave in the same way as in PMIPv6 – in other words, they perform mobility management related signalling on behalf of the MNs.

When an MN attaches to a domain, it acquires an IP address (i.e., HoA1) from the first network to which it attaches (e.g., Net1). The MN can then use HoA1 for its communication, irrespective of the network it is visiting in the domain. When the CN(s) sends packets destined for HoA1to MN, the packets first arrive at the CN's RM, assuming that the CN is also in the same domain as the MN. For example, in Figure 3-1, the packets first arrive at RM3. The CN's RM receives the first packet, and then uses its cache memory to find the location information of the MN. If it does not find the MN location information, it uses its routing table to route the packet to the RM, where the HoA1 belongs (based on the destination address information in the

packet). This procedure reduces excessive signalling that might have been caused by the CN's RM attempt to locate the MN(s) using query messages, the tunnelling overhead, as well as the packet delay for the first packets. However, the first packets may experience a long routing path if the MN is no longer in the network anchoring HoA1. On the other hand, if the location of the MN is found at the cache memory of RM3, then RM3 will tunnel the packets directly to the destination RM (i.e., the RM currently serving the MN).

Once the RM of the HoA1 receives the packets, it interacts with the LM in its network to determine whether the MN is still in its network; or if the MN has moved to a visited network. If the MN is still in the network, the RM forwards the packets to the serving MAG, which then delivers the packets to the MN. If the MN has moved away from the RM's network to the visited network, the RM uses the mapping of its LM to tunnel the incoming packets to the new RM serving the MN in the visited network. Meanwhile, the old RM extracts the IP address of the CN's RM from the source address of the incoming packets; and it then informs the CN's RM about the new RM currently serving the MN.

When the CN's RM learns about the new RM serving the MN, it temporarily stores the mapping of the route between the MN's HNP and the address of the new RM serving the MN. Subsequent packets are directly tunnelled to the new RM serving the MN. This mitigates the long routes usually caused by triangle routing. This procedure helps to optimize the route, especially when both the CN's network and the MN's visited networks are closer to each other, but far away from the network anchoring the HoA1. The cached mapping information on the CN's RM may time out – if there are no more packets flowing to the MN.

The following sub-sections describe the DM-RMG principle of operation in terms of the MN registration procedure, the data flow, and the handover mechanism (using a signalling call-flow diagram). The example of DM-RMG functional architecture in Figure 3-1 is used for purposes of discussion.

3.3.1 MN Registration Procedure and Data Flow after Registration

(a) MN registration procedure

Figure 3-2 illustrates the MN registration procedure. When an MN first attaches to the

DM-RMG domain, for example at Net1, the mobility client at MAG11, to which the MN first attaches performs the access authentication and obtains the MN profile, and then proceeds to register the MN to RM1 in a similar way, as in PMIPv6 [17] (steps 1 and 2). MAG11 sends the PBU message to RM1 on behalf of the MN. Then, RM1 in consultation with LM1, allocates the MN a HNP (e.g., HNP1), stores the binding between HNP1 and MAG11, and returns the PBA message to MAG11 (step 3). On receiving the PBA message, MAG11 sends a router advertisement (RA) to the MN containing HNP1 (step 4). The MN then uses HNP1 to configure an IP address (i.e., HoA1) through either a stateful or a stateless address autoconfiguration mechanism [65]. Given that Net1 has allocated HNP1, from which the MN configures HoA1, Net1 plays the role of the home network for the HoA1.

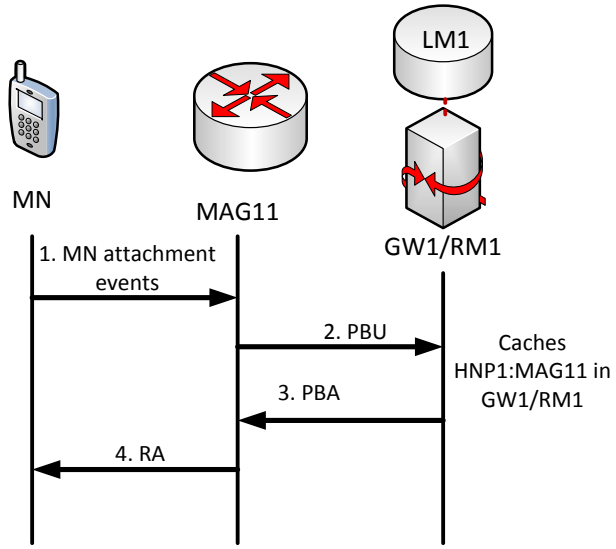


Figure 3-2 MN registration signalling flow of DM-RMG

(b) Data flow after the MN registration

After the MN has successfully configured the HoA1, the MN can send or receive packets to/from the CN using HoA1. In Figure 3-3 , the CN located in Net3 sends data packets to the MN, while the MN is situated in Net1. The packets are destined for MN through HoA1; and they will first arrive at MAG31 (step 1), which then forwards the packets to RM3 (step 2). Upon the packet arriving at RM3, RM3 checks its cache memory (step 3) and finds no HoA1 binding information. RM3 then routes the packets to RM1, using its routing table (based on HoA1 address information) (step 4). As the first packet enters Net1 through RM1, LM1 does not show

that MN has moved to a visited network (step 5). Therefore, RM1 tunnels the packets to MAG11, which then delivers the packets to the MN (steps 6 and 7).

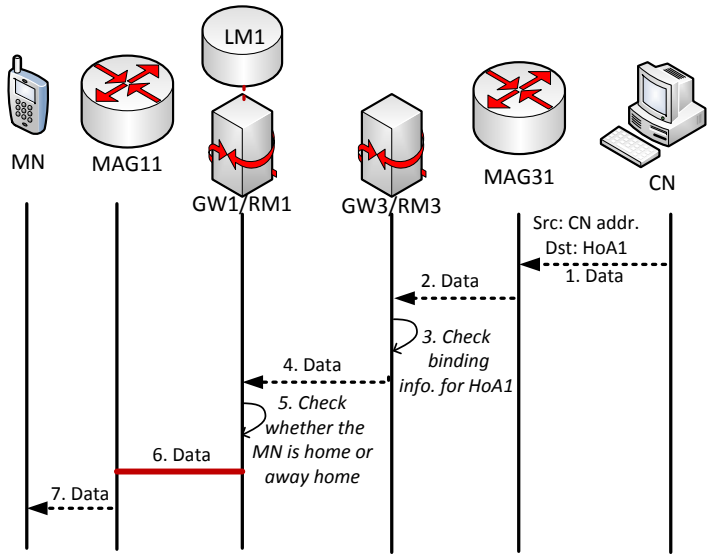


Figure 3-3 Data packet flow when the MN is in home network

3.3.2 Handover Procedure to Visited Network

Since DM-RMG comprises different networks, the MN may undergo handover from one network to another. For example, Figure 3-4 shows a scenario where the MN performs handover from Net1 (where the MN has configured HoA1) to Net2; and it needs to preserve the ongoing communication established by using HoA1.

When the MN moves to Net2, it attaches to MAG21, which detects its attachment to the access link (steps 1). MAG21 then acquires the MN's ID and HNP1, and presents them to RM2 by sending a PBU message (step 2). On receiving the PBU message, RM2 caches the binding between HNP1 and MAG21, and finds out whether it has binding information for HNP1 by interacting with LM2. It will find out that the HNP1 does not belong to Net2, and hence needs to locate the home network of the HNP1 (step 3). Thereafter, it sends a notification message to RM1 through a modified PBU message with a new defined flag (step 4). This is a new proposed message sent from the RM of the visited network to the RM of the home network of the MN's communicating IP address(es).

The message notifies this home network about the attachment of MN to the visited

network. This allows forwarding of ongoing communication to the visited network, where the MN is located. The message includes the MN's ID and HNP1. Since each network manages a unique set of HNP, the distributed database management system of LM in each network knows to which network the MN's HNP belongs; and this enables RM2 to learn about RM1, using HNP1 information.

When RM1 receives the modified PBU message, it extracts the source address from the PBU message (i.e., RM2 IP address); and it stores the mapping between the HNP1 and the RM2 address in LM1 (step 5). Thereafter, it confirms the routing path updates to RM2 through a PBA message, which includes HNP1 (step 6). On receiving the PBA message, RM2 sends to MAG21 the PBA message including HNP1 (step 7). This allows the MN to maintain the same IP address. When MAG21 receives the PBA message, it sends the RA (step 8) with HNP1 to the MN; hence MAG21 appears as it is in Net1 to the MN.

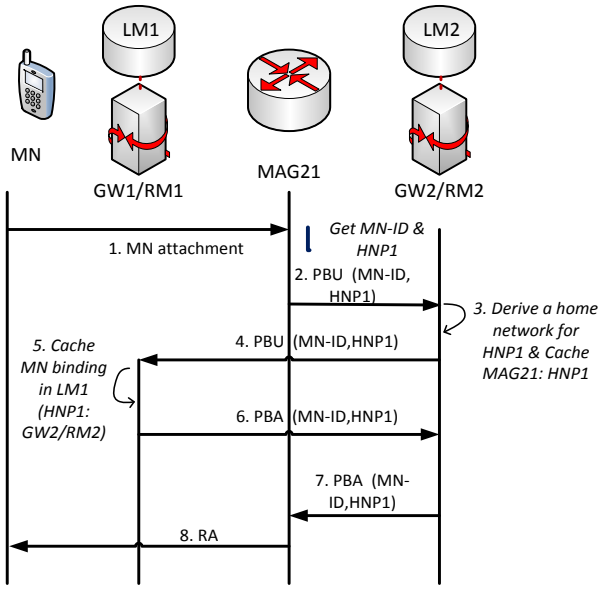


Figure 3-4 Signalling call-flow diagram when an MN moves to a visited network

3.3.3 Data Flow after Handover and Route Optimization Procedures

The DM-RMG provides route optimization support for ongoing communication. To achieve the route optimization, a new message which modifies the PBU message in PMIPv6 has been introduced. This new message is described in this sub-section. When the MN undergoes handover from Net1 to Net2 with ongoing communication, RM3 continues forwarding packets to

Net1, because it would be unaware of the MN's movement to Net2.

Figure 3-5 represents the data flow from the CN to the MN after the MN has performed handover to Net2 with ongoing communication; and it is then used to discuss the proposed mechanism for optimizing the route.

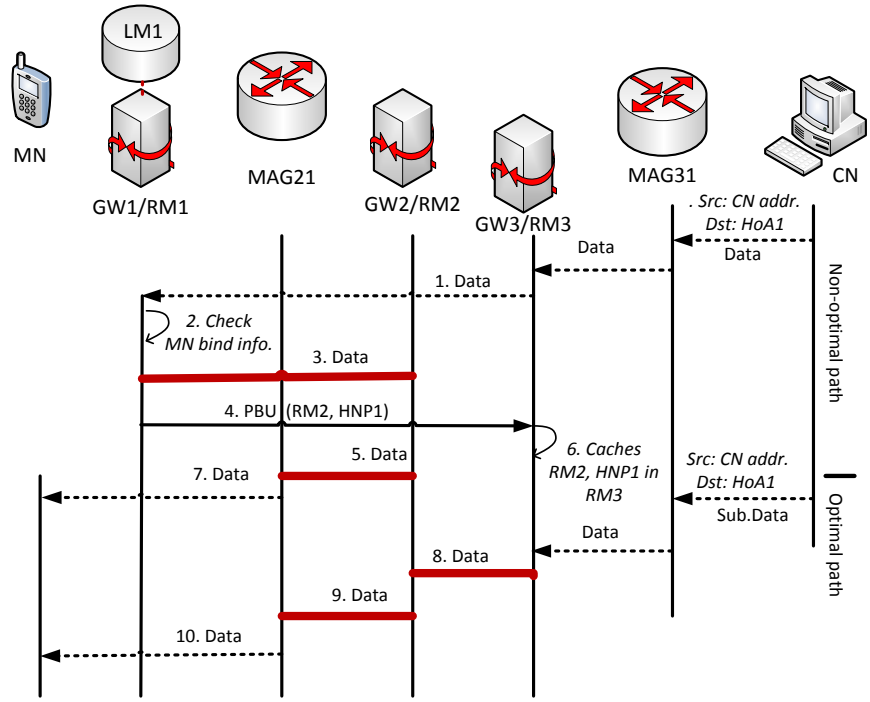


Figure 3-5 Data flows from a CN to an MN before route optimization and after route optimization

When a packet arrives at GW1 after the MN has performed handover to Net2, RM1 finds out (from LM1) the binding of the HoA1 to RM2 address; and it then tunnels the packets to Net2 (steps 1-3). The packet arrives at RM2, which then delivers the packet to the MN through MAG21 (steps 5 and 7). Meanwhile, RM1 sends the location information, PBU (RM2, HNP1) (step 4), to RM3 based on the source IP address found in the received data packets. It does this in order for RM3 to update the routing information. This is a new message, which modifies the PBU message in PMIPv6. The message is delivered from the MNs' communication IP address-anchoring network (e.g., Net1) to CNs' network (i.e., Net3); and it then informs RM3 about the new location of the MN.

This notification allows direct packet routing from Net3 to the RM currently serving the MN (i.e., RM2); and it facilitates route optimization for ongoing communication. The message

includes the IP address of RM2 and the HNP1, to enable updating of the routing path. As RM3 receives the binding update information from RM1, it temporarily stores the MN's current location information on its cache memory, and updates the route to RM2 (step 6). RM3 then proceeds to tunnel subsequent packets to RM2 (steps 8). When the packets arrive in Net2, RM2 decapsulates the packets, and it tunnels them to MAG21, which then delivers them to the MN (steps 9 and 10). Hence, after the location information has been processed at RM3, the packet routing path will be CN-GW3/RM3-GW2/RM2-MN. The stored information in RM3 is timed out when there are no more such packets flowing.

3.4 DM-RMG Extension with TMAG to Support Seamless Handover

3.4.1 Introduction and Motivation

Assume that the MN that initially moved from Net1 to Net2 performs a subsequent handover to another visited network, such as Net4, and may still need to preserve the ongoing communication established when using HoA1 (as illustrated in Figure 3-6).

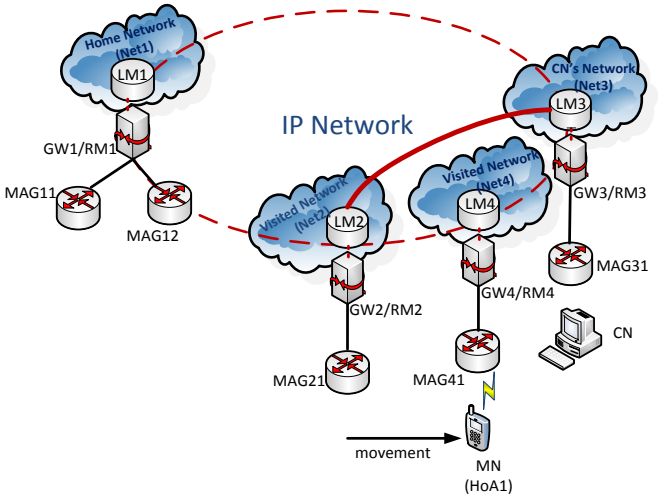


Figure 3-6 MN performing subsequent handover from Net2 to Net4

As the MN moves from Net2 and attaches to Net4, the handover procedures described in Section 3.3 will be executed. This will need Net4 to determine the anchoring network for HoA1 (i.e., Net1), based on HNP1 information, and then send a location notification message to Net1 (as illustrated in Figure 3-7 (steps 1-4)). This location notification will enable the forwarding of

MN's packets to Net4. Given that the MN has moved to Net2, Net1 will need to notify Net2 about the new location of the MN. This is to enable Net2 to learn about Net4, so that it can tunnel ongoing communication to Net4 (steps 7 and 9 in Figure 3-7). Since Net1 may be far away from the visited networks (i.e., Net2 and Net4), communication to Net1 (i.e., MN's communicating IP address anchoring network) when the MN roams between visited networks may result in a high handover delay, and high packet loss. This may cause a long communication disruption. Therefore, in this scenario, the handover mechanism for DM-RMG is extended, in order to reduce the handover delay and packet loss when the MN roams between visited networks.

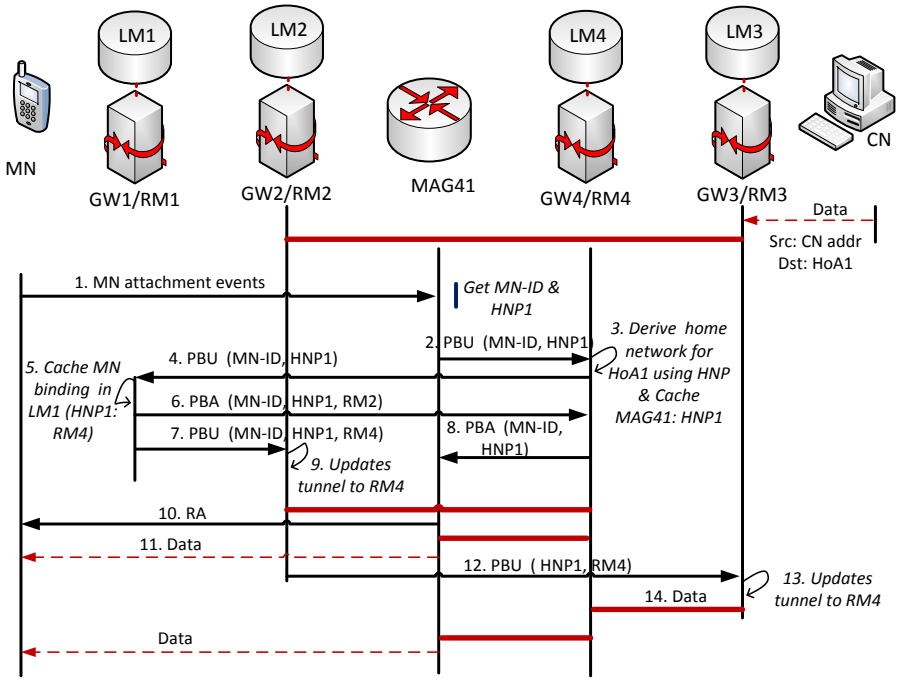


Figure 3-7 Signalling call-flow when the MN performs handover from Net2 to Net4

One possibility for improving the handover delay in this scenario is to enable the MN to cache the binding information of the networks it has visited. However, this violates the network-based mobility feature of not involving the MN in mobility management. Therefore, in this thesis a new mechanism is introduced. This new mechanism uses a Tracking MAG (TMAG) to mitigate delay caused by communication with the MN's IP address-anchoring network (e.g., Net1) when the MN performs handover between the visited networks. The handover mechanism with TMAG enhances the proposed DM-RMG scheme by providing seamless handover support

to the MN(s), while maintaining the network-based mobility feature.

The TMAG, as defined in this thesis, is a MAG, which is situated in the overlapping region shared by different networks. It connects to all networks that share the overlapping region (as shown in Figure 3-8) and its operation is limited within the overlapping region. Since the TMAG is connected to the networks that share the overlapping region, it enables the newly visited network, to which the MN is attaching, to learn about the MN's old visited network(s) – without involving the MN's communicating IP address anchoring network(s) (i.e., the network[s] that the MN has configured IP address used for ongoing session[s]).

In order to support this, the TMAG (with the help of the newly introduced message) notifies the new visited network about the address of the MN's old visited network(s) before the MN is completely attached to the new visited network, while the MN is still served by TMAG in the overlapping region. The new visited network, therefore, learns about the MN's old visited network(s) in advance. This allows it to notify the MN's old visited network(s) about the attachment of the MN to its network, which means that ongoing communication to the MN can be forwarded. To achieve this, a new message, which modifies the PBU in PMIPv6, is introduced.

The message is delivered from the RM in the new visited network to the RM in the MN's old visited network. It includes the MN's HNP, the address of the RM in the new visited network, and a new flag. The RM in the new visited network sends this message, as soon as it receives the modified PBU message from TMAG for registration of the MN. This modified PBU message also tells the RM in the new visited network about the address of RM in the MN's old visited network. When this happens, the RM in the visited network updates its forwarding route and tunnels the MN's packets to the RM serving the MN, while the MN is still being served by TMAG. Thus, the MN is allowed to continue receiving packets, as it detaches from the old visited network, and attaches to the new visited network. This feature can reduce the handover delay and packet loss when the MN performs handovers between visited networks.

Given that the TMAG may be connected to more than one visited network, the TMAG needs to determine the visited network the MN is going to handover to, so that it can forward the MN's old visited network information to the appropriate new visited network. To achieve this, the TMAG is equipped with a database that maintains the geographical location information of

its access points (AP) in relation to the networks connected to it. As the MN attaches to TMAG, and moves away from it, the TMAG traces the AP to which the MN is attaching. Using the database information, the TMAG determines the MN's location and movement direction. This helps to locate the visited network to which the MN is going to hand over. It is assumed that there are no ping-pong effects in the MN movements.

Next, the details of the extended handover operation with TMAG supports are presented, along with the signalling call-flow diagram.

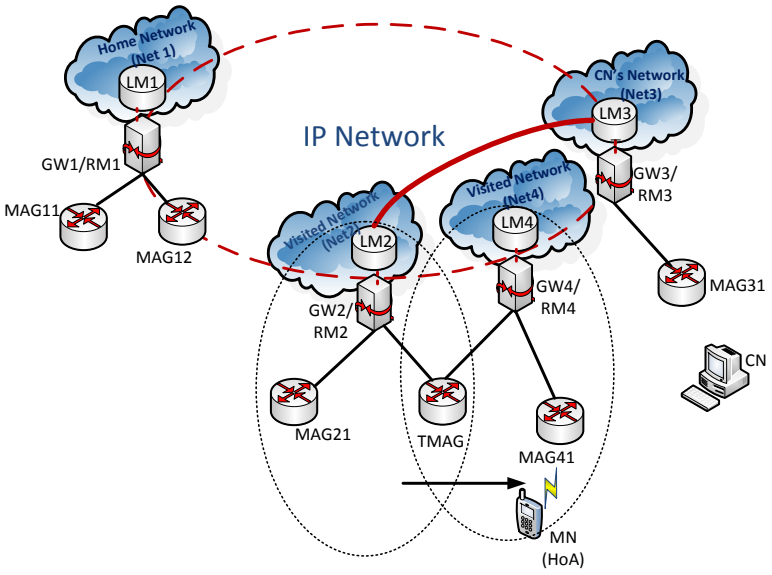


Figure 3-8 TMAG in overlapping region shared by GW2/RM2 and GW4/RM4 networks

3.4.2 Handover Mechanism with TMAG

This sub-section describes the handover operation, which extends DM-RMG handover procedures with TMAG support. The description starts from the point when the MN enters the TMAG region; and it ends when the MN is able to receive packets through the new visited network. The signalling call flow in Figure 3-9 shows the handover operation procedures with TMAG. In the figure, the MN is performing handover from Net2 to Net4 (as illustrated in Figure 3-8). The MN is moving linearly through the overlapping region towards Net4, while its established optimized path for ongoing communication is through Net2. As the MN enters the overlapping region, TMAG will detect the MN's attachment to its access link (step 1); and it will then perform an access authentication procedure for intra-network handover in Net2, as in [17].

After successful authentication, the TMAG receives the MN's profile, including the address of RM2. It then sends the PBU message to RM2 (step 2), in order to update the MN's location. When RM2 receives the PBU, it updates its packet forwarding information to lead to TMAG, and then responds to TMAG with a PBA message (step 3). The PBA includes the HNP that was advertised to the MN during the attachment to Net2; and this HNP belongs to Net1. Subsequently, the MN continues receiving packets through TMAG using its HoA1.

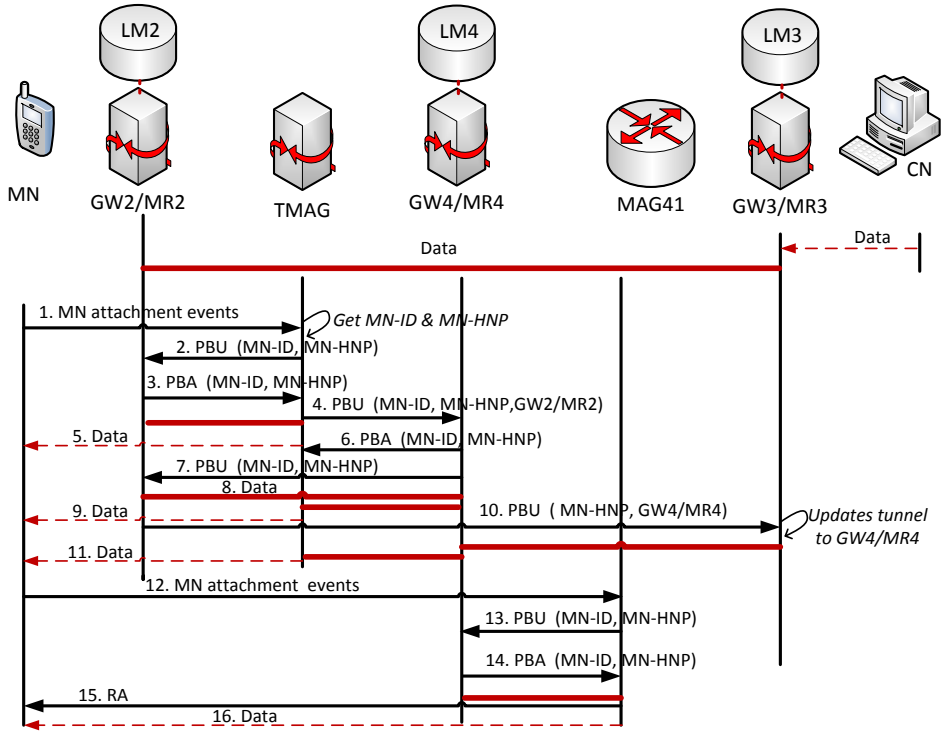


Figure 3-9 Signalling call-flow when the MN performs a handover from Net2 to Net4 networks with TMAG configured in the overlapping region between these networks

Upon receiving the PBA, the TMAG determines the network to which the MN is likely to handover (e.g., Net4), using the location tracking approach, as described in subsection 3.4.1 above. Thereafter, the TMAG sends a modified PBU message to RM4 (step 4), including the address of the current serving RM (i.e., RM2). The modified PBU is the newly introduced message that modifies the PBU message in PMIPv6, and is delivered by the TMAG to the candidate visited network to which the MN is likely to perform a handover (i.e. Net4). This message informs RM4 in Net4 about the MN's HNP and the address of RM2, which is currently serving the MN in Net2. Furthermore, the message also requests RM4 to register the MN.

When RM4 receives the PBU from the TMAG, it extracts the address of RM2, and sends a handover notification message to RM2, by utilizing the modified PBU message (step 7). This enables RM2 to update its forwarding route, and hence to forward packets towards RM4. As RM2 receives the notification message, it caches the mapping of HNP1 to RM4 in its cache memory; it then builds a tunnel to RM4, and tunnels the MN's packets to RM4 (step 8). After that, the packets flow from the CN to the MN as follows:

CN→GW3/RM3→GW2/RM2→GW4/RM4→TMAG→MN

Meanwhile, RM2 informs RM3 to forward packets directly to RM4 (step 10). It can be seen that the tunnel between RM2 and RM4 is created, while the MN is still receiving packets from RM2 through TMAG (step 5). Moreover, the MN continues receiving packets from RM2 through RM4 to TMAG (steps 8 and 9), while still remaining capable of receiving packets from RM2 through TMAG. This smooth transition of the MN's packets from RM2 to RM4, when the MN is performing the handover from Net2 to Net4, reduces the packet loss and handover delay.

When the MN attaches to MAG41 in Net4, it only undergoes intra-network handover (steps 12- 14), since TMAG also belongs to Net4.

3.5 Simulation Environment and Scenarios

The following section provides an overview of the network simulator version 2 (ns-2) used to implement and evaluate the performance of the proposed scheme. The simulation scenarios used to model the proposed scheme, as well as the simulation configuration parameters and the simulation environment, are also given.

3.5.1 Network Simulator Version 2 Overview

Network simulator version 2 [66], commonly known as ns-2, is a discrete event driven network simulator popularly used to simulate wired and wireless networks. It is an open source network simulator written in C++ and OTCL programming languages. Object-oriented C++ provides fast execution; and it used to implement the protocol to be simulated. OTCL (Object Tool Command Language) provides simple and flexible scripts, which allow easy and quick modifications of simulation scenarios, such as simulation setup and the configuration of different models. These two languages are linked together through an interface called TclCl.

The simulator supports different network components, traffic, routing algorithms, and protocols (i.e., TCP and UDP). This provides a platform for network researchers to develop and simulate different network models. The researchers can modify or extend existing ns-2 models in C++ to suit their simulation requirements.

Simulation scenarios are implemented using OTCL scripts, which involve the creation and configuration of nodes, link creation, network setup and run of the scripts. The simulation results are collected in a trace file that includes all the events that have occurred. Text-processing tools, such as AWK can then be used to extract the desired data from the trace file for the analysis.

The simulation of the proposed scheme has been carried out using ns-2 (release 2.29) with NIST mobility package [67]. Ns-2.29 with NIST mobility package provides additional modules that can implement PMIPv6.

3.5.2 Simulation Scenarios

This sub-section presents the implementation and simulation scenarios used to evaluate the performance of the DM-RMG scheme. The ns-2.29 with the NIST mobility package introduced above has been extended to model and simulate DM-RMG scenarios. Two scenarios have been investigated, as explained below.

The first scenario investigates a situation when the MN is located away from its communicating IP address-anchoring network, in a visited network that is closer to the CN's network. Two routing paths followed by the packets when travelling from the CN to the MN were investigated. These paths are named the *sub-optimal path* and the *optimal path*. The *sub-optimal path* represents the routing mechanism used in static anchoring DMM schemes (which is also a mechanism used in centralized IP mobility management, such as PMIPv6).

In the *sub-optimal path*, the packets sent to the MN while away from its communicating IP address-anchoring network always pass via this anchoring network, which then tunnels them to the network that the MN is currently visiting, as shown in Figure 3-5. The *optimal path* is provided by the DM-RMG routing path optimization mechanism. It is achieved by DM-RMG optimizing the route – soon after the MN has performed a handover to the visited network (as

described in Section 3.3, with the illustration in Figure 3-5). So, the packets travel directly from the CN's network to the MN's currently visited network. Both the *optimal* and *sub-optimal paths* that the packets follow are simulated under identical traffic loads; and the packet delivery latency performances are compared.

The second scenario investigates the handover delay and packet loss performances of the DM-RMG handover mechanisms. The proposed handover mechanisms have been simulated. (This is discussed in sub-section 3.3.2 with Figure 3-7, and its extension with TMAG support given in Figure 3-9.) In both mechanisms, it is assumed that the MN has moved to a visited network (from its communicating IP address-anchoring network), and that it is performing a subsequent handover (from its initially visited network to a neighbouring visited network), as shown in Figure 3-6. The simulation extends the implementation in Figure 3-10 by adding another visited network and TMAG, as illustrated in Figure 3-6 and Figure 3-8 , respectively.

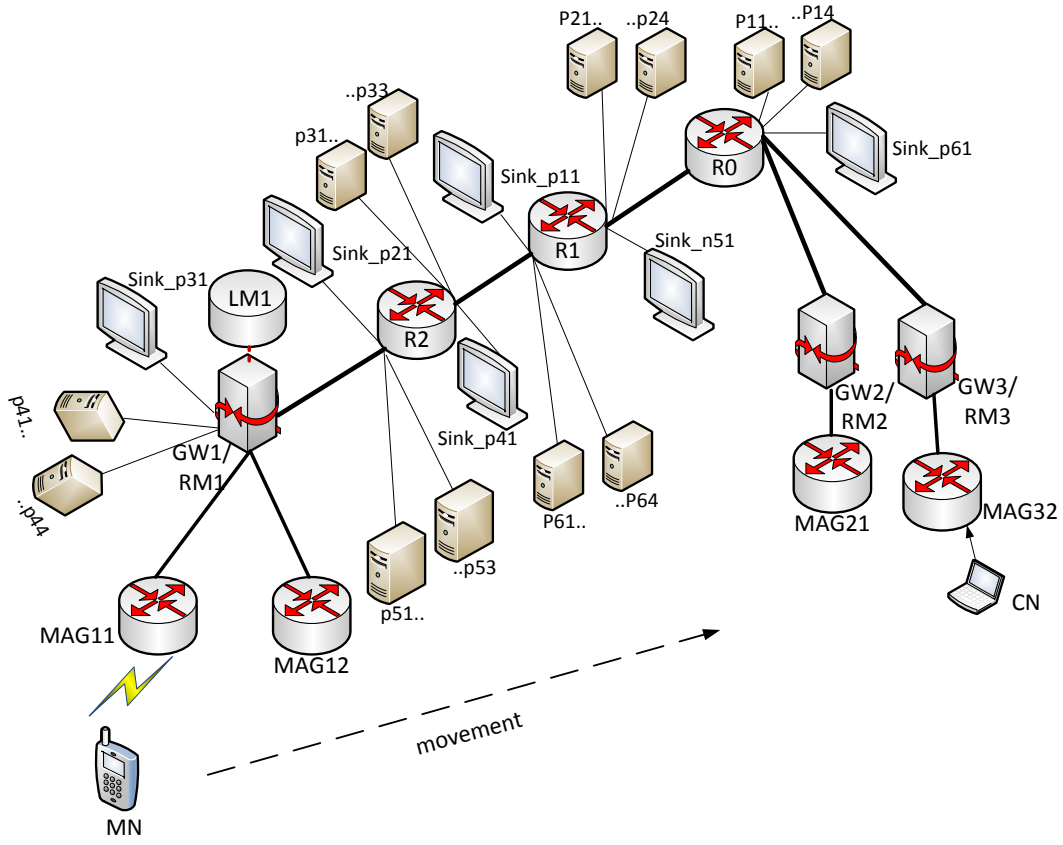


Figure 3-10 Simulated network topology

Figure 3-10 shows the network topology implemented in the ns-2 simulator. In the

simulation, the LM and the RM are co-located at gateway routers (GW/RM), and the RM function is distributed in all the gateways. The GW1/RM1 network presents the MN's communicating IP address-anchoring network; GW2/RM2 network is the MN's visited network; and GW3/RM3 network is the network where the CN resides.

To demonstrate that the GW1/RM1 network is located far away from GW2/RM2 network, intermediate routers (R0, R1, and R2) are placed in the path to the GW1/RM1 network, as illustrated in Figure 3-10. To make the simulation closer to real Internet scenarios, four nodes are connected to the inputs of each intermediate router (to both sides). The nodes p11,…, p14 through p61,…, p64 generate background traffic with an exponential distribution to congest the buffers at the output of the routers to the link input, by sending traffic to the sink/null station connected to the next router.

Each node generates a traffic load of 18.75Mbps under heavily loaded network conditions. A total load of 75Mbps (75% of the transmission capacity) has been generated at the buffer of the intermediate routers in both output sides to the link. The background traffic is received by the sink (null) station connected to the next router. For example, sink station sink_p11 frees the background traffic generated from nodes p11, p12, p13 and p14.

All the wired links are configured with bandwidth of 100Mbps. The wired link delay between intermediate routers is set to 0.25ms; and for the rest of the wired links, a delay of 0.1ms is used. The CN transmits the Constant Bit Rate (CBR) traffic over UDP of packet size 1000bytes every 0.01s to the MN. The MN is configured to move in a horizontal line at a speed of 30m/s from GW1/RM1 network to GW2/RM2 network. These configuration parameters are arbitrarily chosen for the purpose of investigating the behaviour of the proposed scheme through simulation.

To investigate the impact of the topological distance of the MN visited network from its communicating IP anchoring network on packet delivery latency, the topological distance from GW/RM1 network to GW2/RM2 network is varied by increasing the number of the intermediate routers in the path to GW1/RM1 network. Three, six, nine and twelve intermediate routers have been used to emulate the increase in topological distance between the GW2/RM2 network and GW1/RM1 network. During these changes, the background traffic nodes are increased and configured in a similar manner to generate congestions in the network at each intermediate router.

To demonstrate the handover delay improvement of the proposed mechanism, the TMAG has been configured in the overlapping region of GW2/RM2 and GW4/RM4 networks, as shown in Figure 3-8. The TMAG has connections to both GW2/RM2 and GW4/RM4; and it is carefully configured, so that the power it transmits is limited to the overlapping region. A new flag named *tmag-flag* is added to extend the PBU sent from TMAG to GW4/RM4 to include the address of GW2/RM2 (in order to differentiate it from the normal PBU message).

3.6 Simulation, Analytical, and Performance Evaluations

The DM-RMG design focuses on distributing the RM function at the gateways in different networks to achieve DMM. Through this functional distribution of the RM, the MN is locally served by the closest RM function, so that long routes are avoided. Moreover, the proposed route optimization mechanism further reduces the long routes for the ongoing traffic. With this in mind, end-to-end delay is considered as the main performance metric. The end-to-end delay is an important parameter affecting user satisfaction with regard to services (i.e., the quality of service (QoS)). End-to-end delay measures the packet delivery latency, which is the time the packet takes from the source node to the destination node. Handover delay and packet loss are other important parameters studied in the evaluation, because they affect the performance of the delivered service. The handover delay measures the latency when the MN is not able to receive or send packets. When the handover delay is large, it causes packet loss and disrupts the communication between communicating nodes. Therefore, the end-to-end delay, the handover delay, and the packet loss performance metrics are used to evaluate the performance of DM-RMG.

The next sub-sections discuss the analytical evaluation, as well as the simulation results.

3.6.1 Analytical Evaluation for End-to-end Delay

The analysis considers two routing paths taken by the packets: the DM-RMG optimized path, and the sub-optimal path, which is normally the routing path for the static anchoring DMM schemes and the centralized IP mobility management schemes. In the following section, the two routing paths are simply referred to as the optimal path and the sub-optimal path, respectively. The evaluation analyses the time it takes for the packet generated at the CN to reach the MN,

following the above-mentioned paths. In the sub-optimal path, the packets always traverse the MN's communicating IP address-anchoring network. In the optimal path, the packets follow the optimal path; hence, they could bypass the MN's communicating IP address anchoring network.

Figure 3-11 illustrates the paths that the packets transmitted from the CN follow to reach the MN. In this figure, the red-dashed line indicates the sub-optimal path; and the blue-dashed line indicates the optimal path.

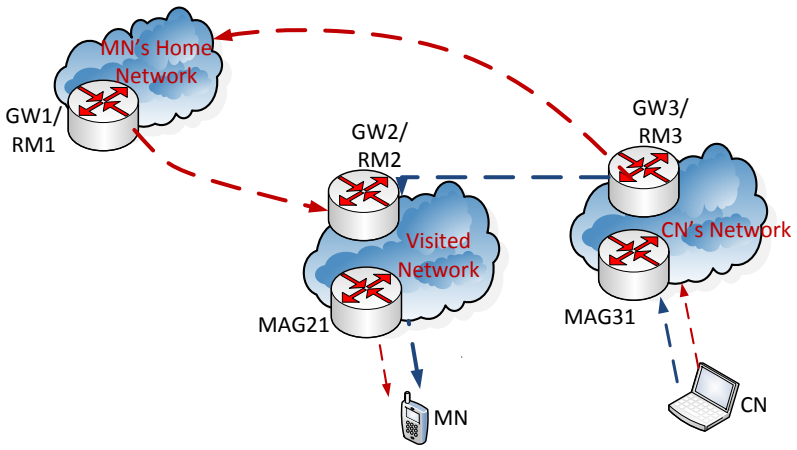


Figure 3-11 The sub-optimal and the optimal paths that the packets follow from the CN to the MN

a. End-to-end delay for the sub-optimal path

From Figure 3-11, a packet sent from CN to MN encounters the following delays: The packet needs time t_{MAG-MN} to reach MAG31. MAG31 then sends the packet to GW3/RM3, which takes the time of t_{MAG-GW}. Then, RM3 forwards the packet to GW1/RM1, incurring $t_{GW3-GW1}$ delay. The packet is then tunnelled from GW1/RM1 to GW2/RM2 with a delay of $t_{GW1-GW2}$. Then, RM2 encapsulates the packet to MAG21, where the MN is currently connected, with a delay of t_{GW-MAG}. Lastly, a delay of t_{MAG-MN} is incurred to deliver the packet from MAG21 to the MN. The packet also incurs processing delays, $P_{GW/RM}$, at GW3/RM3, GW1/RM1 and GW2/RM2. Accordingly, the end-to-end delay can be expressed as:

$$t_{end-to-end} = 2t_{MAG-MN} + 2t_{MAG-GW} + 2t_{GW1-GW2} + P_{GW3/MR3} + P_{GW1/MR1} + P_{GW2/MR2} \quad \textbf{(3.1)}$$

It is assumed that the MN's communicating IP address-anchoring network is located far

away from the MN's visited network; while the CN's network is closer to the MN's visited network, such that $t_{GW1-GW2} \approx t_{GW1-GW3}$.

b. End-to-end delay for the optimal path

A packet generated at the CN that follows the optimal path is tunnelled between the CN's network and the MN's visited network. That is, $t_{GW3-GW1}$ and $t_{Gw1-GW2}$ delays are avoided. Hence, the end-to-end delay can be presented as:

$$t_{end-to-end} = 2t_{MAG-MN} + 2t_{MAG-GW} + t_{GW3-Gw2} + P_{GW3/MR3} + P_{GW2/MR2} \qquad \textbf{(3.2)}$$

From (3.1) and (3.2), it may be noticed that the packet that follows the optimal path does not encounter significant propagation and processing delays – caused by traversing the route through the MN's communicating IP address-anchoring network. Moreover, the delay caused by nodes that may be located in this path towards the anchoring network is also avoided in the optimal path.

3.6.2 Simulation Results and Performance Analysis for the End-to-end Delay

This sub-section analyses the impact of network load and topological distance on packet delivery latency for both optimal and sub-optimal paths. It also analyses the variations of the packet delivery latency, as the MN moves from the traffic anchoring network to the visited network.

3.6.2.1 Effect of Network Load on End-to-end Delay

Figure 3-12 illustrates the effect of the network load (which has been simulated as background traffic) on the end-to-end delay – for both the sub-optimal path and the optimal path. It was observed that the end-to-end delay for the packet travelling through the sub-optimal path increases, as the background traffic load increases. Furthermore, it is also noticed that there is increased randomness in the end-to-end delay for the sub-optimal path. The reason is that the packet travelling through the sub-optimal path encounters many intermediate nodes and an increased path length. Both of these contribute to the transmission delay at each intermediate node, and to the propagation delay on the link between the nodes. Furthermore, the increased

background traffic load causes an additional delay due to processing and queuing at the intermediate nodes.

The randomness is caused by the queuing delay variation at each intermediate node along the path to the MN's communication IP anchoring network, as the background traffic load is changed. This has an effect on the QoS for real time streaming applications, voice and video, since if a packet is delayed past its play out time, the packet is effectively lost [68]. Optimizing the path is then necessary to alleviate this problem.

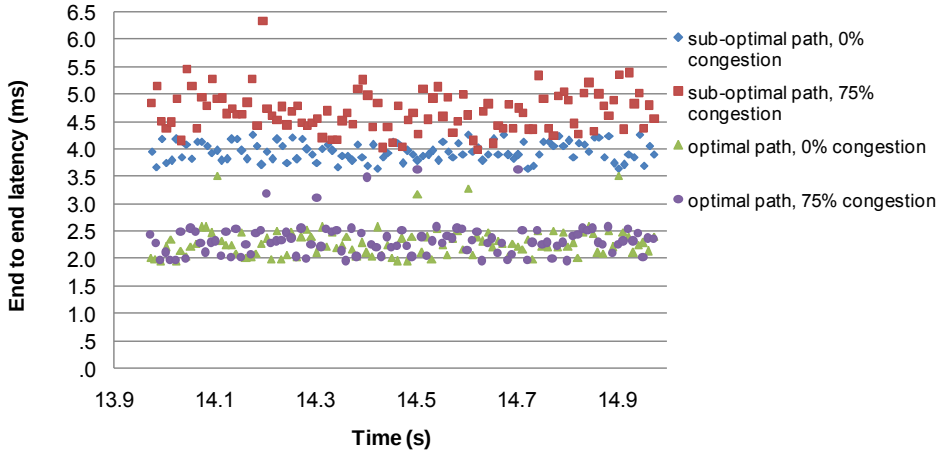


Figure 3-12 Comparison of end-to-end delay between the optimal and sub-optimal paths with change in the background traffic load, with three intermediate routers on the path to MN's communicating IP address-anchoring network

In contrast, the end-to-end delay for the packets following the optimal path is less, and is not greatly influenced by the background traffic. This is because the optimal path has fewer intermediate nodes, and hence, less transmission and propagation latencies, and reduced queuing effects.

3.6.2.2 Impact of Increased Topological Distance on End-to-end Latency

Figure 3-13 and Figure 3-14 demonstrate the impact of the increased topological distance of the MN's communicating IP address-anchoring network from the MN's visited network on the end-to-end delay. The result is obtained by varying the number of intermediate routers in the path to the MN's communicating IP address-anchoring network under identical background traffic loads. Figure 3-13 shows the variation from three routers to six routers; and Figure 3-14 shows the variation from three to twelve routers. It may be observed that, as the MN moves far

away from its communicating IP address-anchoring network to the visited network, the end-to-end delay for the packets that travel through the sub-optimal path becomes large. Also, the summary of the statistics of the end-to-end delay (given in Table 3-1) verifies the increase in the end-to-end delay of the sub-optimal path.

Table 3-1 shows that when three intermediate routers are placed in the path towards the MN's communicating IP address-anchoring network, the average end-to-end delay of the sub-optimal path is approximately 50% higher than that of the optimal path. It becomes approximately 81% higher, when twelve intermediate routers are used. The reason is that as the topological distance increases, the packet encounters a significant number of the intermediate nodes, and the path consequently, becomes longer. Both of these factors contribute to the total transmission delay and the propagation delay on the link between the nodes. It was also observed that there is an increase in randomness in the end-to-end delay, as shown in Figure 3-13. This effect is due to the queuing delay variation at each intermediate node, when the background traffic load of 75Mbps (75% of transmission capacity) is used.

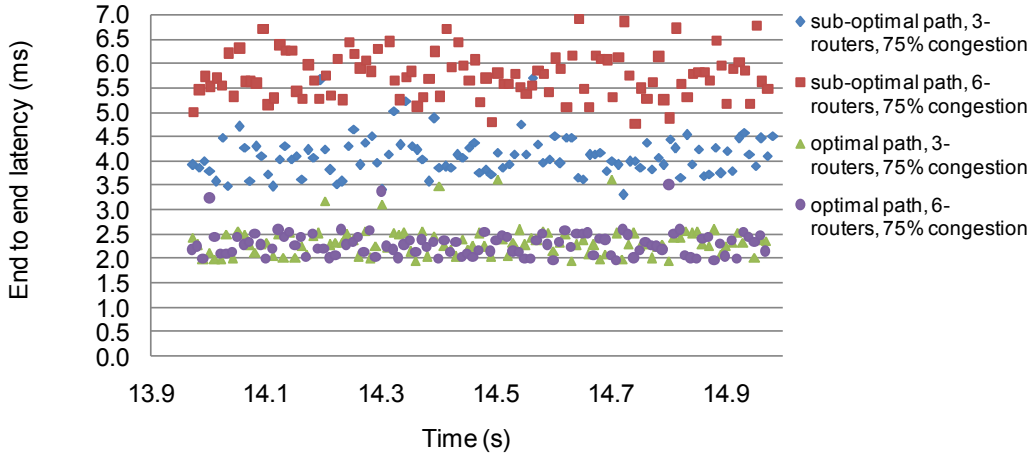


Figure 3-13 The influence of the increased distance between the MN's communication IP anchoring network and the visited network between the optimal and sub-optimal paths under the same network background traffic load

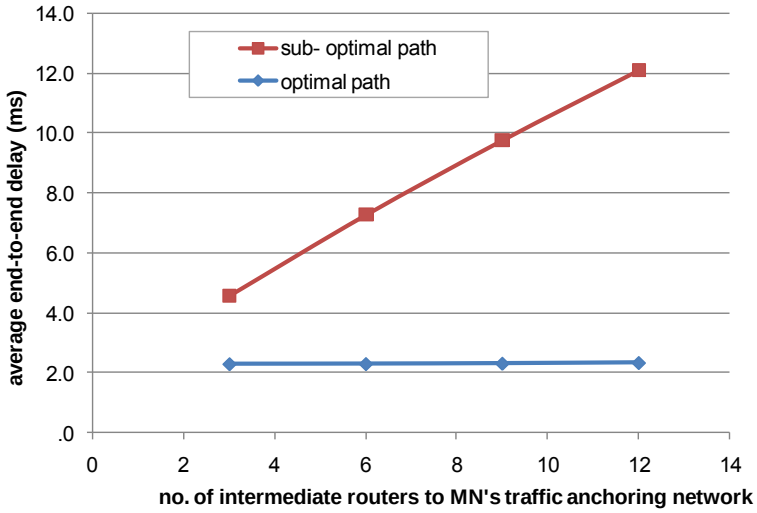


Figure 3-14 The impact of the distance between the MN's communicating IP address anchoring network and MN's visited network on the end-to-end delay of the optimal and sub-optimal paths

It is also observed that the packet that passes through the optimal path is not influenced by the increased distance, because the packets do not follow the sub-optimal route through the MN's communication IP address-anchoring network, which involves many intermediate nodes.

Table 3-1 Summary of the statistics of the end-to-end delay

No. of intermediate routers	sub- optimal path				optimal path			
	Min.(ms)	Avg.(ms)	Max.(ms)	stdDev(ms)	Min.(ms)	Avg.(ms)	Max.(ms)	stdDev(ms)
3	3.743	4.584	6.144	.358	1.957	2.303	6.144	.324
6	6.093	7.286	9.429	.527	1.957	2.308	9.429	.450
9	8.501	9.780	12.214	.611	1.957	2.320	12.214	.591
12	10.200	12.120	14.776	.700	1.957	2.344	15.375	.741

3.6.2.3 End-to-end Delay Variation during MN Movements

Figure 3-15 shows the effectiveness and the fast-path optimization of the DM-RMG scheme after the MN performs the handover. The figure shows the end-to-end delay variation for packets sent to the MN before handover, during handover, and after handover to a visited network. The variation is shown for both the optimal and the sub-optimal paths. It is observed that both paths experience similar end-to-end delay before the handover; and all the packets sent to the MN during handover duration are lost. It is also observed that after the MN has performed handover, there are few packets that experience longer latency in DM-RMG.

The reason is that after the handover, DM-RMG uses the first packet that travels to the

MN's communicating IP address anchoring network to trigger the route optimization process. Thus, some packets will traverse the traffic anchoring network before the CN's network is updated about the network in which the MN is located. However, it can be noted that only the first few packets after the handover traverse the sub-optimal path. This confirms that the DM-RMG path optimization scheme is faster and more efficient.

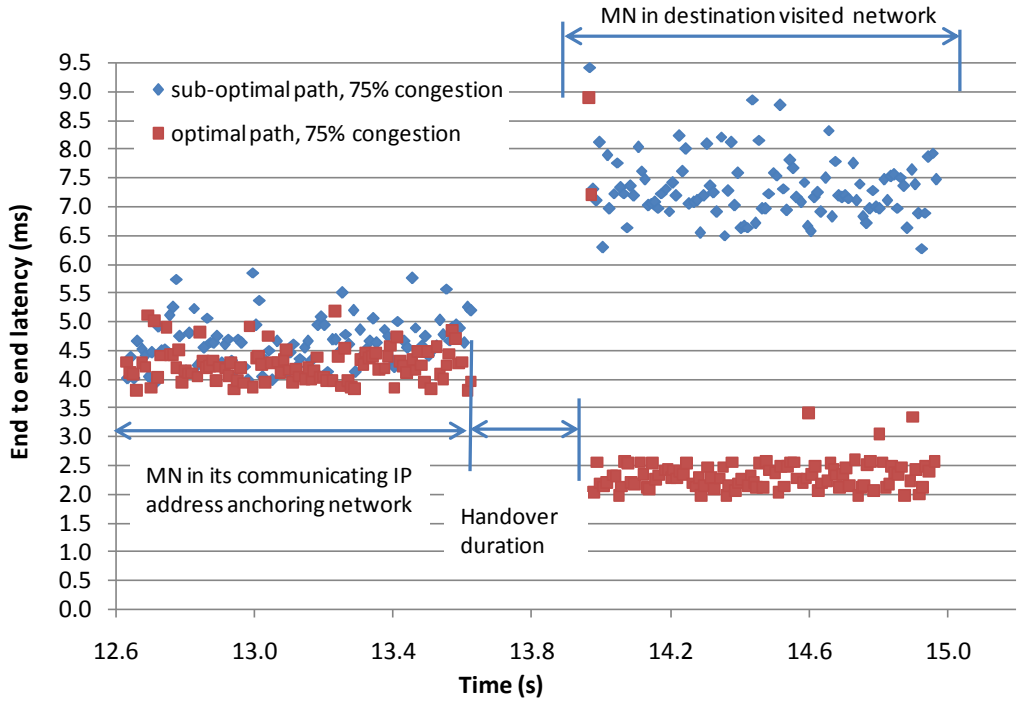


Figure 3-15 Comparison of packet delivery latency before and after the MN's handovers to a visited network between the optimal path and the sub-optimal path

Moreover, it is observed that after handover completion, the average end-to-end delay of the optimal path becomes 43% less, when compared with the latency before the handover. Also, the sub-optimal path experiences an increase of 58% in the average end-to-end delay. The reason is that the CN's network is closer to the network that the MN is visiting. That is, after the MN has performed the handover to the visited network, and the route optimization is performed, the packets follow a route with fewer intermediate nodes towards the network that the MN is visiting. In contrast, the packets following the sub-optimal path continue travelling through the MN's communicating IP address-anchoring network (even after the handover) before they are delivered to the network the MN is visiting. In other words, a delay due to forwarding packets from the MN's communicating IP address anchoring network to the visited network causes the

increase in the end-to-end delay for the sub-optimal path after the MN has performed the handover to the visited network.

3.6.3 Analytical Performance Evaluation for the Handover Delay

This analysis considers handover delay when the MN performs handover from one visited network to another, as illustrated in Figure 3-6. The analysis investigates two scenarios. In the first scenario, TMAG is configured in the overlapping region. In the second scenario, the TMAG has not been configured. The analysis then compares these two scenarios. The original proposed handover mechanism for DM-RMG does not utilize TMAG features, (as shown in Figure 3-7). Figure 3-7 shows that the handover includes several different delays:

- The MN attachment delay to MAG41 (t_{attach});
- The delay in sending the PBU from MAG41 to GW4/RM4 for MN registration (t_{MAG-GW});
- The delay in sending the location notification message to MN's communicating IP address-anchoring network (i.e., GW1/RM1 network) ($t_{Gw4-Gw1}$);
- The delay when GW1/RM1 informs the old visited network (GW2/RM2) that it must update the tunnel towards the new visited network ($t_{Gw1-Gw2}$);
- The delay when the data packet travels from GW2/RM2 to GW4/RM4 ($t_{Gw2-Gw4}$);
- The delay incurred by the data packet from GW4/RM4 to MAG41 (t_{MAG-GW}); and finally;
- The delay incurred in delivering the data packet from MAG41 to the MN (t_{MAG-MN}).

Since the GW1/RM1 network is assumed to be located very far away from the visited networks, so $t_{GW4-GW1} \approx t_{Gw1-Gw2}$. Accordingly, the total handover delay is given as

$$H_{delay} = t_{attach} + 2t_{MAG-Gw} + 2t_{Gw4-Gw1} + t_{GW2-GW4} + t_{MAG-MN} \quad \textbf{(3.3)}$$

In Figure 3-9, the handover mechanism deploys the TMAG in the overlapping region, which continues delivering data packets to the MN as the MN moves between the visited networks. When the MN attaches to TMAG, the TMAG delivers to MN the packets it has received from the serving visited network (a visited network from which the MN is about to detach). Once a tunnel is built between this serving visited network and the new visited network (a visited network the MN is handing over to), the serving visited network tunnels the packets to the new visited network, which then forwards the packets to TMAG. Finally, the TMAG delivers to the MN the packets it receives from the new visited network. So, the MN continues receiving packets, as it moves from an old visited network to a new visited network. Consequently, the MN receives the first packets through the new visited network, while it is still attached to TMAG.

As the MN leaves the TMAG, it only undergoes an intra-network handover, because the TMAG also belongs to the new visited network. Thus, the handover mechanism with TMAG achieves zero handover delay for inter-network handover (the handover from one visited network to another one).

However, although the scenario with TMAG mitigates the handover delay, it achieves this at a cost of increased signalling overhead. For example, in updating the routing location information at the old visited network when the MN moves to the new visited network, only four messages are exchanged between the involved nodes in the scenario without TMAG; five messages are exchanged in the scenario with TMAG.

3.6.4 Simulation Results and Performance Analysis for Handover Latency and Packet loss

This sub-section investigates the improvement gains and constraints for introducing TMAG. This is analysed in terms of the internet-work handover delay and packet loss. Various factors, such as the traffic pattern and the MN speed are used to analyse the improvements and constraints.

3.6.4.1 Influence of Traffic Pattern on Internetwork Handover Latency and Packet Loss

Figure 3-16 and Figure 3-17 show the impact of traffic pattern on inter-network handover delay and packet loss for the configuration with TMAG and the configuration without TMAG. That is, the handover delay the MN experiences when it moves from one visited network to another. Various traffic patterns have been obtained by varying the time interval between the data packets. It is observed that the variation of traffic patterns has no effect on the handover delay for both configurations. This is because the handover delay is affected by the delay of mobility signalling.

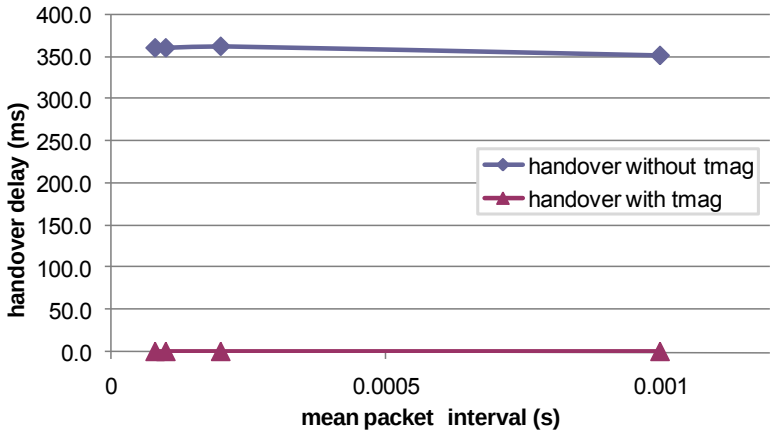


Figure 3-16 Impacts of traffic patterns on the handover delay

The configuration with TMAG has zero handover delay because during the handover between visited networks, the TMAG continues delivering to the MN the packets from the serving visited network (a visited network from which the MN is about to detach). Also, while the MN is still attached to the TMAG, the packets are forwarded from the serving visited network to the new visited network (a visited network to which the MN is going to handover), and then to TMAG – which then forwards the packets to the MN. Because the MN receives the first packets from the new visited network, while it is still attached to TMAG, the MN experiences no communication disruption, as it moves from one visited network to another. Moreover, when the MN moves away from TMAG to another MAG in the new visited network, it only experiences an intra-network handover because the TMAG is also a MAG that belongs to the new visited network. In contrast, when the MN performs handover in the configuration

without TMAG, the MN undergoes an inter-network handover. This handover incurs delays due to mobility signalling messages exchanged from the new visited network to the MN's communicating IP address-anchoring network, and then from this anchoring network to the old visited network. These mobility-related signals are exchanged, in order to allow packet forwarding from the old visited network to the new visited network; but the MN experiences a significant communication disruption, before it starts receiving its packets through the new visited network. This results in the increased handover delay. Moreover, this delay increases as the distance to the MN's communicating IP address anchoring network from the visited networks increases.

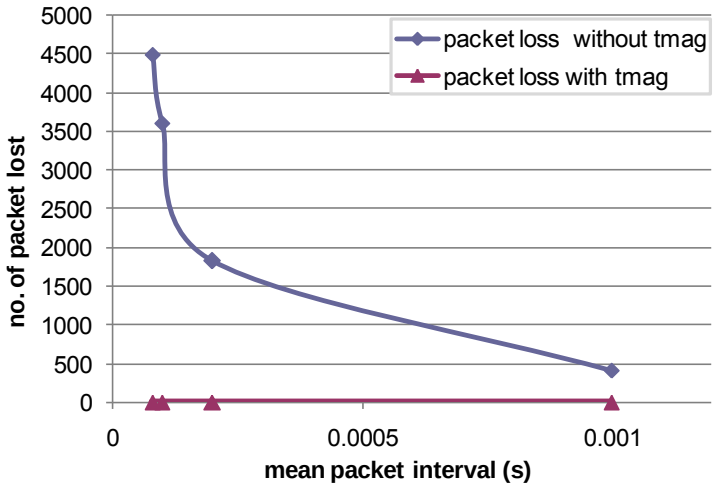


Figure 3-17 Impacts of traffic patterns on packet loss

On the other hand, the configuration without TMAG experiences high packet loss when the packet inter-arrival time is small. The reason is that a small inter-arrival time between packets indicates that the CN sends out packets at a higher rate. Many packets arrive during the handover period, when the MN is not able to receive packets, and are lost as a result of this, as is shown in Figure 3-17.

3.6.4.2 Impact of MN's Speed on Internetworks Handover Delay and Packet Loss

Figure 3-18 and Figure 3-19 show the impact of the MN's speed on the inter-network handover delay and the packet loss for both scenarios configured with TMAG and without TMAG. It is observed that the MN's speed does not affect the handover delay, and the packet

loss for the configuration without TMAG. The configuration with TMAG has zero handover delay; and it has zero packet loss, when the MN's speed is low. However, when the MN's speed is high (i.e., above 60m/s), both the handover delay and the packet loss increase. This is because when the MN is moving at a high speed, the MN crosses the overlapping region within a short time, (that is, before the TMAG is able to finish its operation). Consequently, when the MN attaches to a new visited network, the handover operation without TMAG is utilized, and the handover delay becomes increased, resulting in higher packet loss.

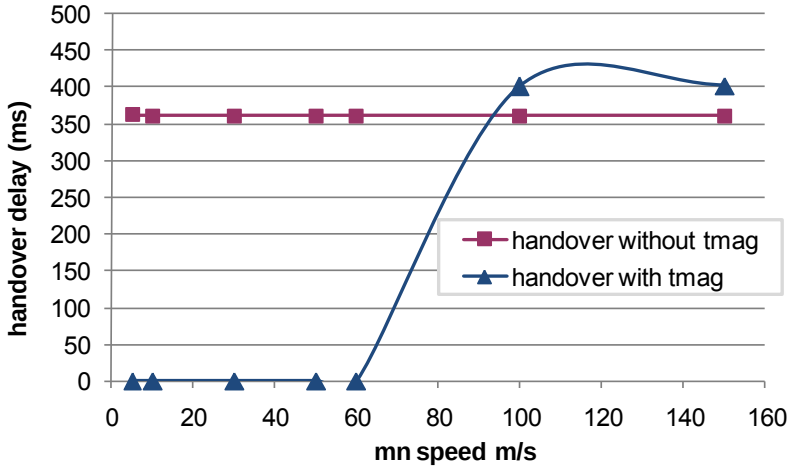


Figure 3-18 Impacts of MN's speed on the handover delay

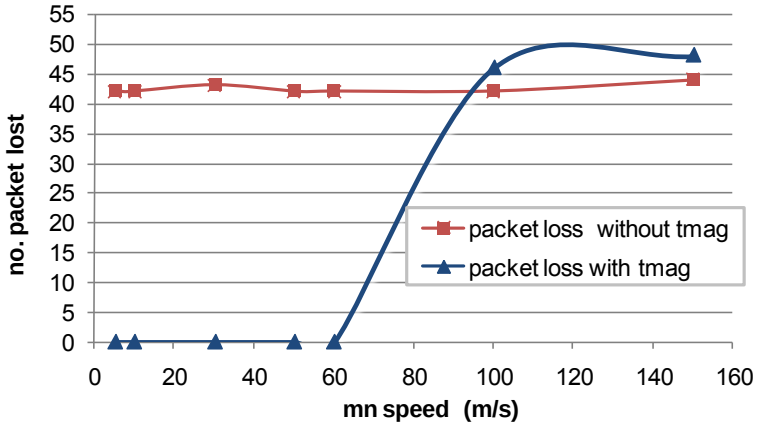


Figure 3-19 Impacts of MN's speed on packet loss

3.7 Comparative Qualitative Analysis of DM-RMG with Other DMM Schemes

DMM is a hot topic in both the research community and the IETF [21]; and some new schemes have already been proposed. In this section, DM-RMG scheme is compared to the most referenced schemes emanating from the research community.

In [60], the mobility management function of the home agent in MIPv6 is decomposed into the control plane and the data plane; and only the data plane is distributed at each access router. With respect to the concept of RM function, the scheme brings this function to the first hop router of the MN (i.e., at the access router). Both this scheme and DM-RMG split the mobility management functionality and place them in different network entities; and both schemes consider route optimization for ongoing communication. However, the scheme proposed in [60] is a host-based mobility management scheme; while the DM-RMG scheme proposed in this thesis is a network-based mobility management scheme. Moreover, the scheme in [60] establishes a direct communication between MN and CN – at the cost of jeopardizing the MN's location privacy, where the control plane node reveals the MN's current care-of-address to the CN, whereas the DM-RMG does not.

A network-based partially DMM scheme for PMIPv6 is presented in [57]. The scheme places the mobility routing function at each MAG; and it introduces a central mobility database node for location management. Unlike DM-RMG proposed in this thesis, this scheme does not implement route optimization for ongoing communication, which can lead to large end-to-end delay when the MN is far away from the traffic anchoring point with long-lasting traffic.

In [52][55], two mobility schemes that maintain the three mobility management functions on each access router, have been proposed. The two schemes bring the mobility management function closer to the MN, at the first hop router (the access router). In comparison, the DM-RMG proposed in this thesis places the RM function at the gateway router. DM-RMG and the scheme proposed in [52] do not need the MN to implement mobility client function; but [55] needs the MN to implement mobility client function. Moreover, the two schemes proposed in [52][55] experience long routes and high end-to-end delay for long-lasting traffic, as the MN moves far away from the initial point of its traffic establishment; whereas the DM-RMG

implements route optimization for ongoing communication, which reduces end-to-end delay.

3.8 Summary

This chapter has provided a detailed discussion on the design for a novel network-based DMM scheme that splits the mobility management function of an LMA in PMIPv6; and then distributes the routing management (RM) function at the gateway router of different networks. The scheme is named DM-RMG. The DM-RMG uses a mechanism to optimize the data path when an MN moves far away from its communicating IP address-anchoring network to a visited network, and this has also been discussed. Furthermore, the handover enhancement of DM-RMG using tracking MAG (TMAG) to provide seamless handover support for the MN, which performs subsequent handover between visited networks, has been presented.

The chapter has also explained the simulation environment and the implementation of the proposed DM-RMG scheme. The network simulator (ns-2), which is used for developing the simulation model and the performance evaluation of the proposed DM-RMG scheme has been briefly explained. The simulated network topology for the DM-RMG scheme and the simulation parameters haves also been discussed. Furthermore, the chapter has presented and analysed the results obtained from the simulation of DM-RMG model. A performance evaluation of DM-RMG has been given, in which DM-RMG has been compared with static traffic anchoring DMM schemes, as well as centralized IP mobility management schemes.

The performance analysis show that DM-RMG reduces the long packet delivery latency caused by triangle routing, when compared with static traffic anchoring DMM schemes. What this means is that network service providers can deploy DM-RMG to the gateways of their networks, in order to mitigate the effects of triangular routing, as the MN(s) roams between the different networks. The proposed scheme, incorporated with optimized handover mechanism, can achieve seamless handover, when an MN moves between the visited networks.

Chapter 4 Network-based Distributed Mobility Management for Network Mobility: NDM-RMG

4.1 Introduction

This chapter presents a new network-based distributed mobility support design for moving networks (i.e., a group of nodes moving together in a vehicle), which is named Network-based Distributed Mobility Management for Network Mobility (NDM-RMG). NDM-RMG extends DM-RMG (which has been presented in Chapter 3) to provide distributed mobility support for network mobility (NEMO). The motivation for the NDM-RMG design is given in this chapter. Additionally, the literature review on NEMO is also given. The chapter also discusses the design of NDM-RMG, together with the detailed principle of operation of NDM-RMG. Thereafter, it gives the analytical functions used to evaluate the performance of the proposed scheme, the simulation environment, and the simulation results. Moreover, the chapter discusses the simulation carried out using ns-2.

Finally, the results are presented in this chapter to show the effectiveness of the proposed scheme in terms of reducing packet overhead, packet delivery latency, the binding update cost, and the packet delivery cost.

4.2 Motivation and Design Approaches for NDM-RMG

As discussed in Chapter 1, the development of portable devices with wireless Internet access capability and the advancement in wireless technologies have made it possible for users to access Internet services anytime, anywhere, which produces more data traffic than ever. As a result, the number of mobile users is rapidly growing; and the volume of data traffic is also increasing at an exponential pace. There is also a growth in the number of users interested in accessing Internet services from moving vehicles, such as cars, trains, aircraft, and buses.

To support the mobility of a moving network, IETF has standardized Network Mobility (NEMO) Basic Support protocol (NBSP) [24]. NBSP is built on the principles of MIPv6. It comprises one or more mobile routers. A mobile router is the default gateway of a mobile network (known also as NEMO). Packets destined to the mobile network nodes (MNNs) in the mobile network are encapsulated through a bi-directional tunnel established between the mobile

router and its home agent (HA).

When a mobile network moves to a visited network, the mobile network may connect to another mobile network and form a nested NEMO. As the number of nested mobile networks increases, the pinball route problem occurs [69].

NBSP is based on MIPv6 principles; and it employs a centralized mobility management approach that leads to problems [14], such as:

(i) Sub-optimal routing because NBSP does not employ direct routing between an MNN and its correspondent node(s). The problem becomes severe when the NEMO has multiple nested mobile networks, which leads to longer packet delivery delay and high packet overhead – due to the pinball routing problem;

(ii) The HA maintains the mobility context of each mobile router and the MNNs; and the traffic destined to the MNNs is first routed to the HA for packet redirection. Thus, the scheme has scalability issues, and also may lead to a single point of failure (i.e., in the HA), which will cause a service outage for a large number of MNNs.

Therefore, NBSP cannot efficiently handle the increase in both the number of mobile users and the data traffic volume. Alternatively, the mobility management based on the DMM approach [14][12][5] can provide an efficient scheme for NEMO.

This chapter, therefore, develops novel network-based DMM schemes for non-nested and nested NEMO scenarios – with the goal of mitigating the sub-optimal routing, high packet overhead, and long packet delivery latency, especially in a nested NEMO scenario, due to the employment of a centralized mobility management approach in NBSP. The two schemes are based on the concept of decomposing the LMA logical functions in PMIPv6, as discussed in Chapter 3, and co-locating the RM with the gateway routers in different networks, so as to mitigate the pinball routing problem and the packet overhead in NEMO.

Additionally, the schemes co-locate a prefix delegating router (DR) function [70], with LM, in order to enable the delegation of a set of HNP (i.e., mobile network prefix, MNP) to a mobile router – for use inside its NEMO. Furthermore, the schemes extend the mobile router functionality by including the DR and Requesting Router (RR) functions [70]. The RR allows

the mobile router to get MNP that is topologically correct with respect to its currently attached infrastructure network, which is then used to assign IP prefix(es) to nodes inside NEMO. The DR function enables the parent mobile router to delegate a subset of these topologically correct prefixes to other mobile routers attached behind its network in nested NEMO.

These additional functions facilitate mitigating the pinball routing problem in nested NEMO basic support protocol. Moreover, the network-based features of the proposed schemes avoid packet tunnelling over the wireless link; thereby improve wireless link resource utilization.

4.3 Related Work

The NEMO Basic Support protocol (NBSP) [24] defines a mobile network as a network segment that can move and attach to an arbitrary point in the routing infrastructure. It is composed of at least one mobile router (MR) and multiple MNNs. The MR provides access to the MNNs for connecting to the external infrastructure; and it is in charge of the mobility management of MNNs inside the mobile network.

The NBSP requires a home agent (HA) similar to the one in MIPv6; the HA registers the mobile network's location and forwards packets to it. To achieve these functionalities, the HA performs three basic logical functions [24][23][50]: (i) Allocation of home address (HoA) to a mobile router (or an MNN); (ii) Location management (LM): managing and keeping track of the internetwork location for the mobile router (or MNN), which involves the mapping of the HoA to an address where the mobile router (or MNN) is reachable; and (iii) Routing management (RM): intercepting packets to/from the HoA of the MNN and forwarding the packets, based on the internetwork location information from the LM, either to the destination or to some other network element that knows how to forward them to the destination.

In the following paragraphs, the operation of NBSP with respect to the logical functions, are explained.

Figure 4-1 illustrates the NBSP architecture [24] in a three-network scenario, namely Net1, Net2, and Net3. Net1 is the home network of mobile network1 (NEMO1), a network from which the mobile router1 (MR1) obtains its HoA, i.e., P1::/mr1(HoA11), and a set of prefixes (also called mobile network prefixes [MNPs]), i.e.,P11::/64, for the nodes inside NEMO1. The MR1 uses the MNP, P11::/64, to assign IP address(es) to the MNNs inside NEMO1; and this

MNP remains unchanged when NEMO1 moves away from home.

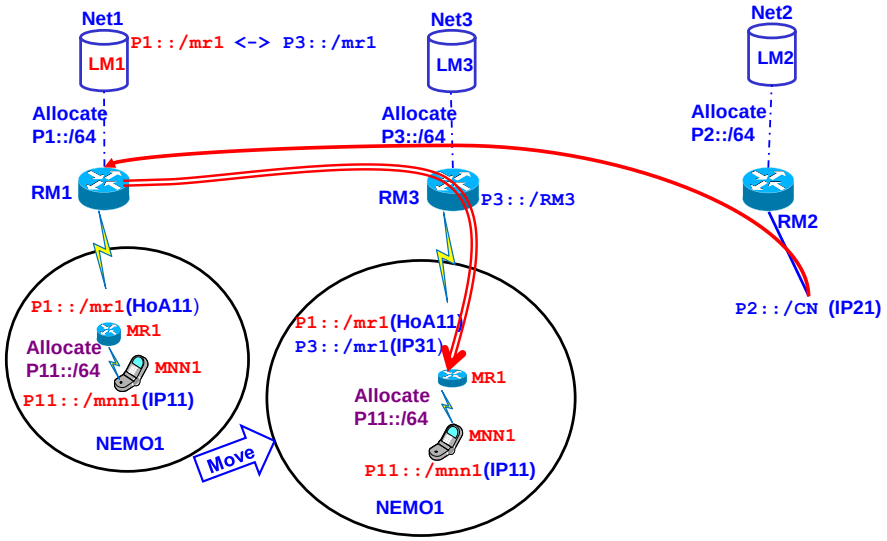


Figure 4-1 NEMO Basic Support Protocol Operation

As depicted in Figure 4-1, when NEMO1 moves away from home and attaches to a visited network, e.g., Net3, MR1 configures a care-of-address (CoA) from Net3, e.g., P3::/mr1(IP31). Thereafter, MR1 informs Net1 that it needs to update its current location, the CoA, through an exchange of binding update (BU) and binding acknowledgment (BA) messages, which may also include the MNP. The LM1 of Net1 then creates a binding entry that links MR1's HoA and MNP to the CoA. Subsequently, a bi-directional tunnel is established between RM1 and MR1, in order to carry the traffic for MNNs in NEMO1. After that, if for example CN in Net2 sends packets destined to MNN1, they are intercepted by RM1, and then tunnelled to MR1, based on the binding information stored in LM1.

Upon the packets' reception and de-capsulation, MR1 delivers the packets to MNN1. In general, all data packets from/to MNNs in NEMO1 are routed via RM1, which causes a triangular routing and tunnelling overhead to each transmitted data packet. It is important to note that these logical functions in NBSP, reside in a single entity (i.e., the HA).

The NBSP also considers cases in which a NEMO moves to a visited network and attaches to another NEMO, in order to reach the infrastructure. In this case, a nested mobile network is formed (i.e., nested NEMO), as shown in Figure 4-2 . In the figure, mobile routers MR1, MR4, and MR5 represent individual NEMOs; and their home networks are Net1, Net4,

and Net5, respectively. Each mobile router configures its HoA and obtains an MNP for its mobile network from their corresponding home networks. For example, MR5 has obtained P51::/64 prefixes from Net5 to be employed as the MNP.

The mobile networks in this example all move to Net3. During the attachments, MR1 connects to the infrastructure (i.e., RM3); MR4 connects to MR1; and MR5 connects to MR4. Hence, MR5 configures a CoA, P41::/mr5(IP41), from MR4's MNP. Similarly, MR4 configures a CoA, P11::/mr4(IP31), from MR1's MNP. As expected, MR1 also configures a CoA, P3::/mr1(IP31), assigned by Net3. After the configuration of the CoA, each mobile router proceeds to notify the new location to its home network. Despite the new addressing assignments, MNNs do not perceive any changes, since mobile routers continue advertising the same MNPs.

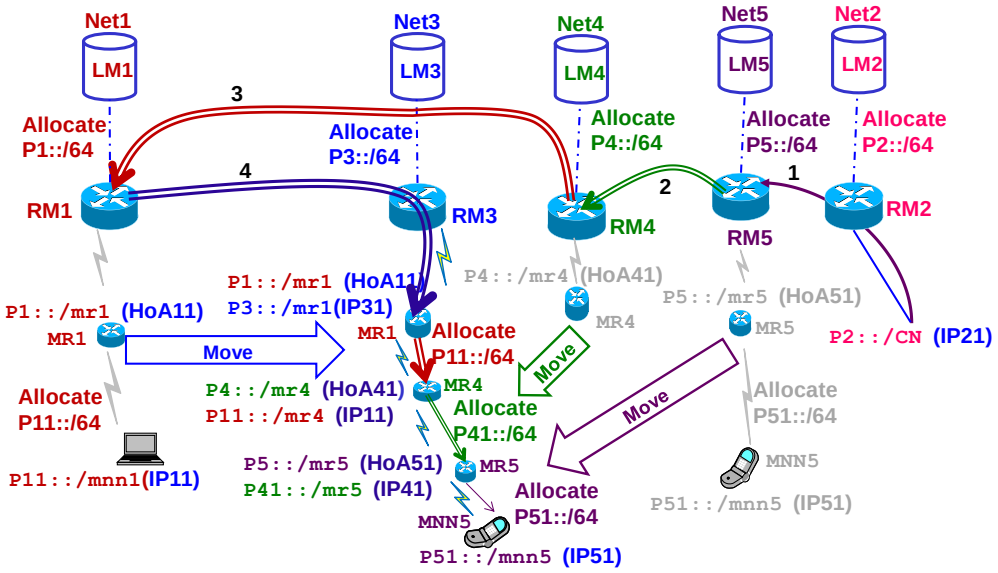


Figure 4-2 Pinball routing in Nested NEMO Basic Support Protocol

Let us assume that the CN in Net2 sends a packet destined to MNN5 in NEMO5. The packet first gets to Net5, where the IP prefix of NEMO5 is anchored (step1). LM5 then queries the binding information for NEMO5, to find that it is attached to NEMO4. Next, RM5 encapsulates the packet with the new IP destination P41::/mr4(IP41), and tunnels it to NEMO4's home network, i.e., Net4 (step 2). The same process is repeated at Net4, where the packet is encapsulated again to be forwarded to Net1, and from there to Net3 (steps 3 and 4). At Net3, the packet is forwarded to the final destination through the nested NEMO. As a result of the nested

scenario, the packet has to pass through three tunnels (steps 2 through 4), causing a high overhead due to the encapsulation process.

In addition to packet overhead, the path followed by the packet is long and sub-optimal, because of the pinball routing [71] that appears when packets traverse all the mobile routers' home networks before reaching their final destination. That is the effect of the centralized mobility management approach employed by NBSP, where all the mobility management logical functions are confined to the HA. Distributed Mobility Management (DMM) is an approach that can address the problems identified in NBSP [5][12]. DMM distributes the logical functions to different networks elements, while bringing them closer to the mobile nodes. Based on the concept of DMM, two DMM schemes have been proposed for NEMO [72][73].

In [72], a distributed mobility scheme based on NEMO for MIPv6 network has been proposed. The scheme is developed for both non-nested and nested NEMO. The HA in MIPv6 is distributed to each access router, the default gateway of NEMO. Every time a NEMO moves to a different network, the mobile router configures a new address, registers it with the HA where the traffic is anchored, and obtains a mobile network prefix from the new network. New IP sessions use the MNP assigned in the new network. However, the downside of this scheme is that the tunnel created over the wireless link for ongoing connections has a negative impact on the wireless bandwidth efficiency. Moreover, an ongoing long-lasting traffic may be subject to long routes because of the static traffic anchoring (e.g., long-range movement using vehicles that cover across multiple anchors).

Another distributed mobility scheme to support mobile networks in vehicular scenarios and flat network architectures has been proposed in [73]. The scheme employs network-based mobility, and combines the local mobility anchor (LMA) and mobile access gateway (MAG) functions of Proxy MIPv6 in each access router. The scheme introduces a proxy router to manage the MNs mobility in the mobile network, and a central session database to facilitate the reachability of the MNs and the proxy router. The MNs configure new addresses from the different networks that the mobile network visits; and it uses the new address for new IP sessions. However, the scheme has not covered the nested NEMO problem.

4.4 Architecture of the Proposed NDM-RMG Scheme

4.4.1 Architecture Overview

This section presents the proposed network-based DMM schemes for non-nested NEMO and nested NEMO scenarios (non-nested NDM-RMG and nested NDM-RMG). Considering the DMM requirement of re-using existing mobility protocols [14], NDM-RMG schemes decompose the logical functions of the LMA in PMIPv6 [17] to location management (LM), routing management (RM), and home network prefix (HNP) allocation functions, in a similar way to the DM-RMG that was developed in Chapter 3; however, DM-RMG does not address mobility for moving networks (NEMOs). In addition, NDM-RMG schemes co-locate the RM function at the gateway of each network, which in a flat network architecture may coincide with the access router. In this way, the data-plane routing function for mobile network nodes is served by the local RM at the network gateway.

Figure 4-3 shows an example of the architecture of non-nested NDM-RMG, which comprises a large domain divided into three networks: Net1, Net2, and Net3. Each network has the mobility management functions distributed as follows: (i) An LM server; (ii) an RM co-located with the gateway router (GW); and (iii) multiple access routers with mobility client functions equivalent to the MAGs in PMIPv6. Each network owns a unique IP prefix block from which it delegates a set of prefixes to the attached mobile networks. In this example, Net1 owns P1:: /64 prefixes; Net2 owns P2::/64 prefixes; and Net3 owns P3::/64 prefixes. Figure 4-3 shows an example of the mobile network NEMO1 attached to its home network, Net1.

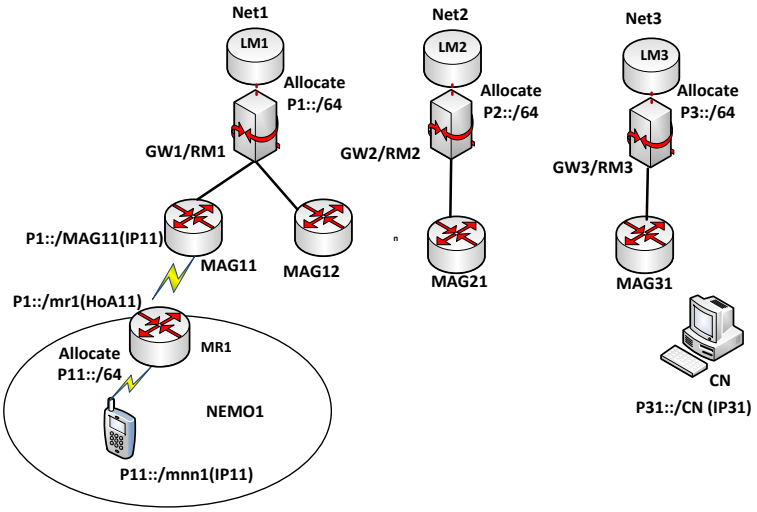


Figure 4-3 NDM-RMG architecture with RM co-located at the gateways for non-nested NEMO

Each location management server, LM1, LM2, and LM3, is co-located with the HNP allocation function, so that it delegates, upon request, a set of prefixes to the mobile router(s) attached to its network. The servers also maintain the mapping between the HNP and the IP address of the RM located at the visited network, where the NEMO is attached. For example, if NEMO1 in Figure 4-3 moves and attaches to Net2, LM1 keeps the mapping between the MR1's HNP and the address of RM2. These LM servers constitute a distributed database of LMs in the domain; and they can be either virtually or physically deployed.

On the other hand, the RM keeps the binding information between the mobile router's HNP and the address of the MAG that is currently serving the mobile router. To achieve an internetworking mobility routing, each RM interacts with the LM to retrieve the location information of the mobile router attached to its network. Moreover, the data plane routing functions for the MNNs are performed by the RM co-located at the gateway in the visiting network. This alleviates the need for MNN's traffic to traverse the mobile router's home network, and thereby addresses the pinball routing problem.

As NDM-RMG is based on PMIPv6, which has no mobility support for mobile networks, NDM-RMG co-locates the Delegating Router (DR) function with the LMs, so they can delegate IP prefixes to the mobile router(s), in order to be assigned to the MNNs. NDM-RMG also extends the mobile router functionality with the DR function, as well as with a Requesting

Router (RR) function, as defined in [70]. The RR allows the mobile router to obtain a topologically correct IP prefix, which is then advertised to the mobile network nodes. The DR function enables the parent mobile router to delegate a subset of these topologically correct prefixes to mobile routers attached to the mobile network, in the case of nested NEMO. Moreover, every mobile router detects the attachment of the MNNs to its access network, in the same way that a MAG detects the MN attachment in PMIPv6.

Every MAG is collocated with the DHCPv6 relay function [74], in order to handle the prefix request message sent from a mobile router. As a result, the MAGs are able to relay prefix request messages to the DR in their networks. Apart from the DHCPv6 relay functions, the MAGs behave in the same manner as in PMIPv6; hence, they can perform mobility management signalling on behalf of the mobile routers that attach to the access networks.

The following sub-section describes in detail the operation and signalling mechanisms of NDM-RMG – for both non-nested and nested NEMO scenarios.

4.4.2 Mobile Router Registration and Prefix Acquisition Procedures

The signalling call flow for the attachment and registration procedures for a mobile router (MR1) is illustrated in Figure 4-4. In the figure, the NDM-RMG architecture illustrated in Figure 4-3 is used to describe the registration and the prefix acquisition procedures. When MAG11 detects MR1's attachment, it obtains the MR1's identifier, MR1-id. Next, MAG11 notifies RM1 about MR1's attachment by means of a PBU message. RM1 queries LM1, and then latter allocates a HNP to MR1, i.e., P1::/64(mr1). After that, RM1 stores the mapping between the assigned prefix and MAG11's IP address; this is followed by the sending of a PBA to MAG11 [17]. Upon PBA's reception, MAG11 records the entry for P1::/64(mr1) and MR1-id in its binding update list (BUL), followed by the sending of router advertisements (RA) messages to advertise the allocated prefix to MR1. Lastly, MR1 uses the advertised prefix to configure its HoA, P1::/mr1(HoA11).

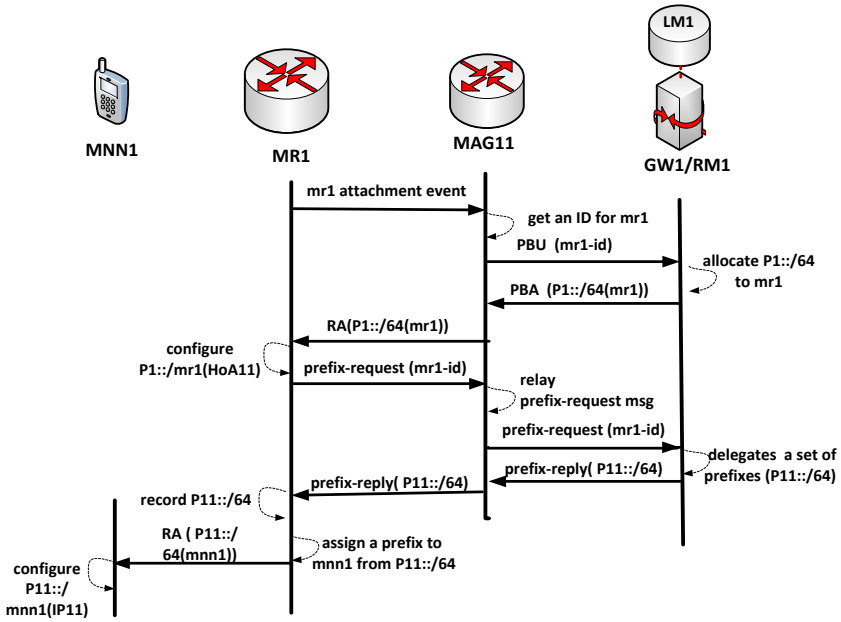


Figure 4-4 Mobile router and MNN attachment and registration in the NDM-RMG domain

After MR1 has successfully configured its home address, it runs the router requesting function to send a prefix-request message to the DR (collocated with the LM1). Once the prefix request message is received by MAG11, it proceeds to relay the message to RM1. RM1 interacts with LM1 and the DR at LM1 allocates a prefix that is a subset of MR1's HNP (i.e., P11::/64). Thereafter, RM1 forwards the delegated prefix to MR1 in a prefix-reply message, which is then relayed by MAG11 to MR1. At MR1, the prefix-reply message reception triggers the creation of a prefix pool for the handling of the prefixes. At this point, MR1 has topologically correct prefixes to be advertised to the mobile network nodes.

When the MNN1 attaches to MR1, MR1 detects the attachment and allocates a prefix, P11::/64(mnn1), from its prefix pool. It also creates an entry that associates this prefix with the MNN1-id. Lastly, MR1 sends an RA message to advertise the prefix to MNN1. Once MNN1 configures a valid IP address, it may initiate an IP session with the CN.

4.4.3 Handover Mechanism for a Non-Nested NDM-RMG Scenario

Again, the example of the network architecture, as presented in Figure 4-3, is considered in the following discussion. Let us assume that NEMO1 has moved from Net1 to Net2, while the

MNN1is still having an active communication with CN in Net3. This is illustrated in Figure 4-5, which also shows the detailed handover operation with a signalling call flow.

When NEMO1 attaches to Net2, MAG21 and RM2 perform the regular exchange of PBU/PBA messages to assign and later advertise a HNP to MR1, i.e., P2::/64(mr1).

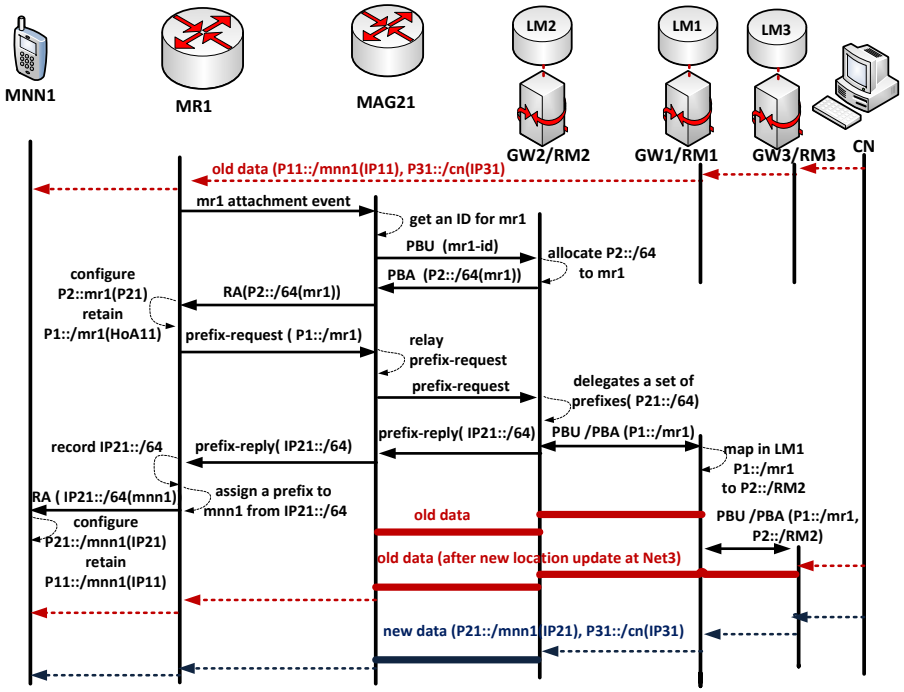


Figure 4-5 Mobile network handover signalling call flow from Net1 (RM1) to Net2 (RM2) for non-nested NDM-RMG

When MR1 receives the RA message, it realizes there has been a change of access network; MR1 proceeds to configure an address from the new prefix, e.g., P2::/mr1(IP21); while at the same time, it keeps the old address configured from Net1, in order to maintain the ongoing session. The router-request/router-reply messages are also exchanged for the MR1 to obtain prefixes for the MNNs. However, in this scenario, the prefix-request message also carries the MR1's old address. This information serves two purposes: (i) To enable the MAG21 to create a mapping between the old and new addresses of MR1, in order to allow for the forwarding of traffic towards MR1 for ongoing sessions; and (ii) to allow the RM2 to learn about the prefix from the network from which MR1 has detached.

When RM2 learns about such a prefix, it records the mapping between MR1's old address and MAG21 address, so as to facilitate routing for MNNs' active session(s). RM2 also

interacts with LM2, in order to locate the network to which MR1's old address belongs. Since each network owns unique IP prefixes, LM2 knows that the owner of MR1's old prefix information is LM1 (with the help of the LM's distributed database). This allows RM2 to learn about RM1 for tunnel establishment. Subsequently, RM2 communicates with RM1 to establish the bi-directional tunnel that will carry the ongoing session(s) for mobile network nodes attached to MR1. Such a communication employs a modified PBU, which includes RM2's address, MR1's old prefix, and an LM-flag. Upon reception, LM1 stores the mapping between MR1's old prefix and RM2's address.

When MR1 receives the prefix-reply message from RM2, the assigned prefix is marked as new in the prefix pool, whereas the old prefix is marked as old. After that, MR1 advertises the new prefix to MNN1 (i.e., P21::/64(mnn1)). MNN1 configures a new address from P21::/64(mnn1); but it also maintains the address from the old prefix, P11::/mnn1(IP11). The new address is used to establish new IP sessions to/from MNN1; whereas the old address is used for the continuity of ongoing active sessions.

After the configuration is completed, when RM1 receives packets for MNN1 (i.e., destined to P11::/mnn1(IP11)), it retrieves the MNN1's location from LM1, which should point towards the RM2's address. Thereafter, RM1 updates its routing entry with this mapping information, so that incoming packets are encapsulated towards RM2, and then de-capsulated at RM2 for tunnelling towards MAG21. In the same way, MAG21 de-capsulates the packets and forwards them through the interface to which MR1 is attached. Thus, packets are delivered to MR1 without tunnelling over the wireless link. MR1 then forwards the packets to their final destination, MNN1.

The routing for ongoing sessions can be further improved if RM1 informs RM3 (i.e., CN's home network) about MNN1's new location. In this way, packets destined to MNN1's old address are encapsulated from Net3 to Net2, thereby completely bypassing Net1. This alleviates the pinball routing problem.

In the case of new IP sessions, MNN1 uses the newly configured IP address, P21::/mnn1(IP21) to establish such communications.

4.4.4 Handover Mechanism for a Nested NDM-RMG Scenario

Figure 4-6 shows the architecture for a nested NDM-RMG scenario with four networks, Net1, Net2, Net3, and Net4. In the figure, NEMO1 has moved from Net1 to Net2, and NEMO4 has moved from Net4 to Net2, forming a nested NEMO (i.e., MR4 attaches to MR1). After NEMO1's movement, MR1 configures an address from P2::/mr1(IP21), and obtains a set of IP prefixes, e.g., P21::/64, to be used inside the mobile network. When NEMO4 moves to Net2 and attaches to MR1, MR1 detects the attachment and advertises the prefix P21::/64 to MR4. Next, MR4 configures an IP address, e.g., P21::/mr4(IP21), while retaining its home address, P4::/mr4(HoA41).

After IP address configuration, MR4 sends the prefix-request message to MR1, so as to obtain a set of prefixes for its own mobile network. The message includes MR4's home address information, P4::/mr4(HoA41), in order to allow RM2 to learn about Net4, and to be able to notify Net4 about the attachment of NEMO4. When MR1 receives the prefix-request message, it acts as a DR and delegates a topologically valid set of prefixes to MR4, i.e., P211::/64. Meanwhile, MR1 also sends a modified PBU to MAG21, namely a nested PBU (nPBU), which informs Net2 about the attachment of MR4 as a mobile network. The message includes options for MR1's HNP, P2::/mr1(IP21), and the old address of MR4, P4::/mr4(HoA41).

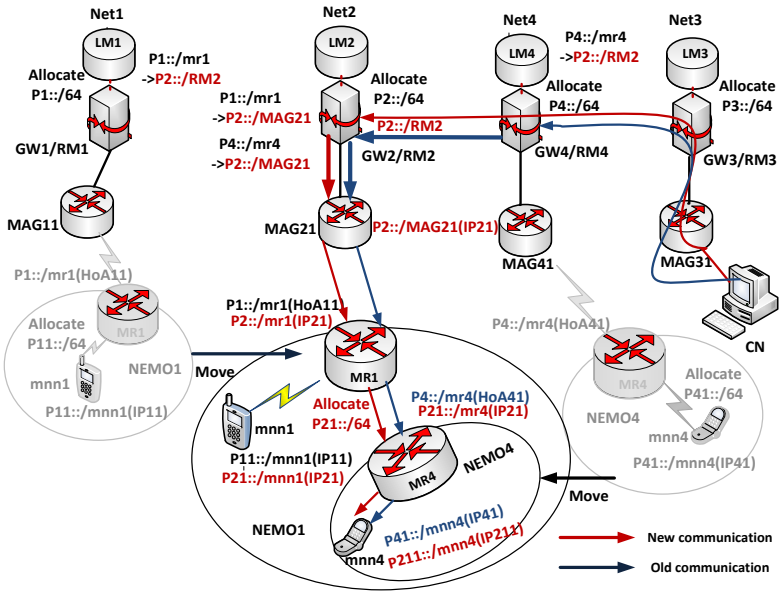


Figure 4-6 A NDM-RMG framework with RM co-located at the GWs for nested NEMO

When MAG21 receives the nPBU, it stores the mapping between P4::/mr4(HoA41) and P2::/mr1(P21), and then relays the message to RM2. Upon receiving the message, RM2 records the mapping between P4::/mr4(HoA41) and the MAG21 address, as shown in Figure 4-6. With the help of LM2's distributed LMs database, RM2 locates Net4 as the owner of MR4's address, P4::/mr4(HoA41). Thus, RM2 notifies RM4 about the new location of MR4, and the latter requests LM4 to create an association between P4::/mr4(HoA41) and RM2's address, before acknowledging to RM2. Next, RM2 responds to MR1 with an nPBA, which confirms the successful tunnel establishment in the direction of (towards) Net4. The nPBA message is then relayed by MAG21 to MR1. Upon receiving the message, MR1 records the mapping between MR4's old address, P4::/mr4(HoA41), and MR4's HNP, P21::/64(mr4).

When MR4 receives the prefix-reply message, it records the prefix in the prefix pool, as discussed in sub-section 4.4.3. Such a prefix can be assigned to mobile network nodes in NEMO4. For example, MR4 advertises P211::/64(mnn4), and MNN4 configures the address P211::/mnn4(IP211), (while keeping P41::/mnn4(IP41) for the ongoing communications). If MNN4 establishes a new IP session, it uses P211::/mnn4(IP211), so as to not traverse its home network, Net4. In the case of the ongoing sessions, these sessions traverse Net4, but avoid passing through Net1. This is not what happens in NEMO Basic Support in nested NEMO scenarios [24]. The proposed scheme may further reduce the pinball routing problem by allowing RM4 to inform RM3 about the address of RM2. Subsequently, RM3 tunnels packets for MNN4 directly to RM2; because of this, the packets no longer traverse Net4, which also reduces the processing burden on the network.

4.5 Analytical Performance Evaluation

In this section, the performance of NDM-RMG is analysed and compared with NBSP [24] and NEMO-based DMM (N-DMM) [72]. N-DMM is compared with NDM-RMG because both schemes use DMM concepts to address NEMO problems, especially for the complex nested NEMO scenario. Compared to N-DMM, NDM-RMG decomposes and distributes PMIPv6 functionality, whereas N-DMM distributes the HA of MIPv6.

In the performance evaluation, analytical equations for end-to-end delay, packet delivery cost and binding update cost are developed and used in the analysis. Also the packet overhead is used as an additional performance metric in the analysis. The analysis focuses on the nested

NDM-RMG scenario, where the pinball routing problem is most prevalent, because the major focus of NDM-RMG scheme is to mitigate the pinball routing problem in NEMO. In the analysis, the case of one nesting level presented in Figure 4-6 is first considered; and thereafter, more than one nesting level is considered, so as to examine the performance gain of the NDM-RMG scheme. For easy comparison with NBSP, the HAs in NBSP are assumed to be located in the same place as RMs in NDM-RMG. Similarly, it is assumed that access routers (ARs) in NBSP and HAs in N-DMM are positioned at the same place as MAGs in NDM-RMG.

4.5.1 Packet Overhead Analysis

Packet overhead is defined as the amount of control information carried on a data packet to facilitate mobility management. The additional overhead has an impact on the protocol performance, as it causes additional processing delay, and wastes wireless bandwidth, which is scarce and expensive. An overhead of 40 bytes per packet is used in the analysis. It is assumed that the nested mobile network node (e.g., MNN4) has ongoing sessions, namely the handoff traffic, before its mobile network performs a handover.

Following the example shown in Figure 4-6, if NBSP is employed, packets are first encapsulated by HA_Net4, and then by HA_Net1, which leads to a total of 80 bytes extra overhead in the path between CN and MR1. In contrast, N-DMM builds a tunnel between the HA located at MAG41 and MR4, so it incurs an extra overhead of 40 bytes due to packet encapsulation in the link between MAG41 and MR4. In these schemes, the overhead is carried over the wireless link towards MR1. However, NDM-RMG encapsulates packets with 40 bytes extra overhead in the wired link between RM4 and RM2, and then between RM2 and MAG41. MAG41 forwards the packet to MR1 without any extra overhead.

With an m-level of nesting, NBSP and N-DMM have an extra overhead of $40(m+1)$ bytes and 40 bytes, respectively, over the wireless link. In contrast, NDM-RMG has zero byte overhead on the wireless link. Thus, NDM-RMG performs better in terms of packet overhead carried over the wireless link. However, NDM-RMG incurs 40 bytes extra packet overhead in the wired link, the same as N-DMM.

4.5.2 End-to-end Packet Delay Analysis

The end-to-end latency analysis is performed for the traffic established by MNN4 before NEMO4 performs the handover to Net2. The end-to-end latency is defined as the time elapsed from when the packet is transmitted from CN's network until it is received by MNN4. It involves the propagation delay incurred in the various links through which the packet passes. The end-to-end latency analysis for NBSP, N-DMM, and NDM-RMG (with and without route optimization, RO), is calculated according to (4.1) through (4.4), based on the scenario presented in Figure 4-6. The notations and their corresponding parameter values are presented in Table 4-1.

$$T_{NBSP} = t_{RM3/HA3-RM4/HA4} + t_{RM4/HA4-RM1/HA1} + t_{RM1/HA1-MAG21/AR21} + t_{MAG21/AR21-mr1} + t_{mr1-mr4} + t_{mr4-mnn4} \quad \textbf{(4.1)}$$

$$T_{N-DMM} = t_{MAG31/HA31-MAG41/HA41} + t_{MAG41/H41-MAG21/HA21} + t_{MAG21/HA21-mr1} + t_{mr1-mr4} + t_{mr4-mnn4} \quad \textbf{(4.2)}$$

$$T_{NDM-RMG_{noRO}} = t_{RM3/HA3-RM4/HA4} + t_{RM4/HA4-RM2/HA2} + t_{RM2/HA2-MAG21/AR21} + t_{MAG21/AR21-mr1} + t_{mr1-mr4} + t_{mr4-mnn4} \quad \textbf{(4.3)}$$

$$T_{NDM-RMG_{RO}} = t_{RM3/HA3-RM2/HA2} + t_{RM2/HA2-MAG21/AR21} + t_{MAG21/AR21-mr1} + t_{mr1-mr4} + t_{mr4-mnn4} \quad \textbf{(4.4)}$$

Table 4-1 Parameter notations and values [75][76]

Parameter Notation	Meaning	Value
$\mathbf{S_{bu}}$	BU message size (bytes)	72
$\boldsymbol{S_{pbu}, S_{npbu}}$	PBU message size (bytes)	76
$\boldsymbol{S_{prm}}$	Prefix (DHCPv6) request message size (bytes)	96
$\boldsymbol{t_{RM/HA-RM/HA}, t_{HA-HA}}$	Link delay between HAs, RMs (ms)	100
$\boldsymbol{t_{cn-RM/HA}}$	Link delay between CN/CN network and HA , RM (ms)	100
$\boldsymbol{t_{RM/HA-AR}}$	Link delay between HA and AR (ms)	10-100
$\boldsymbol{t_{mr-mr}}$	Link delay between MRs (mobile routers)	5

	(ms)	
$t_{AR-mr}, t_{MAG-mr}, t_{MAG/HA-mr}$	Link delay between MAG/AR and MR (ms)	5
t_{mr-mnn}	Link delay between MR and MNN (ms)	5
t_{RM-MAG}	Link delay between RM and MAG (ms)	15
$t_{MAG/AR-MAG/AR}, t_{MAG/HA-MAG/HA}$	Link delay between MAGs/ARs (ms)	10
$p_{RM/HA}, p_{MAG}, p_{mr}$	Processing delay of HA, RM, MAG, MR (ms)	10
m	Number of levels of nesting in nested NEMO	

In order to evaluate the performance of the developed scheme, in terms of reducing the pinball routing problem, the above analysis is extended to an m-level of nesting. The processing delay caused by various mobility management elements through which the packet has to traverse is also considered in the evaluation. It is assumed that CN is sending packets to a mobile network node (MNN) located in an m-level nested mobile network. Thus, the end-to-end latency for NBSP can be expressed according to (4.5).

$$T_{NBSP} = \sum_{j=1}^{m+1}(p_{RM/HA}^{j} + p_{mr}^{j}) + \sum_{j=1}^{m}(t_{RM/HA-RM/HA}^{j} + t_{mr-mr}^{j}) + t_{cn-RM/HA} + t_{RM/HA-AR} + t_{AR-mr} + t_{mr-mnn} \quad \textbf{(4.5)}$$

N-DMM and NDM-RMG bypass the route used by NBSP, and their end-to-end latencies are given by (4.6), (4.7) and (4.8), respectively.

$$T_{N-DMM} = t_{cn-MAG/HA} + p_{MAG/HA} + t_{MAG/HA-MAG/HA} + t_{MAG/HA-mr} + \sum_{j=1}^{m+1} p_{mr}^{j} + \sum_{j=1}^{m} t_{mr-mr}^{j} + t_{mr-mnn} \quad \textbf{(4.6)}$$

$$T_{NDM-RMG_{noRO}} = t_{cn-RM/HA} + 2p_{RM/HA} + t_{RM/HA-RM/HA} + t_{RM/HA-MAG} + p_{MAG} + t_{MAG-mr} + \sum_{j=1}^{m+1} p_{mr}^{j} + \sum_{j=1}^{m} t_{mr-mr}^{j} + t_{mr-mnn} \quad \textbf{(4.7)}$$

$$T_{NDM-RMG_{RO}} = t_{cn-MAG} + t_{MAG-RM/HA} + t_{RM/HA-RM/HA} + 2p_{RM/HA} + t_{RM/HA-MAG} + p_{MAG} + t_{MAG-mr} + \sum_{j=1}^{m+1} p_{mr}^{j} + \sum_{j=1}^{m} t_{mr-mr}^{j} + t_{mr-mnn} \quad \textbf{(4.8)}$$

4.5.3 Packet Delivery Cost Analysis

As the mobile network becomes nested, the amount of encapsulation grows with the increase in the number of nesting levels. These multiple encapsulations add extra bits that cause additional transmission and processing costs, due to the various mobility entities traversed by the packet. The packet delivery cost (PDC) is calculated as the sum of transmission cost and processing cost. The transmission cost is computed as the product of the data packet size and the link latency of the various link types the packets traverses. The IPv6 header of 40 bytes is used in this analysis. Thus, considering the schemes' operational mechanisms, and using (4.5) through (4.8), the packet delivery cost for each scheme can be derived, as given in (4.9) to (4.12).

$$PDC_{NBSP} = \sum_{j=1}^{m+1}(p_{RM/HA}^{j} + p_{mr}^{j}) + \sum_{j=1}^{m}((S_{data} + 40j)\cdot t_{RM/HA-RM/HA}^{j} + (S_{data} + 40(m-j+1))\cdot t_{mr-mr}^{j}) + t_{cn-RM/HA}\cdot S_{data} + (S_{data} + 40(m+1))\cdot(t_{RM/HA-AR} + t_{AR-mr}) + t_{mr-mnn}\cdot S_{data} \quad \textbf{(4.9)}$$

$$PDC_{N-DMM} = t_{cn-MAG/HA}\cdot S_{data} + p_{MAG/HA} + (S_{data} + 40)\cdot t_{MAG/HA-MAG/HA} + (S_{data} + 40)\cdot t_{MAG/HA-mr} + \sum_{j=1}^{m+1} p_{mr}^{j} + \sum_{j=1}^{m}(S_{data} + 40)\cdot t_{mr-mr}^{j} + t_{mr-mnn}\cdot S_{data} \quad \textbf{(4.10)}$$

$$PDC_{NDM-RMG_{noRO}} = t_{cn-RM/HA}\cdot S_{data} + 2p_{RM/HA} + (S_{data} + 40)\cdot t_{RM/HA-RM/HA} + (S_{data} + 40)\cdot t_{RM/HA-MAG} + p_{MAG} + t_{MAG-mr}\cdot S_{data} + \sum_{j=1}^{m+1} p_{mr}^{j} + \sum_{j=1}^{m} t_{mr-mr}^{j}\cdot S_{data} + t_{mr-mnn}\cdot S_{data} \quad \textbf{(4.11)}$$

$$PDC_{NDM-RMG_{RO}} = t_{cn-MAG}\cdot S_{data} + (S_{data} + 40)\cdot t_{MAG-RM/HA} + (S_{data} + 40)\cdot t_{RM/HA-RM/HA} + 2p_{RM/HA} + (S_{data} + 40)\cdot t_{RM/HA-MAG} + p_{MAG} + t_{MAG-mr}\cdot S_{data} + \sum_{j=1}^{m+1} p_{mr}^{j} + \sum_{j=1}^{m} t_{mr-mr}^{j}\cdot S_{data} + t_{mr-mnn}\cdot S_{data} \quad \textbf{(4.12)}$$

4.5.4 Binding Update Cost Analysis

When a mobile network moves and attaches to a visited network through either a direct attachment or another mobile network(s), the mobile router of that network needs to send a BU message to the mobility anchor in the home network (or traffic anchoring node) to inform about its present location. Network resources, such as transmission and processing power consumed by the BU messages comprise the binding update cost (BuC). The BuC is defined as the product of the BU message size and the link delay of the various link types traversed by the message. The BuC caused by the deepest mobile router in an m-level nested scenario is investigated. Equations (4.13) to (4.16) show the BuC incurred in transferring the BU packet from the deepest mobile router (MR) at the m-level of nesting. In NBSP, each MR traversed in the path from the deepest MR toward the HA encapsulates the packet to its corresponding HA with an overhead of 40 bytes. The BuC of the schemes is derived, according to equations (4.13) to (4.16).

$$BuC_{NBSP} = \sum_{j=1}^{m}(S_{bu}+40(j-1))\cdot t^{j}_{mr-mr} + (S_{bu}+40m)\cdot(t_{RM/HA-AR}+t_{AR-mr}) + \sum_{j=1}^{m}(S_{bu}+40(m-j))\cdot t^{j}_{RM/HA-RM/HA} \quad \textbf{(4.13)}$$

$$BuC_{N-DMM} = (t_{MAG/HA-MAG/HA}+t_{MAG/HA-mr})\cdot S_{bu} + \sum_{j=1}^{m} S_{bu}\cdot t^{j}_{mr-mr} \quad \textbf{(4.14)}$$

$$BuC_{NDM-RMG_{noRO}} = S_{prm}\cdot t_{mr-mr} + \sum_{j=2}^{m} S_{npbu}\cdot t^{j}_{mr-mr} + S_{npbu}\cdot t_{mr-MAG} + S_{npbu}\cdot t_{MAG-RM} + S_{pbu}\cdot t_{RM/HA-RM/HA} \quad \textbf{(4.15)}$$

$$BuC_{NDM-RMG_{RO}} = S_{prm}\cdot t_{mr-mr} + \sum_{j=2}^{m} S_{npbu}\cdot t^{j}_{mr-mr} + S_{npbu}\cdot t_{mr-MAG} + S_{npbu}\cdot t_{MAG-RM} + 2S_{pbu}\cdot t_{RM/HA-RM/HA} + S_{pbu}\cdot t_{cn-MR} \quad \textbf{(4.16)}$$

4.5.5 Numerical Results and Discussion

In this sub-section, the performance of the proposed schemes is evaluated, based on the analytically derived equations. The numerical results are generated in MATLAB (R2009b),

using the parameter values, as given in Table 4-1.

Figure 4-7 shows the impacts of the different levels of nesting on the end-to-end latency of the three schemes. The resulting end-to-end latency for NBSP grows quickly, because the packet traverses the HAs of parent mobile routers, as given in (4.1) and (4.5). In contrast, both NDM-RMG and N-DMM bypass the home networks of the parent mobile routers, as given in (4.2), (4.3), (4.4), (4.6), (4.7) and (4.8). Thus, both schemes achieve a smaller end-to-end latency compared with NBSP; hence, they mitigate the effect of pinball routing. The end-to-end latency for N-DMM with a link delay of 10ms between HAs/ARs is slightly better than that of the proposed scheme with RO. However, if the MNN has a long-lasting traffic session, and its mobile network moves far away from the traffic anchoring point (as shown in Figure 4-7 with a link delay of 100ms), subsequently, the N-DMM scheme will result in a higher end-to-end latency compared to NDM-RMG with RO, because N-DMM employs a static traffic anchoring mechanism.

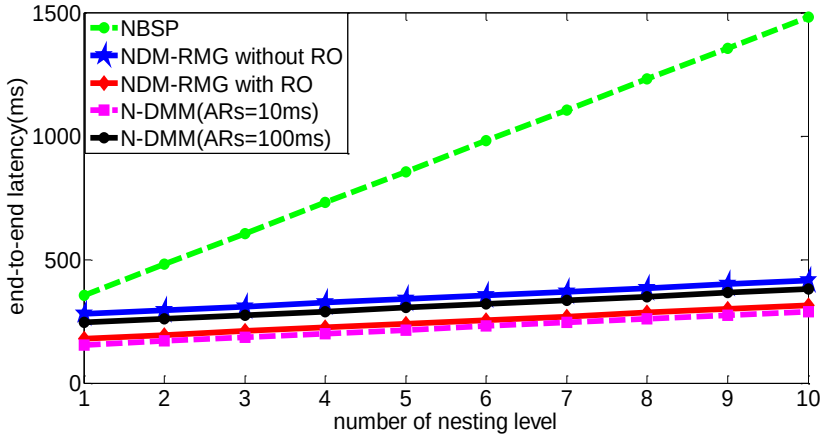


Figure 4-7 End-to-end delay with different levels of nesting

Figure 4-8 shows the trends of packet delivery cost for various levels of nesting. A data packet size (S_{data}) of 50 bytes is used for the calculations. It can be seen that NBSP has a high PDC compared with other schemes, because it suffers from the pinball routing problem; so the packets encounter multiple encapsulations through a long path, and many processing delays. Comparing NDM-RMG with N-DMM, NDM-RMG has a slightly higher PDC. However, when the mobile network moves far away from the traffic anchoring point, the PDC for N-DMM becomes high, (as is shown in the figure for a link delay of 100ms). This is due to N-DMM not optimizing the route for ongoing traffic.

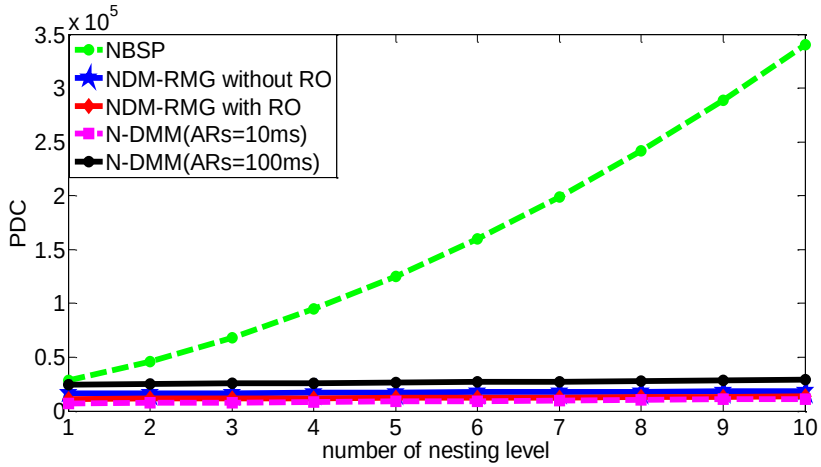


Figure 4-8 Packet delivery cost with different levels of nesting

Figure 4-9 illustrates the impact of different degrees of level of nesting on BuC. The BuC for NBSP is high; and it increases very quickly compared with other schemes. This is attributed to: (1) The encapsulation of the BU messages by each mobile router in the path between the deepest mobile router and their corresponding HAs; and (2) the link delay between HAs. On the other hand, the 'NDM-RMG scheme with RO' has a slightly higher BuC when compared with the N-DMM scheme, and also when compared with 'NDM-RMG without RO'. This is due to the additional BU messages needed by 'NDM-RMG scheme with RO' to optimize the route at the CN's network. In addition, it can be seen that, as the mobile network moves far away from the traffic anchoring network in N-DMM, the BuC increases as well.

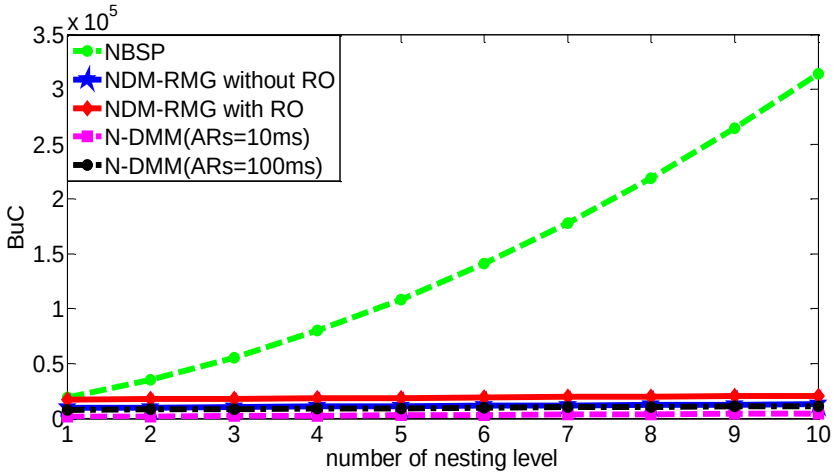


Figure 4-9 Binding update cost with different levels of nesting

4.6 Simulation Evaluation in ns-2

4.6.1 Simulation Scenario in ns-2

This section discusses the simulation performed using ns-2 [66] with the NIST mobility package for PMIPv6 [67]. The simulator has been used to model NDM-RMG and NEMO basic support protocol (NBSP). Figure 4-10 demonstrates the topology used in the simulation with two levels of nesting. The configuration parameters used in the simulation are shown in the figure. Routers (router1, router2 and router3) are fixed routers, which emulate the mobile routers. They emulate the nesting configuration.

In Figure 4-10, the correspondent node (CN) is a constant bit rate (CBR) UDP source. The CN is configured to send to MNN3 (a sink node) data with a packet size of 1000 bytes generated at packet intervals of 0.01sec.

The link between routers emulates a wireless link (IEEE 802.11b) with link delay and bandwidth shown in the figure.

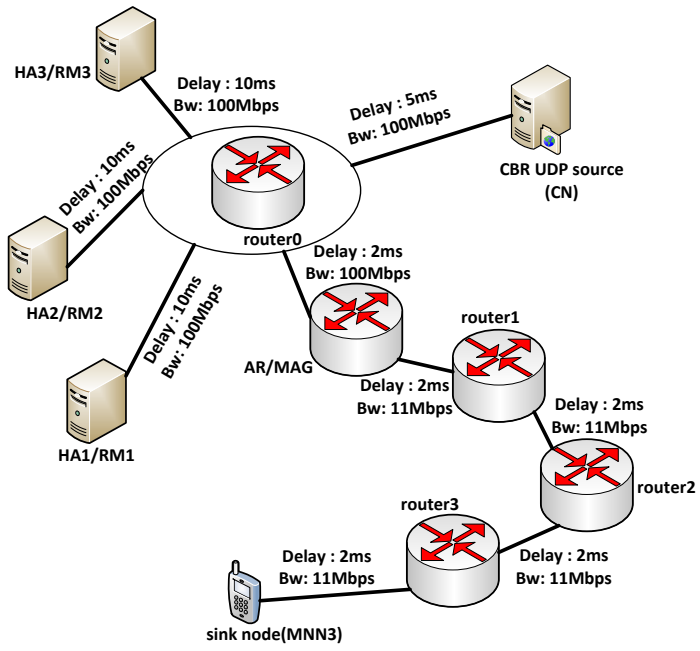


Figure 4-10 Simulated nested NEMO topology

Using ns-2 simulation, the performances of the NDM-RMG and NBSP schemes are

evaluated and compared in terms of packet delivery latency and packet overhead over the wireless link. The impacts of the number of levels of nesting and the distance between HAs/RMs on packet delivery latency and packet overhead are analysed.

4.6.2 Simulation Results and Analysis

Figure 4-11 shows a variation of the packet delivery delay for the packets sent to a node at the m^{th} level of nesting (sink node [MNN3]), according to the number of levels of nesting. It compares the NDM-RMG and the NBSP schemes. The number of levels of nesting is varied from 0 to 4 (0 implies that a mobile network has moved to a visited network, but no nesting is formed). The delays between HAs/RMs are fixed at 20ms. Whereas the packet delivery delay of the NBSP scheme is significantly affected by the level of nesting, it can be seen that the packet delivery delay for NDM-RMG scheme is only slightly affected by the level of nesting. Thus, the NDM-RMG scheme addresses the pinball routing problem because it uses distributed mobility management functions and a routing optimization mechanism.

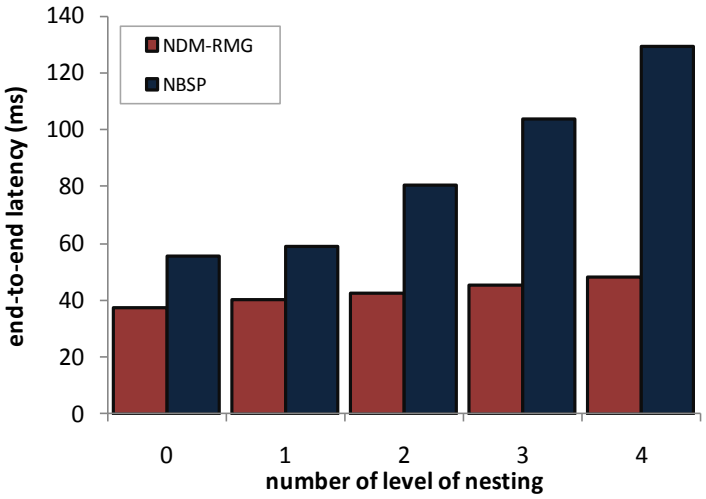


Figure 4-11 The impact of the number of levels nesting on Packet delivery delay

Figure 4-12 shows the impact of the distance between the HAs/RMs on end-to-end delay of the two schemes. During simulation, the number of levels of nesting has been set to 4; and the distance between the HAs/RMs has been varied between 10ms and 100ms. The packet delivery latency for the packets sent from the CN to the MNN3 has been measured. The results show that the packet delivery latency increases as the distance between HAs/RMs increases – for both

schemes. However, the NDM-RMG scheme experiences only a slight increase in terms of packet delivery latency, as the distance between the HAs/RMs increases; whereas the NBSP scheme experiences a significant increase in the packet delivery latency. This is because the NDM-RMG scheme optimizes the data route; and it is therefore, less affected by the pinball routing problem.

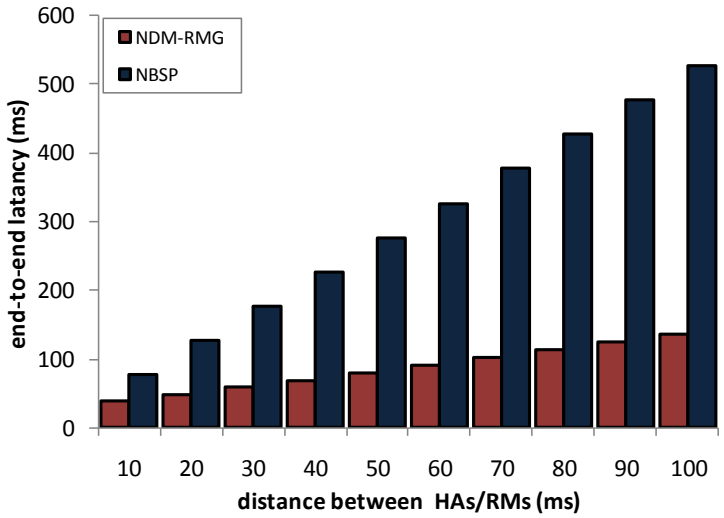


Figure 4-12 End-to-end delay, according to the distance between HAs/RMs

Figure 4-13 shows the data packet size transmitted from the CN and the measured packet size at the top-level mobile router (a size of the packet received at router 1), as the number of levels of nesting is varied during the simulation. The number of levels of nesting has been varied from 0 to 4. It can be seen from the figure that the size of the packet for NBSP increases as the level of nesting increases. This is because of the encapsulation due to pinball routing, which adds an extra header to the packets. As a result, the extra header of the packet is pushed to wireless link, which consumes the wireless link resources unnecessarily. In contrast, the NDM-RMG scheme has no packet overhead over the wireless link due to network-based feature used in the design of the scheme.

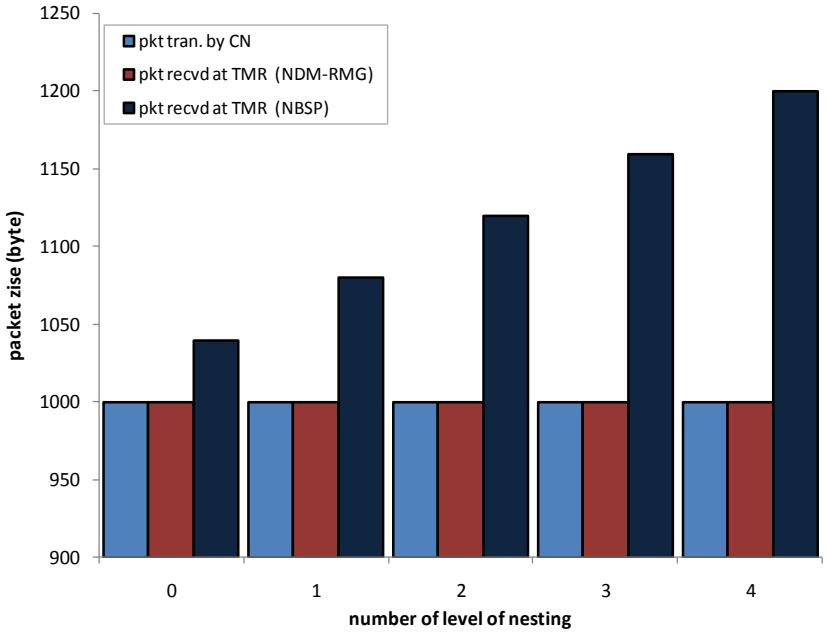


Figure 4-13 Measured packet size in the wireless link

4.7 Summary

The chapter has presented novel network-based distributed schemes for both non-nested and nested NEMO scenarios, which are named NDM-RMGs. The schemes decompose the LMA entity in PMIPv6, and distribute the mobility routing function to the gateways of different networks. The schemes combine this distribution with a delegating router function co-located with the location manager and the mobile routers, which helps to solve the pinball routing problem in the standard NEMO Basic Support protocol.

The detailed functioning of the scheme has been discussed, with signalling diagrams. The performance analysis shows that NDM-RMG mitigates the packet header overhead. The results also show that NDM-RMG scheme reduces packet delivery latency, PDC, and BuC – when compared with the NBSP scheme. Although, N-DMM has a slightly smaller packet delivery delay, PDC and BuC in cases where the mobile nodes are close to traffic anchoring networks, NDM-RMG outperforms N-DMM, when the mobile network moves far away from the anchoring network. The NDM-RMG is network-based DMM schemes; whereas N-DMM is a host-based DMM scheme. Thus, the N-DMM scheme inherits most of the shortfalls of host-based mobility support.

Chapter 5 Network-based DMM with Distributed Routing Management at Access Routers: DM-RMA

5.1 Introduction and Motivation

Chapter 3 and Chapter 4 discussed the proposed network-based distributed mobility management schemes that address the limitations of the current host and network IP mobility management protocols, respectively. The schemes decompose the logical functions of PMIPv6 and co-located the RM function at each of the gateway routers of the different sub-networks. It is, however, necessary to tunnel traffic between the RM and the access router (AR) in these schemes.

This chapter presents a new scheme that follows a similar concept of decomposing the logical functions of PMIPv6 to LM, RM and HNP allocation. However, the scheme proposed in this chapter co-locates the RM and HNP allocation functions at the access routers. This design brings the mobility logical functions (RM and HNP allocation) closer to the MN; and it prevents the need for tunnelling between RM and AR. Importantly, the mobility management remains in the network. Furthermore, the scheme addresses the limitations of centralized IP mobility management and avoids the static traffic anchoring in currently proposed DMM schemes.

This new scheme is called network-based distributed mobility management, with RM and HNP allocation functions distributed to the access routers (DM-RMA). DM-RMA can provide mobility support for a mobile node, as well as a mobile network. However, the thesis only discusses mobility support for the mobile node. The early preliminary version of DM-RMA scheme has already been presented in [51].

DM-RMA dynamically anchors the data traffic at the respective anchoring access routers of the MN during mobility events. This avoids the traffic bottlenecks and a single point of failure, as well as releases the load burden from the centralised mobility anchor, such as LMA. Moreover, it optimizes the route for ongoing MN communication, such that both handover and newly established communications are optimally routed. Additional mobility functions, such as MN tracking, updating and node information query are performed at the access routers. These

functions allow the access routers to track the MN's movement and to learn the MN's preconfigured mobility information. Furthermore, the design of DM-RMA allows the MN to configure different IP addresses. The MN uses the new address to establish new communication(s), and the old address(es) to maintain active communication(s).

The chapter discusses the design and functional operation of DM-RMA. An analytical model is developed for the performance evaluation of the DM-RMA scheme in comparison with other related distributed mobility schemes in the literature. The analytical model is used to evaluate the impact of network topology, session length, session-to-mobility ratio, and the probability of hand off traffic on packet delivery, tunnelling, signalling, and total costs.

Furthermore, the performance of the DM-RMA scheme is evaluated through discrete event simulation conducted in ns-2. In the ns-2 simulation, the impacts of DM-RMA on packet end-to-end delay, location update delay, hand off delay, and packet loss, are investigated.

5.2 Related Work

Network-based DMM inherits network-based mobility management features that many of the network operators consider for protocol deployment [13][16]. According to DMM requirements presented in [14], PMIPv6 can be extended to work in a distributed manner; hence it can thereby achieve network-based DMM.

It is possible to decompose and distribute the logical mobility management functions of the centralized LMA in PMIPv6 to different networks – while bringing them to the edge of the networks – so that they are closer to the users. Then, network-based distributed mobility management can be achieved by distributing the RM function to different networks. In order for the packet to be intercepted by the closest RM, anycast can be used [53], or the RM can be located at a network element through which the packets to/from the MN must pass [50][23].

In [23], a network-based distributed scheme comprising a single large domain divided into sub-networks is presented. The scheme co-locates the RM at each gateway of the different sub-networks. Simulation results show an improvement in packet delivery latency under various network load conditions. However, it is necessary to tunnel traffic between the RM and the AR in this network-based scheme.

In [77] the distributed PMIP (D-PMIP) scheme has been proposed. The scheme integrates

different PMIPv6 domains in a fully distributed (DF-PMIP), or a partially distributed (DP-PMIP) scheme. DP-PMIP scheme adds a common Inter-domain Central Mobility Database (ICMD) to which the LMA in each PMIPv6 domain performs another level of proxy binding update when the LMA receives such an update as a MN performs handover to a different PMIP domain, thereby receiving a new IP address in the new domain for new application sessions. The ongoing sessions from the previous domain utilize tunnels between the LMAs of the old and new domain.

To facilitate the setting up of the tunnel between the old and new LMAs, the ICMD may notify the new LMA about the old LMA in a relay (DP-PMIP-R) approach; or it may notify the old LMA about the new LMA in a locator (DP-PMIP-L) approach; or it may simultaneously inform both the new and old LMAs in a proxy (DP-PMIP-P) approach.

DF-PMIP also employs tunnels between the LMAs; but it lacks the ICMD. However, DP-PMIP outperforms DF-PMIP in terms of signalling cost, hand over latency and tunnel usage.

The intra-domain mobility of both D-PMIP schemes in each domain is centralized, as in PMIP, and therefore inherits centralized mobility management disadvantages. Hence, the tunnelling overhead between MAG and LMA remains. The inter-domain mobility also lacks route optimization, and is centralized at the original LMA. Furthermore, the DP-PMIP introduces a centralized control plane to manage the distributed PMIPv6 domains, which may scale with difficulty, as the number of domains and the number of MNs increase.

Partially distributed and fully distributed schemes for PMIPv6 are proposed by [57]. The partially distributed scheme distributes the data plane of the LMA to the mobility access gateways (named mobility anchor and access router – MAAR). The control plane is maintained in a centralized database, which works as the proxy for PMIPv6 signalling between the old and new MAARs. The fully distributed scheme distributes both control and data planes to each MAAR. In both schemes, when the MN moves from the old MAAR to the new MAAR with ongoing communication, a tunnel is built between them to route the ongoing communication. However, the schemes have not considered route optimization for ongoing communication. Consequently, when an MN with long-lasting traffic moves far away from the traffic anchoring MAAR, the resulting routing path becomes longer, and the end-to-end delay becomes larger.

5.3 DM-RMA Overview and Operation Mechanism

This section describes the DM-RMA scheme proposed in this chapter; and it explains its design approach, as well as the operation mechanism.

The DM-RMA scheme splits the logical functions of LMA in PMIPv6 into Location Management (LM), Routing Management (RM), and HNP allocation functions in a similar way, as presented in Chapter 3. Thereafter, it co-locates the RM and HNP allocation functions at the distributed access routers with a mobility client function, which is referred to as Mobile Access Gateway (MAG) in PMIPv6.

The DM-RMA constitutes a large domain, which is partitioned into distributed networks, as shown in Figure 5-1, with an example of three networks: network 1 (Net1), network 2 (Net2) and network 3 (Net3). Each network consists of: (i) An LM server; (ii) a gateway (GW), which is a normal IP router without mobility support functionality; and (iii) distributed MAGs, where RM is collocated.

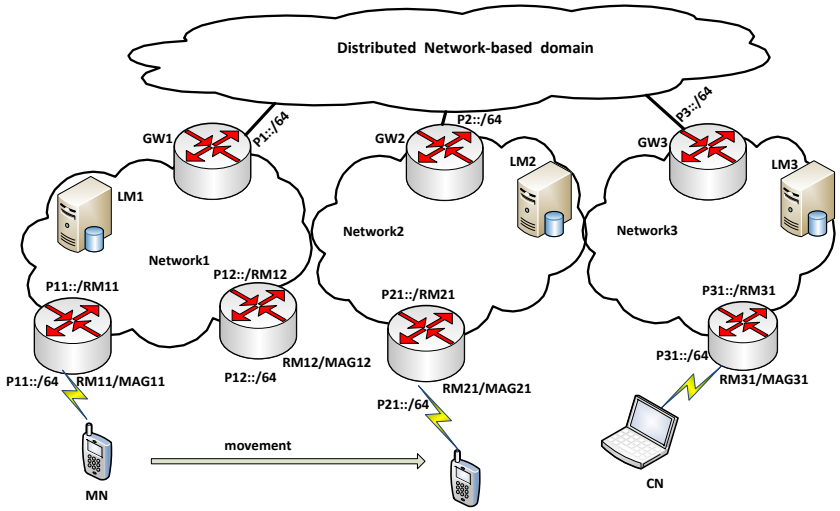


Figure 5-1 DM-RMA architecture

Each network owns a unique prefix block from a large domain. For example, Net1 owns P1::/64; Net2 owns P2::/64; and Net3 owns P3::/64. The prefix of each network is further sub-divided into unique subsets of prefixes; and each subset of prefixes is assigned to a particular MAG in the network, so as to allow for normal routing. For example, RM11/MAG11 in Net1 is assigned P11::/64, which is a subset of P1::/64; and RM21/MAG21 in Net2 is assigned P21::/64

from P2::/64. Each MAG allocates a unique prefix to an MN that attaches to its network from the assigned subset of prefixes. In order to allow MAGs to distinguish the intra-network handover from inter-network handover, each MAG in these distributed networks keeps a prefix table of other MAGs in the same network. This further improves the intra-network handover latency, and facilitates the discovery of the previous network(s).

Each server LM1, LM2, or LM3 in each network Net1, Net2, or Net3, respectively, keeps the mapping of MN's home prefix to the address of the RM, to which the MN is currently attached in that network. These servers form a distributed database of the overall LM functionality. The LMs can be virtually or physically deployed.

Each MAG has an HNP allocation function, and is responsible for registering the prefix it assigns to an MN to the LM server in its network. To allow for internetworking mobility routing, every RM interacts with the LM server located in its network, in order to find the MN location information whenever it realizes that the MN's preconfigured address(es) belong(s) to a different network(s), such as in an inter-network handover. In contrast, when the MN undergoes handover within the same network, the RM uses its prefix table to locate the previous RM(s) where the MN's address(es) were configured.

The data-plane routing is served by the local RM in the MAG, to which the MN is attached. Moreover, the scheme optimizes the route for ongoing session(s), whereby the RMs of the CNs are notified about the RM currently serving the MN. The traffic, therefore, will not need to route via the original RM anchoring the traffic, thus improving the traffic delivery latency.

Based on the DMM and IPv6 concepts, DM-RMA allows the MN to configure and use multiple IP addresses from the prefixes advertised by the different MAGs it visits. The principle of operation of the DM-RMA is detailed in the sub-sections below.

5.3.1 Initial MN Registration and Communication Establishment

Figure 5-2 shows the initial attachment and establishment of communication. When MAG11 detects the attachment of an MN to its network, it gets the MN identity (ID) using PMIPv6 procedures. It then uses HNP functionality to assign to the MN a unique prefix (i.e., P11::/64(mn)). Thereafter, MAG11 creates a binding entry for the MN, and sends the router

advertisement (RA) message to the MN including P11::/64(mn). Meanwhile, it registers the MN to LM1 (through the PBU message), and LM1 records the association of mn-id, P11::/64(mn) and P11::/MR11.

Upon receiving the RA, the MN configures from P11::/64(mn) an IP address P11::/mn, which is used as the session identifier when the MN communicates with a CN in Net3, for example. As a packet from the CN destined to P11::/mn reaches RM31 (i.e., RM31/MAG31 in Figure 5-1), there is no binding information for P11::/mn from its binding cache or prefix table, because P11::/mn belongs to a different network. The packet is, therefore, routed to GW1 in Net1, then to RM11, and finally to MN – using the usual routing table without tunnelling.

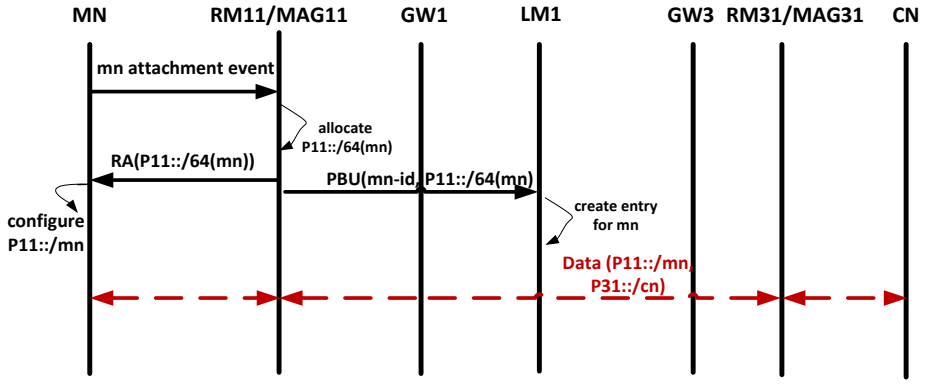


Figure 5-2 Initial attachment and session establishment procedures

5.3.2 Handover to another Network

Figure 5-3 shows the signalling call flow for the MN that was initially attached to RM11 in Net1, as it moves and attaches to RM21 in Net2, with reference to Figure 5-1. As the MN enters Net2, MAG21 detects the attachment, and it acquires the MN-ID and MN prefix P11::/mn. From the prefix table, MAG21 finds out that the MN is moving to Net2 from a different network. It allocates a prefix P21::/64(mn), and sends the RA message with this prefix to MN. Meanwhile, MAG21 sends to LM2 a modified PBU with a newly defined flag to register the MN and to request the IP address of the RM anchoring the MN's old prefix, P11::/64(mn). This is to allow forwarding of the MN's old traffic towards RM21.

With the LM distributed database, LM2 knows that the owner of MN's old prefix information is the LM1, to which it then forwards the request through a modified PBU message.

This includes the address P21::/ RM21 of RM21 and the old prefix, with a LM-flag, so as to resolve the RM address for this prefix.

Upon receiving the PBU message with the LM-flag, LM1 looks up the prefix to locate RM11; and it notifies RM11 about RM21 (P21::/ RM21) through a modified PBU. As LM1 replies to LM2 with a modified PBA with the RM11 address, LM2 also replies with this address to RM21. RM21 then communicates with RM11 to establish a bi-directional tunnel between them, in order to carry the ongoing traffic from/to MN when using this old prefix.

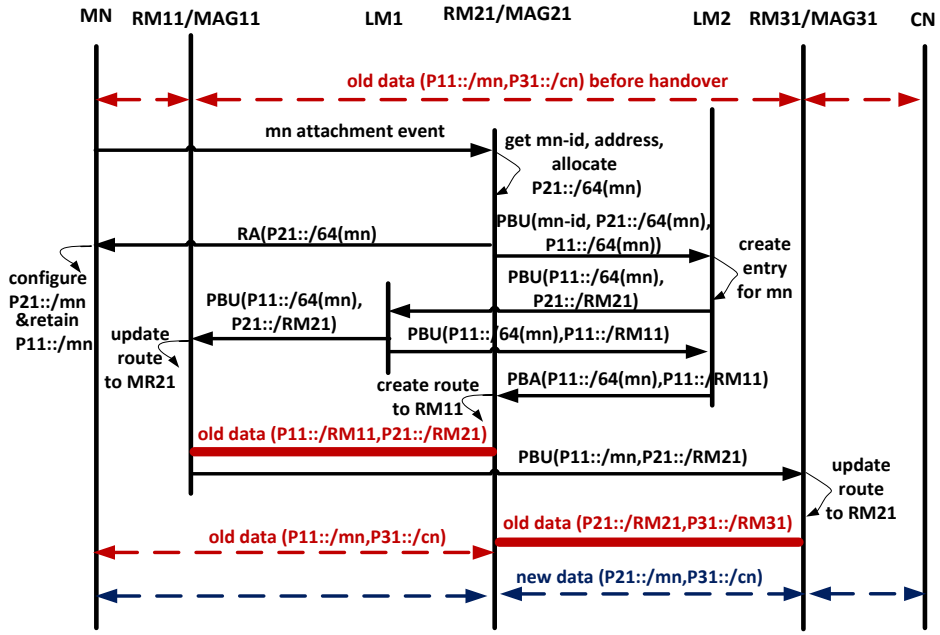


Figure 5-3 Handover signalling call flow and new session setup

While the P11::/mn packets are tunnelled between RM11 and RM21, RM11 notifies RM31 (where the CN is anchored) about the address of RM21 through PBU message. RM31 then caches the new location of the MN for the ongoing session. Subsequently, the packets are tunnelled directly between RM31 and RM21 – in order to avoid triangle routing, and to reduce the end-to-end latency.

The new communication established after the MN has moved to Net2 uses the new address P21::/mn as the session identifier; and therefore, its traffic is routed directly to/from RM21. So, the proposed scheme optimally routes both the ongoing and the newly established communications.

5.3.3 Handovers within the Same Network

Within the same network, e.g., Net1, the MN may change its point of attachment, e.g., moving from RM11 to RM12. Then MAG12 (where RM12 collocates) detects the MN attachment, acquires MN-ID and the configured IP address P11::/mn, and assigns to MN a new prefix, i.e. P12::/64(mn). However, MAG12 will use the prefix table of all the RMs in Net1 to look up RM11 (using the old prefix). It then runs RM12 to initiate the set up of the tunnel with RM11 to carry ongoing traffic of the old prefix. It also updates the MN location to LM1.

Upon receiving the RA message from MAG12, the MN also configures a new address, P12::/mn, which it uses for newly established sessions, in order to route directly without using the tunnel.

To achieve ongoing session continuity, the new RM – where the MN attaches – needs to discover the previous RM, to which the MN was previously attached. Various mechanisms are possible, such as layer 2 handover mechanisms, e.g., utilizing IEEE 802.21 and IPv6 neighbour discover mechanism. Another approach, that does not introduce extra signalling, gets the MN configuration information from the MN uplink traffic. However, the ongoing session may experience long interruptions, when the MN has no packets to send. DM-RMA employs a Node Information Query (NIQ) [78] mechanism. Upon detecting the MN attachment in its network, the MAG exchanges NI Query/Reply messages [64] to get the MN configured IP address, from which it extracts the prefix(es). It then uses either its prefix table, or the distributed LM servers to discover the IP address of the old RM, as described in the previous sub-sections.

5.4 Performance Evaluation

In this section, the performance of DM-RMA is evaluated and compared with a similar DMM scheme, named the D-PMIP (DP-PMIP and DF-PMIP) [77]. The D-PMIP scheme is a network-based scheme, which also employs partitioned networks (i.e., domains) similar to the DM-RMA scheme. D-PMIP performs better than centralized mobility management [77]. However, DM-RMA decomposes and distributes the PMIPv6 functionalities, whereas D-PMIP distributes the whole PMIPv6 domain to different networks.

The performance of the schemes is evaluated by using analytical modelling and discrete event simulation methods. In the analytical evaluation, the protocol costs of DM-RMA and D-

PMIP are analysed and compared in terms of total cost, signalling cost, packet delivery cost, and tunnelling cost. In ns-2 simulation, the schemes are evaluated, using the following metrics: packet end-to-end delay, location update delay, handover delay, and packet loss.

In the following sub-sections, the analytical network model and the mobility model used to evaluate these costs are described. The analytical results and their analysis are discussed in Section 5.5; while the simulation evaluation is presented in Section 5.6.

5.4.1 Network Model

Figure 5-4 shows the network model used for cost modelling and performance analysis.

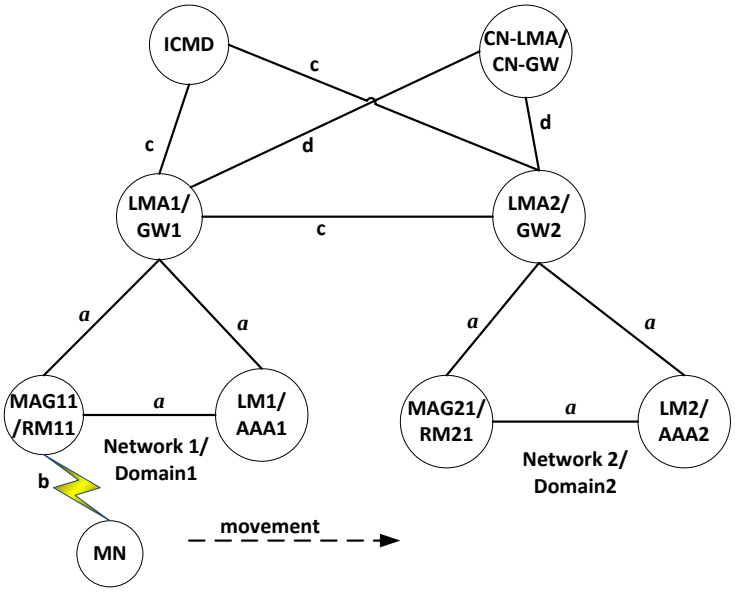


Figure 5-4 A network model used for cost modelling

The model is composed of four networks/domains, each made of either an LMA/GW or an ICMD. It is assumed that the gateway (GW) in different sub-networks of DM-RMA is positioned in the same location as LMA in D-PMIP domains. For a fair analysis and comparison, the network under GW defined in DM-RMA, is assumed to be identical to the network under LMA, as described by D-PMIP. The LM servers in DM-RMA and the Authentication, Authorization and Accounting (AAA) servers used in DF-PMIP are assumed to be co-located.

The CN and MN are assumed to be located in different LMA/GW domains. Furthermore, each domain is assumed to be circular in shape, and consists of N identical circular cells, each

with a MAG integrated with access point function. The area of each cell is $A = \pi R^2$, where R is the radius of the cell.

The nodes are connected through a number of hops. The average distance of connections (hop counts) between different nodes, shown in letters in Figure 5-4, is considered to be the same between different domains, i.e., c= d, and that between adjacent MAGs in the same domain is considered to be identical, and is defined as $g = \sqrt{N}$ [79] , whereas a is the hop count between MAGs and LMA/LM in a given domain.

5.4.2 Mobility and Traffic Models

Based on the principle of operation of DM-RMA and D-PMIP, two kinds of handover are considered: intra-domain handover, when the MN moves to another MAG within the same domain; and inter-domain handover, when the MN moves to another MAG in a different domain. In the rest of this chapter, the term domain also refers to the network under the GW, as defined in DM-RMA.

To derive the mobility model, the thesis assumes the following: (i) The residence time of an MN staying in a particular MAG or in a domain is an exponentially distributed random variable; (ii) the session arrival process to an MN follows a Poisson distribution, i.e., the inter-arrival time is exponentially distributed, with a rate λ_s; (iii) the MN movement follows a Fluid Flow model [80][81], in which the MN moves at an average speed ν and in a direction uniformly distributed over the range [0,2π].

The MAG crossing rate μ_c and the domain crossing rate μ_d are, respectively, expressed as follows [82]:

$$\mu_c = 2\nu/\sqrt{\pi A} \qquad \text{and} \qquad \mu_d = \mu_c/\sqrt{N} \qquad \textbf{(5.1)}$$

Where ν is the average speed (m/s) of the MN.

The rate μ_l of intra-domain handover, where there is an MAG crossing, but no domain crossing, is computed by subtracting μ_d from μ_c as given in (5.2).

$$\mu_l = \mu_c - \mu_d = \mu_c(1 - 1/\sqrt{N}) \quad \textbf{(5.2)}$$

From (5.1) and (5.2), the average number of movements (i.e., location updates) during an inter-session time interval can be evaluated. Then, the average numbers E(N_c) and E(N_d) of movement crossing MAG and crossing domain, respectively, can be expressed as follows [82].

$$E(N_c) = \mu_c/\lambda_s \qquad \text{and} \qquad E(N_d) = \mu_d/\lambda_s \quad \textbf{(5.3)}$$

The average number of movements, for which the MN still remains in the same domain E(N_l) is given as follows:

$$E(N_l) = (\mu_c/\lambda_s)(1 - 1/\sqrt{N}) \quad \textbf{(5.4)}$$

5.4.3 Total Cost

To evaluate the performance of DM-RMA, and to efficiently compare it with D-PMIP, mathematical expressions are formulated for signalling cost, packet delivery cost, total cost, and tunnelling cost. The total cost for a mobility protocol, C_T, is the sum of the signalling cost, C_S, which is due to the signalling used to provide mobility support and the packet delivery cost, C_{PD} , which is due to the data packet transmission to an MN.

$$C_T = C_S + C_{PD} \quad \textbf{(5.5)}$$

5.4.4 Signalling Cost

The location update comprises intra-domain signalling due to intra-domain movement with signalling cost C_S^{intra}, and inter-domain signalling due to inter-domain movement with signalling cost C_S^{inter}. The signalling cost C_S is then expressed as follows:

$$C_S = C_S^{intra} + C_S^{inter} \quad \textbf{(5.6)}$$

The location update incurs packet transmission cost and processing cost on mobility management related entities, i.e., LM and MR. The packet transmission cost is proportional to the distance (in hops or time) between source and destination nodes [82][83]. Assume the

average distance (hops) between network nodes X and Y is $d_{X,Y}$. Then, the transmission cost of a control packet between the nodes X and Y belonging to the wired network part, is $C_{X,Y} = \alpha d_{X,Y}$ – while $C_{MN,MAG} = \beta d_{MN,MAG}$ belongs to the wireless part, where α and β are the unit transmission costs in a wired and a wireless link, respectively.

The location update signalling cost is given by the product of the size of the mobility signalling message and the weighted distance (hops) [79][84]. In addition, the location update signalling cost includes the processing cost of the signalling message in mobility agents (i.e., binding lookup). Following the operation procedures of the schemes, the network model given in Figure 5-4, and considering the average number of movements of the MN (5.3 and 5.4), the intra- and inter-domain signalling costs during the inter-session time interval for each scheme are derived as follows.

(a) Intra-domain signalling cost for DM-RMA

The cost includes cost due to signalling message exchange needed to setup a tunnel between old and new RMs, to advertise the prefix(es), to acquire MN's previous information, and to register the MN to the LM server.

It is assumed that the RA message size, L_{RA}, is equal to the sizes of the router solicitation and NI (node information) request/reply messages. Also, it may be assumed that the PBU message size, L_{PBU}, is equal to the PBA message size. Suppose PC_Z defines the processing cost on the mobility related entity z. Then the signalling cost is computed as follows:

$$C^{\text{int } ra}_{S(DM-RMA)} = E(N_l)(4\beta \cdot b \cdot L_{RA} + 2\alpha \cdot g \cdot L_{PBU} + 2\alpha \cdot a \cdot L_{PBU} + PC_{LM} + PC_{MAG}) \quad \textbf{(5.7)}$$

(b) Inter-domain signalling cost for DM-RMA

Here, the cost includes the cost due to the control message exchange, in order to obtain MN previous information, to advertise prefix(es) to MN, to setup the tunnel between new and old MRs in different domains, and to optimize the route for ongoing session. Accordingly, the inter-domain signalling cost is given as follows:

$$C^{\text{int } er}_{S(DM-RMA)} = E(N_d)(4\beta \cdot b \cdot L_{RA} + 3\alpha \cdot a \cdot L_{PBU} + 2\alpha \cdot c \cdot L_{PBU} + \alpha \cdot d \cdot L_{PBU} + 3PC_{LM} + 3PC_{MAG}) \quad \textbf{(5.8)}$$

The signalling cost for DM-RMA is given by the sum of (5.7) and (5.8).

(c) Intra-domain signalling cost for DP-PMIP

The three approaches, DP-PMIP-R, DP-PMIP-L, and DP-PMIP-P, use the PMIPv6 procedure to manage the intra-domain handover. Therefore, the intra-domain signalling cost can be expressed as follows:

$$C_{S(DP-PMIP)}^{intra} = E(N_I)(2\beta \cdot b \cdot L_{RA} + 2\alpha \cdot a \cdot L_{PBU} + PC_{LMA} + PC_{MAG}) \quad \textbf{(5.9)}$$

(d) Inter-domain signalling cost for DP-PMIP

Also, it is noticed that all three schemes have the same signalling cost for inter-domain handover, which includes intra-domain signalling costs, registration costs to ICMD, and tunnel establishment cost between old LMA and new LMA.

$$C_{S(DP-PMIP)}^{inter} = E(N_d)(2\beta \cdot b \cdot L_{RA} + 2\alpha \cdot a \cdot L_{PBU} + 4\alpha \cdot c \cdot L_{PBU} + 3PC_{LMA} + PC_{MAG} + 2PC_{ICMD}) \quad \textbf{(5.10)}$$

The signalling cost for DP-PMIP is the sum of (5.9) and (5.10).

(e) Intra-domain Signalling cost for DF-PMIP

Here, the cost includes the regular PMIPv6 signalling cost, and the cost to acquire the MN configured information. So, the intra-domain signalling cost is expressed as follows:

$$C_{S(DF-PMIP)}^{intra} = E(N_I)(4\beta \cdot b \cdot L_{RA} + 2\alpha \cdot a \cdot L_{PBU} + PC_{LMA} + PC_{MAG}) \quad \textbf{(5.11)}$$

(f) Inter-domain signalling cost for DF-PMIP

The cost includes the cost due to signalling message exchanges, so as to get MN configuration information, to resolve the MN previous domain using the AAA infrastructure, and to register the MN to LMA in a new domain. Then the intra-domain signalling cost is expressed as follows:

$$C_{S(DF-PMIP)}^{inter} = E(N_d)(4\beta \cdot b \cdot L_{RA} + 4\alpha \cdot a \cdot L_{RADIUS} + 2\alpha \cdot a \cdot L_{PBU} + 2\alpha \cdot c \cdot L_{RADIUS} + 2\alpha \cdot c \cdot L_{PBU} + 4PC_{AAA} + 2PC_{MAG} + 3PC_{LMA}) \quad \textbf{(5.12)}$$

where L_{RADIUS} defines the message size of the RADIUS messages exchanged with AAA.

The signalling cost for DF-PMIP is the sum of (5.11) and (5.12).

5.4.5 Packet Delivery Cost

Packet delivery cost, C_{PD}, defines the cost which a packet incurs during ongoing communication session between CN and the MN. The cost includes the data packets' transmission cost, and the cost to process packets at mobility entities. Accordingly, the packet delivery cost can be computed as the sum of the packet transmission cost and the processing cost. The processing cost includes costs on LMA and MAG (i.e., look-up of mapping table, encapsulation/de-capsulation).

All the schemes considered allow an MN to anchor traffic on different anchors. That is, the old traffic (traffic handed over) and new traffic can be routed differently. For example, when the MN performs a handover to a new domain, the ongoing communication may remain anchored to the anchor in the old domain; while the newly established traffic is anchored to the anchor in the new domain. Therefore, both types of traffic may experience different costs. Hence, the packet delivery cost is composed of costs for old traffic and new traffic.

Let ω and τ define the probability of the old traffic (i.e. a probability that the traffic is hand off traffic) and tunnel overhead size, respectively. So, 1 – ω presents the probability that the traffic is the newly established traffic after the MN has performed its handover to a new domain. Suppose E(S) and L_d are average session lengths measured in packets (i.e., average number of packets during a session) and average data packet size, respectively. The packet delivery cost for the schemes is computed as follows.

(a) Packet delivery cost for DM-RMA

When the MN performs a handover to a new domain with ongoing session, there are the first few packets that traverse the RM in the old domain, which are then tunnelled to the RM in a new domain, while the route is being optimized. After the route is optimized, all the packets are tunnelled directly from the CN's network to the network to which the MN is currently attached. It is assumed that 1/E(S) [79] is a ratio of the packets of the MN's old session that traversed the old RM in the old domain before the route was optimized. That is, 1 – 1/E(S) represents the ratio of the packets of the old session that go through an optimized path.

On the other hand, the path for the new session is optimized; and the packets arrive to RM in the new domain without tunnelling. Hence, the packet delivery cost is expressed as

follows:

$$C_{PC(DM-RMA)} = \omega \cdot (\frac{1}{E(S)} \cdot (\alpha \cdot E(S) \cdot L_d \cdot (d+a) + \alpha \cdot E(S) \cdot (L_d + \tau) \cdot (2a+c) + \beta \cdot E(S) \cdot b \cdot L_d + 2PC_{MAG}) + (1 - \frac{1}{E(S)}) \cdot (\alpha \cdot E(S) \cdot (L_d + \tau) \cdot (d+a) + \beta \cdot E(S) \cdot b \cdot L_d + 2PC_{MAG})) + (1-\omega) \cdot (\alpha \cdot E(S) \cdot L_d \cdot (d+a) + \beta \cdot b \cdot E(S) \cdot L_d) \quad \textbf{(5.13)}$$

(b) Packet delivery cost for DP-PMIP and DF-PMIP

The schemes use the same packet routing path for the old session. Similarly, the packets for the new session established after handover follow a similar routing path in both schemes. Moreover, the schemes do not implement RO (route optimization) for the ongoing session. Hence, both schemes have the same packet delivery cost, which is derived as follows:

$$C_{PC(DP-PMIP)} = \omega \cdot (\alpha \cdot E(S) \cdot d \cdot L_d + \alpha \cdot E(S) \cdot c \cdot (L_d + \tau) + \alpha \cdot E(S) \cdot a \cdot (L_d + \tau) + \beta \cdot E(S) \cdot b \cdot L_d + 2PC_{LMA} + PC_{MAG}) + (1-\omega) \cdot (\alpha \cdot E(S) \cdot d \cdot L_d + \alpha \cdot E(S) \cdot a \cdot (L_d + \tau) + \beta \cdot E(S) \cdot b \cdot L_d + PC_{LMA} + PC_{MAG}) \quad \textbf{(5.14)}$$

5.4.6 Packet Tunnelling Cost (C_{TC})

Data packets are encapsulated and de-capsulated, while they travel through the tunnels between RMs, MAG and LMA, and LMAs. The tunnelling costs indicate the cost due to tunnel overhead; hence, using (5.13) and (5.14), C_{TC} can be expressed as follows:

$$C_{TC(DM-RMA)} = \omega \cdot (\frac{1}{E(S)} \cdot (\alpha \cdot E(S) \cdot \tau \cdot (2a+c) + 2PC_{MAG}) + (1 - \frac{1}{E(S)}) \cdot (\alpha \cdot E(S) \cdot \tau \cdot (d+a) + 2PC_{MAG})) \quad \textbf{(5.15)}$$

$$C_{TC(DP-PMIP)} = \omega \cdot (\alpha \cdot E(S) \cdot c \cdot \tau + \alpha \cdot E(S) \cdot a \cdot \tau + 2PC_{LMA} + PC_{MAG}) + (1-\omega) \cdot (\alpha \cdot E(S) \cdot a \cdot \tau + PC_{LMA} + PC_{MAG}) \quad \textbf{(5.16)}$$

5.5 Numerical Results and Discussion

This section discusses the performance evaluation of DM-RMA. It quantitatively compares DM-RMA with D-PMIP [77], based on the analytical cost functions developed in Section 5.4. The default parameter values used for the analysis are according to [79][84], and are given in Table 5-1. For simplicity, all control messages are assumed to have equal message sizes,

L_s. The numerical results were obtained through implementing the developed analytical cost functions in MATLAB (R2009b) by using the default parameter values given in Table 5-1.

Table 5-1 Parameters used for numerical results

Notation	Meaning	Value
a	Distance between MAGs and LMA/GW (hops)	5
b	Distance between MAG and MN (hops)	1
c	Distance between LMAs/GWs (hops)	15
d	Distance between CN-LMA/GW and LMAs/GWs (hops)	15
N	Number of MAGs in a given domain	$1-25$
$E(S)$	Average session length in packets	30
R	Radius area that MAG cover (m)	100
L_s	Size of the control signaling message (Bytes)	76
L_d	Size of the data packet (Bytes)	10
PC_{LMA}	Processing cost on LMA, LM, ICMD,AAA	24
PC_{MAG}	Procesing cost on MAG/RM	12
α, β	Unit transmission cost wired/wireless	1,1.5
τ	IPv6 tunnel header (Bytes)	40

The impacts of the following metrics are investigated: MN speed, domain size, average number of packets per session, hand off traffic probability, and the session-to-mobility ratio (SMR), on the above mentioned costs. The details of the evaluation are given below.

(i) Impact of MN speed on signalling cost

Figure 5-5 shows the impact of the MN speed on the signalling cost for the DM-RMA, DP-PMIP, and DF-PMIP schemes. The velocity of the MN is varied from 5m/s to 100m/s; while the domain size, N, was set as 15.

The results show that the signalling cost of the three schemes increase linearly with the MN speed. This is because, when the MN moves faster, the handover rate increases, for both intra- and inter-domain handover. Consequently, the frequency of the location updates increases, which subsequently increases the signalling cost. However, the DP-PMP scheme has an average of 20% lower cost than the DM-RMA scheme, and 28% lower cost than the DF-PMIP scheme. DP-PMIP incurs lower cost because of its common centralized control plane; whereas the other schemes employ a fully distributed mobility management approach, and incur extra signalling cost. The extra signalling cost is incurred when locating the MN's old network when inter-

domain handover occurs, which increases as the handover rate increases.

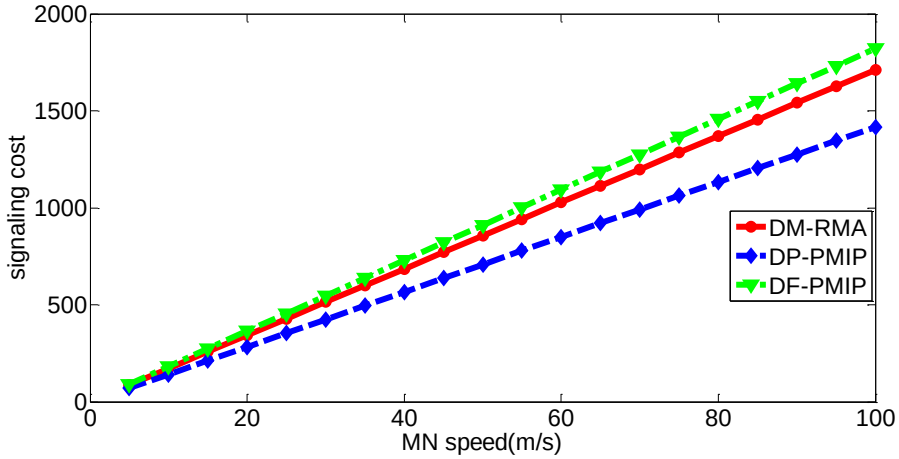


Figure 5-5 The variation of signalling cost with MN speed

(ii) The influence of domain size on signalling cost

Figure 5-6 shows the impact of the domain size on signalling cost. The size of the domain, N, is varied by increasing the number of MAGs from 1 to 25; while the MN moves at a constant speed of 30m/s. Then, both intra- and inter-domains signalling are observed: the combined results of the signalling cost for each scheme is presented in Figure 5-6. From the results, DM-RMA demonstrates a better signalling cost at small domain sizes, when compared with other schemes. However, as the domain size increases, DP-PMIP has lower signalling cost. This is because when the domain size is large, the MN stays longer within a particular domain; and as a result, the intra-domain handover rate increases. Given that DM-RMA distributes the mobility routing function at the access routers, it incurs extra signalling between MAGs during the intra-domain handover. On the other hand, when the domain is small, the inter-domain handover rate increases. Consequently, DP-PMIP incurs extra signalling due to location updates at the common database as well as tunnel setup. Similarly, the DF-PMIP encounters extra signalling to locate the MNs previous network through the AAA infrastructure – in addition to the tunnel setup cost.

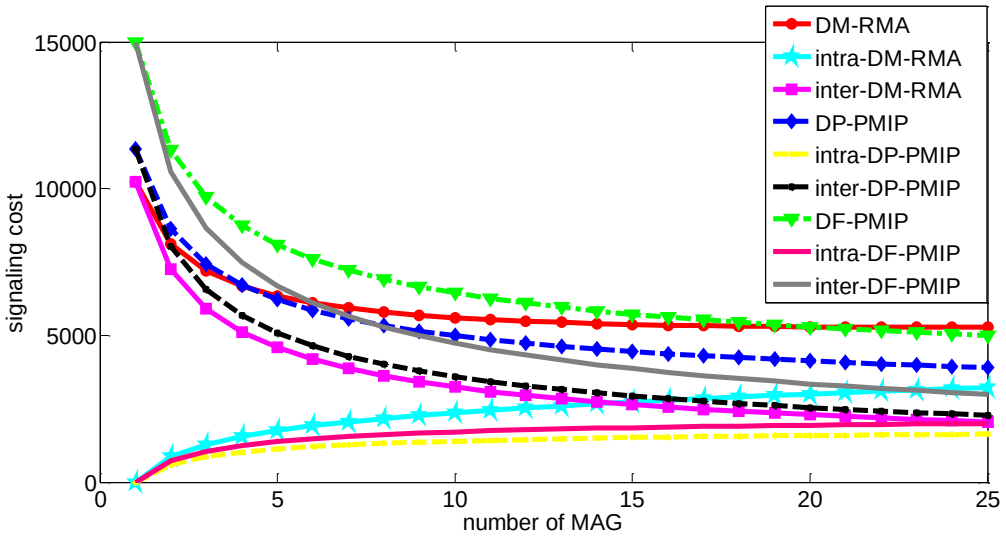


Figure 5-6 The impact of domain size on signalling cost

(iii) *The impact of session length, $E(S)$, on packet delivery and tunnelling costs*

Figure 5-7 and Figure 5-8 show the variation of the packet delivery cost and the tunnelling cost, as the average number of packets during a session, $E(S)$, is varied. The $E(S)$is varied from 5 to 30; while the probability of the hand off traffic, ω, during a session was set as 0.5.

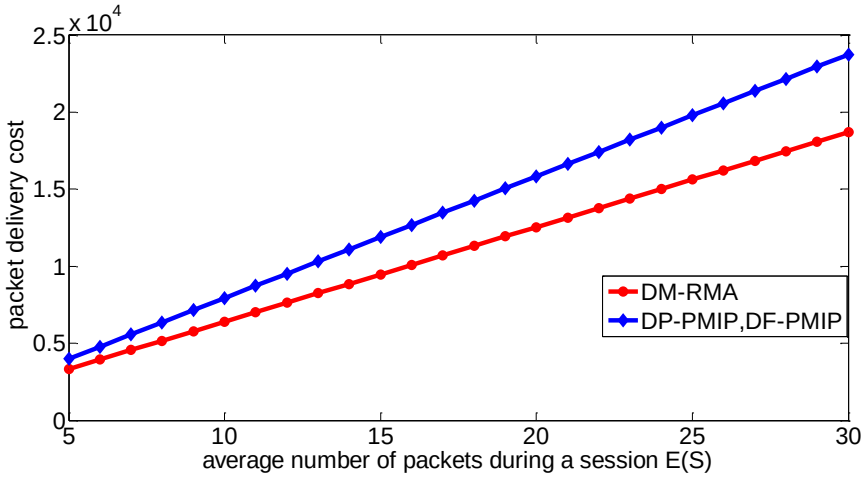


Figure 5-7 The effect of average session length on packet delivery cost

From Figure 5-7, the packet delivery cost of the schemes increases as $E(S)$ increases; however, DM-RMA outperforms both DP-PMIP and DF-PMIP. The reason is that DM-RMA optimizes the route for both hand off and new traffic; and hence, it reduces the packet delivery

cost. Moreover, DM-RMA has lower tunnelling cost, as shown in Figure 5-8, because it alleviates tunnelling for the new traffic, and optimizes the route for the old traffic. Although DP-PMIP and DF-PMIP optimize the route for the new traffic, they still apply PMIPv6 tunnelling to the new traffic within a domain.

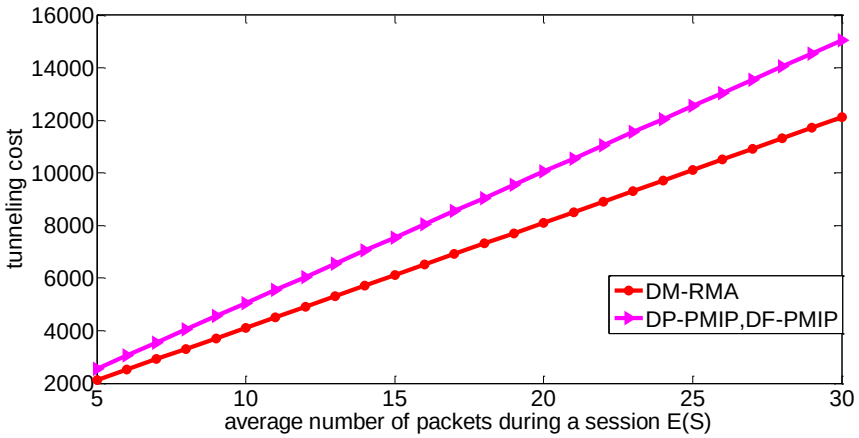


Figure 5-8 The impact of average session length on packet tunnelling cost

(iv) *The impact of hand off probability on packet delivery and tunnelling costs*

Figure 5-9 and Figure 5-10 show the impact of hand off traffic and the new traffic probabilities on the packet delivery cost and the tunnelling cost. The probability of hand off traffic, ω, in a session was varied from 0 to 1; and $E(S)$ was set as 30.

It is observed that the DM-RMA has better packet delivery cost and tunnelling cost compared with DP-PMIP and DF-PMIP, regardless of the percentage change of the hand off traffic, ω. This is because DM-RMA optimizes the path for both old and newly established traffic. However, the performance gap of the packet delivery cost and the tunnel cost decreases as the ω increases, as shown in Figure 5-9 and Figure 5-10, respectively. The reason is that when $\omega = 0$, the ongoing session has only the new established traffic; while when $\omega = 1$, the session consists of only handover traffic. Given that DM-RMA does not use a tunnel for the new traffic, it has better packet delivery cost and tunnelling cost for the new traffic compared with DP-PMIP and DF-PMIP, which both tunnel the new traffic between LMA and MAG in a given domain.

However, in the case of hand off traffic, there is a cost in DM-RMA, due to the first few packets of the hand off traffic that traverse the non-optimal route, while the route is being optimized (in addition to the cost for the packets that are tunnelled directly from CN's network to

current MN's network after the RO). On the other hand, both DP-PMIP and DF-PMIP tunnel the hand off traffic via a non-optimal path between the old anchoring domain and the new anchoring domain to which the MN is currently attached.

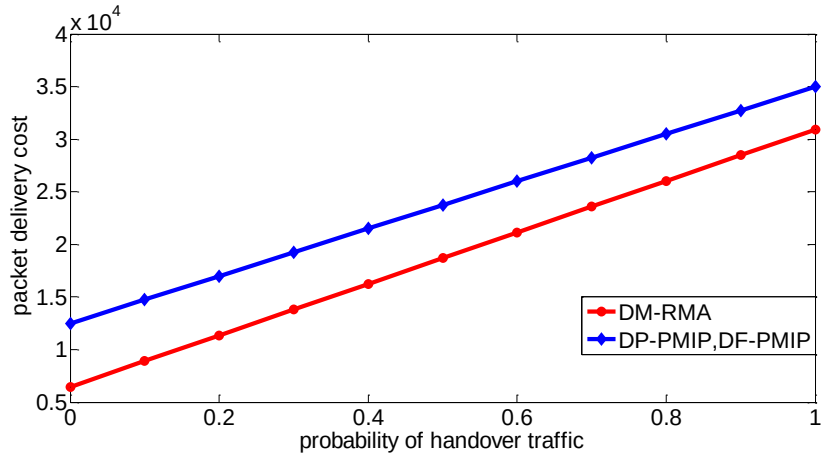


Figure 5-9 The impact of the probability that the traffic is hand off traffic on packet delivery cost

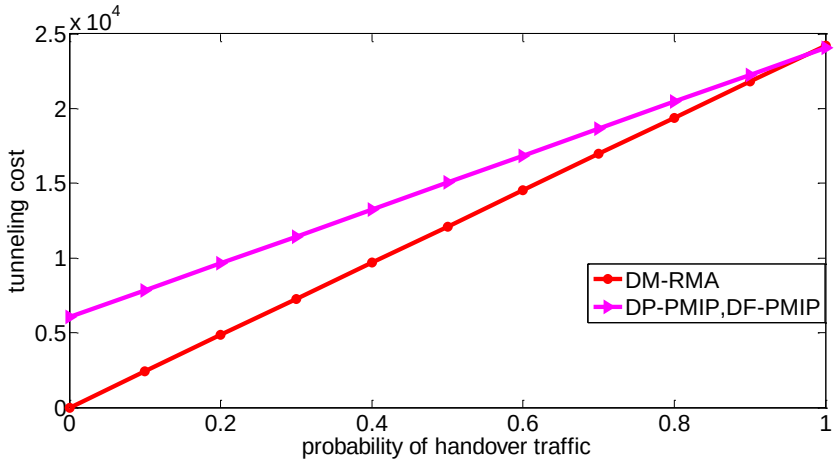


Figure 5-10 The impact of the probability that the traffic is hand off traffic on packet tunnelling cost

(v) *The impact of session-to-mobility ratio on total cost*

Figure 5-11 shows the performance of the total cost (equation 5.5) as a function of an important performance factor called session-to-mobility ratio (SMR). The SMR provides the relative ratio of session arrival rate to user mobility rate (i.e., MAG crossing rate), λ_s/μ_c [82][84].

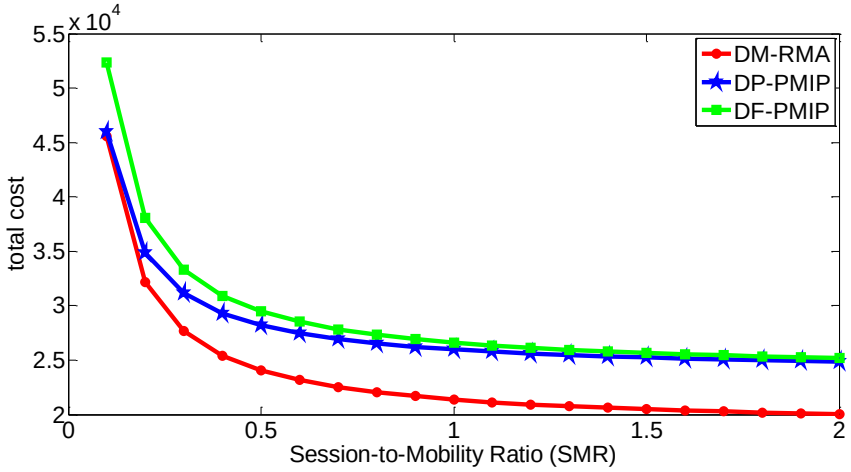


Figure 5-11 The impact of SMR on total cost

From the result, as the SMR increases, the total cost for all the schemes decreases. However, DM-RMA has a better total cost than other schemes. This can be explained by the cost benefit gains in packet delivery costs compared to the other schemes. DP-PMIP outperforms DF-PMIP, when the SMR is small; however, at large SMR, both DP-PMIP and DF-PMIP demonstrate a similar performance. The reason is that at small values of SMR, the mobility is high; hence the signalling cost dominates in the total cost, as shown in Figure 5-5. At large values of SMR, the mobility is low, meaning that the packet delivery cost becomes the major contributor to the total cost over the signalling cost.

5.6 Simulation Evaluation in ns-2

This section describes the simulation environment and the modelling of DM-RMA and DP-PMIP in ns-2 [66] (briefly discussed in Section 3.5.1) with NIST mobility package for PMIPv6 [67]. The performance of the DM-RMA scheme is compared with that of DP-PMIP scheme, which has a better performance in comparison to DF-PMIP and centralised IP mobility approach, such as MIPv6 [77].

The impact of decomposing and distributing the mobility management functions, and the introduction of the route optimization (RO) mechanism as proposed in DM-RMA scheme, are analysed. Two different scenarios are considered. In the first scenario, the impact on handover delay and packet loss for intra-domain handover is evaluated. The second scenario analyses the impact on location binding update latency, packet delivery latency, handover delay, and packet loss for inter-domain handover scenario.

5.6.1 Simulation Scenarios

The simulated network topology is shown in Figure 5-12. The topology is considered to constitute four domains/networks (made of either LMA/GW or ICMD) connected through Router0. Router0 and Router1 represent normal routers without mobility management functions; and they are used to connect different mobility management entities. For reasonable comparison of the schemes, the GW in DM-RMA and the LMA in DP-PMIP are implemented in the gateway of each domain/network. The distances from MAGs to both LMA and LM in a given domain are configured as being the same, for fair comparison.

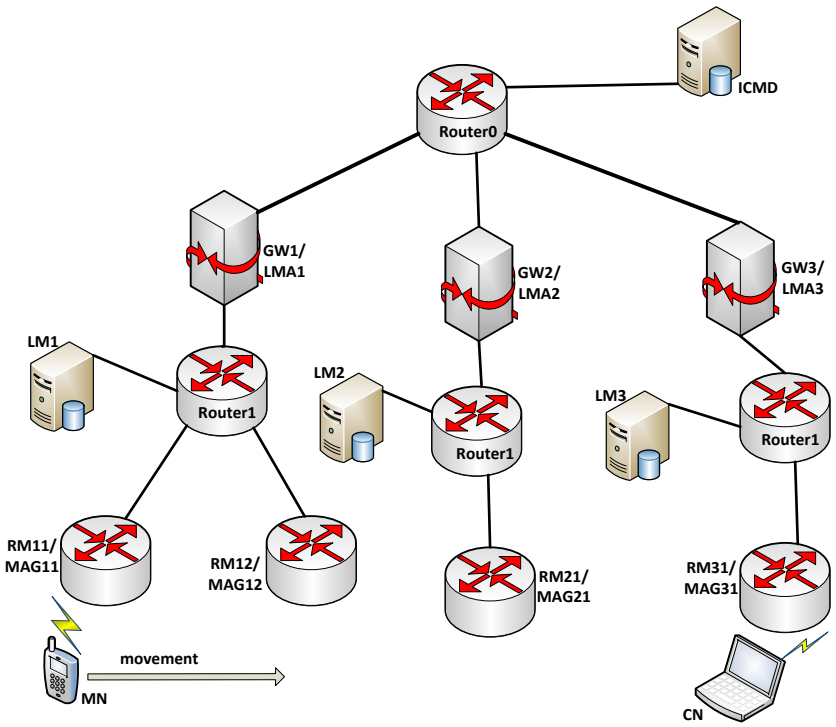


Figure 5-12 Topology used for simulation evaluation

The LM and ICMD run on servers, which implement the extended binding cache entry of the LMA. The RMs are configured as part of MAGs in DM-RMA. Each MAG implements a binding cache, which serves as a prefix table to store the address of the MAGs in the same network. It is assumed that in DP-PMIP, LMAs are aware of the ICMD's address; and it is further assumed that in DM-RMA, MAGs are aware of the LM's address in their respective networks. These are statically configured.

During the simulation, the MN configures an address from MAG11, which is used for communication with CN. Since both schemes optimize the route for the new traffic, the simulation focuses on the handoff traffic; hence the MN does not configure new address(es) in a visited network/domain. The MN movement is classified as intra- and inter-domain handovers. In the intra-domain handover, the MN moves from MAG11 to MAG12; whereas in the case of the inter-domain handover, the MN moves from MAG11 to MAG21.

The MN traffic in DM-RMA is anchored at MAG11, which also takes care of the route optimization at MAG31 during handover, as discussed in Section 5.3. MAG11 is configured to perform normal routing for the MN's traffic received; while the MN is still in its coverage, and to perform tunnelling for the traffic received, while the MN is away. This has been achieved by enabling MAG11 to install the RM function of the LMA, when it is notified about the new location of the MN. That is, the MAG classifier implementation in ns-2 has been modified to behave as an LMA classifier for handover traffic. This allows MAG11 to encapsulate packets of the MN, and to tunnel them to the new location.

In DP-PMIP, the LMA1 is configured to anchor the MN traffic, and to tunnel traffic to LMA2, when the MN performs an inter-domain handover from the GW1/LMA1 network to the GW2/LMA2 network.

The parameters used in the simulation for the wired network are given in Table 5-2. The wireless link has implemented IEEE 802.11b with 11 Mbps data rate. The CN transmitted CBR UDP packets of 1000bytes at 0.005sec packet intervals to the MN. The MN moves at a speed of 30m/s. The parameters were arbitrarily chosen for the purpose of simulation and validation of the performance of the proposed scheme. Hence, they are not a reflection of a specific real-world scenario.

Table 5-2 Parameters for simulation evaluation

Delay Router0-GW/LMA, Router0-ICMD	Delay Router1-GW/LMA, Router1-LM	Delay Router1-RM/MAG	Wired link bandwidth
2ms	1ms	0.25ms	100Mbps

5.6.2 Simulation Results and Analysis

This sub-section discusses the simulation results for both intra- and inter-domain scenarios for all the simulated schemes.

5.6.2.1 Intra-domain Handover Scenario

During the intra-domain handover, the impact of RM distributed at MAGs and the prefix table implemented at the MAGs on handover delay and the packet loss are investigated. From the simulation results (presented in Table 5-3), the average handover delay and packet loss for DM-RMA are 433.47ms and 86 packets, respectively. Yet, DP-PMIP has 441.94ms handover delay and 88 packets loss, respectively. It is important to note that the measured handover delays include layer 3 and layer 2 handover delays.

The relatively low handover delay and packet loss of DM-RMA are due to the fact that when the MN performs intra-domain handover, DM-RMA creates a tunnel between the old RM and the new RM (with the help of the prefix table) to carry the MN traffic without the involvement of a centralized entity, such as LMA in DP-PMIP. This tunnel creation approach reduces the latency that could be incurred (due to mobility signalling exchange) in switching the tunnel at LMA. Although there is only a small gain in terms of handover delay and packet loss, DM-RMA removes the data traffic concentration on the LMA; and it also overcomes other issues related to single points of failure and bottlenecks.

Table 5-3 Summary of intra-domain handover delay and packet loss

Evaluated metrics	DM-RMA	DP-PMIP
Average handover delay (ms)	433.47	441.94
Packet loss (pkts)	86	88

5.6.2.2 Inter-domain Handover Scenario

(a) Location update latency

Location update latency is analysed as two latency components. The first latency

component analysed is the time the new MAG needs to update the MN binding in the location management system (i.e., LM or ICMD), once it detects the MN attachment. The second latency component analysed is the time that has elapsed from when the new MAG detects the attachment of the MN to when the traffic anchoring node (i.e., RM, LMA) in the previous domain is updated about the new location of the MN. These analyses show the impact of decomposing and distributing the mobility functions on the location update latency.

From the simulation results (shown in Table 5-4), it has been observed that the new MAG (MAG21) in DP-PMIP needs 5.27ms on average to update the MN location at ICMD, and 9.28ms is spent to update this new location at LMA1 for the purpose of the tunnelling update. In the case of DM-RMA, it is observed that MAG21 needs 1.26ms to update the MN location at LM2 and 10.55ms to update this location at RM11 for the tunnelling update. DM-RMA uses a shorter time to update the location management system compared with DP-PMIP, because it has built-in distributed LMs. On the other hand, DP-PMIP uses a shorter time for tunnelling update at the previous domain because DP-PMIP updates the tunnel at the LMA, which is located at the gateway; while DM-RMA updates the RM collocated with MAG at the access network level.

Table 5-4 Summary of location update latencies

Evaluated metrics	DM-RMA	DP-PMIP
Location update latency at ICMD/LM2 (ms)	1.26	5.27
Tunnel update latency at LMA1/RM11(ms)	10.55	9.28

(b) Packet delivery latency

Packet delivery latency is measured in terms of end-to-end delay, which is the time the packet takes to travel from the CN to the MN. The packet delivery latency is used to investigate the performance of RM distribution and the route optimization mechanism implemented in DM-RMA, as shown in Figure 5-13 and Figure 5-14. Figure 5-13 shows the variation of end-to-end delay, as the MN performs the handover from MAG11 to MAG21. From the figure, it may be observed that before the handover, both schemes have similar end-to-end delay, and experience packet loss during handover. After the handover, the DM-RMA scheme has an end-to-end

delay, which is similar to the previous value – before handover. (But there are few packets in DM-RMA that experience high end-to-end delay due to traversing RM11 before the route is optimized.) In DP-PMIP, the end-to-end delay increases by about 45% after handover, because DP-PMIP does not optimize the route for the handover traffic; hence packets continue to be routed via the initial anchoring LMA (LMA1). In contrast, DM-RMA optimizes the route, and thus reduces the packet end-to-end delay for the handover traffic. The end-to-end delays before and after handover in DM-RMA appear the same, because the simulation assumes that all domains are placed at equal distances from each other. However, this is not always the case; and in some scenarios, the resulting end-to-end delay after handover may be smaller or larger than the one before the handover.

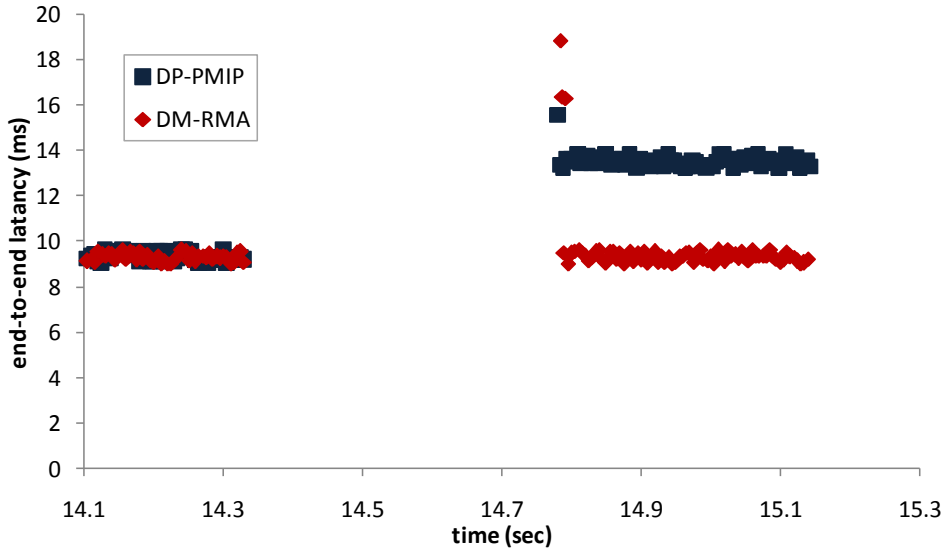


Figure 5-13 Packet delivery latency before, during and after handover

Similarly, Figure 5-14 illustrates the effects of increased distance between the domains/networks on the average end-to-end delay. The distance between domains is varied from 3ms to 15ms; while the other parameters are kept constant. The average end-to-end delay of both schemes is influenced by the variation of the distance between the domains. This is because the time required for the packet to travel from the CN to the MN increases as the CN network moves far away from the network serving the MN. Nevertheless, DM-RMA still outperforms DP-PMIP, because of its built-in route optimization.

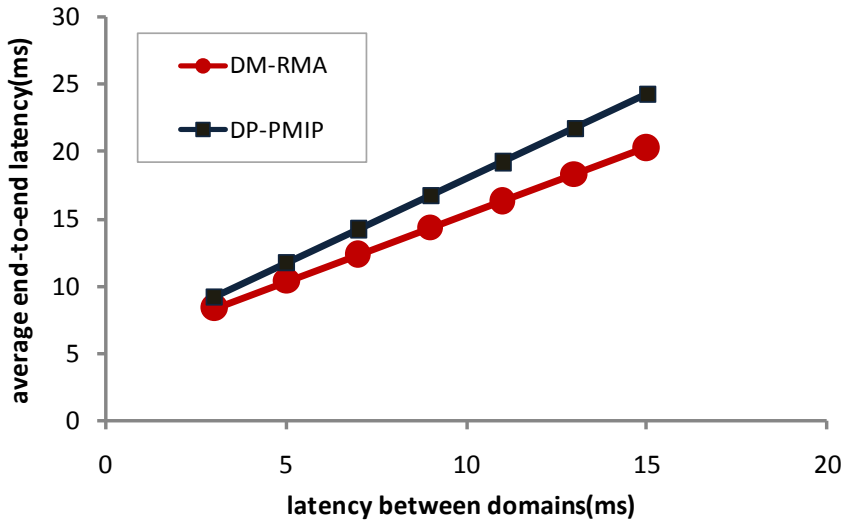


Figure 5-14 The impact of distance between domains on packet delivery latency

(c) Handover delay and packet loss

The impact of the distance between the distributed domains and the mobility functions on the handover delay and the packet lost during handover are investigated. Figure 5-15 shows the handover delay variation, according to the change of distance between the domains when the MN moves between domains. The number of packets lost during handover is represented by the number on the top of each bar. The handover delay in both schemes increases with the increase of the distance between the domains. This is because as the distance between the domains increases, the location update latency to the mobility entity (LMA, LM, RM) increases as well. This is a result of the long time taken by the exchange of mobility signalling (i.e., PBU and PBA) between the mobility entity involved in different domains. Nevertheless, DM-RMA outperformed DP-PMIP in terms of handover delay and packet loss, especially when the domain distance becomes large.

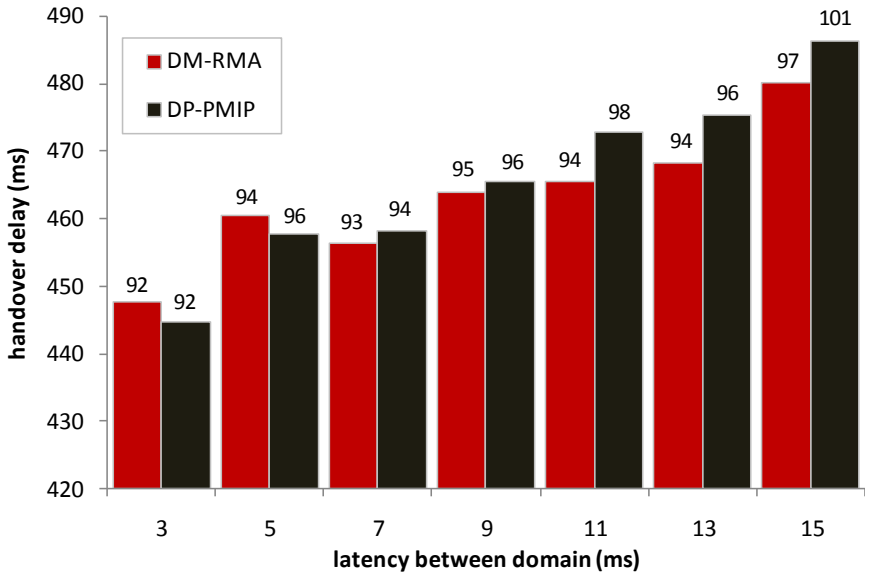


Figure 5-15 Inter-domain handover delay and packet loss

5.7 Summary

The chapter has presented a new network-based DMM scheme that splits the logical function of LMA, and co-locates the RM and the HNP allocation functions in distributed access routers with a mobility client function. The scheme is named DM-RMA. The chapter has discussed and illustrated with appropriate signalling diagrams, the design and the operation mechanisms of the DM-RMA scheme. An analytical model has been developed and used to evaluate the performance of the DM-RMA scheme, in terms of signalling cost, packet delivery cost, tunnelling cost and total cost. The scheme has also been compared with DP-PMIP and DF-PMIP schemes. Moreover, the performance of the proposed scheme has been evaluated through simulations conducted in ns-2.

The analytical results show that the DM-RMA has better packet delivery cost, tunnelling cost, and total cost compared with DP-PMIP and DF-PMIP. However, DP-PMIP outperforms DM-RMA in terms of signalling cost. Moreover, the ns-2 simulation results show that the DM-RMA scheme reduces end-to-end delay, handover delay, and packet loss.

Chapter 6 Route Optimization Mechanism for Proxy MIPv6 using CMAG

6.1 Introduction and Motivation

The Proxy MIPv6, discussed in previous chapters, is a network-based mobility management protocol standardized by IETF. It removes the mobility management functions from the MN, and implements them in the network, thereby addressing MIPv6 limitations. Accordingly, the need for alteration of the MN's protocol stack is alleviated, and the complexity of the MN, as well as the MN energy consumption is reduced.

The PMIPv6 basic standard [17] defines the setup and maintenance of the bidirectional tunnel between the MAGs and the LMA, in order to allow for the forwarding of data packets to/from the MN. The packets sent to/from the MN are encapsulated in this bidirectional tunnel, irrespective of the MN's remote communication end point. The communicating nodes may be attached to the same MAG, or to different MAGs within the same service provider (LMA) domain; yet the packets have to traverse the MN's LMA. This introduces a long routing path and increased packet delivery latency [85][14], especially when the LMA is topologically located far away from MAGs, such as in the core network.

However, the PMIPv6 basic standard does not address the routing path optimization; and instead, it leaves this case as an open problem [44].

Routing path optimization mechanisms can reduce the routing path between the communicating nodes in the PMIPv6 domain, and thereby reduce end-to-end packet delay. Various schemes have been proposed for routing path optimization for PMIPv6 in the literature [47][86][87]. These schemes either use a data packet or a query message to trigger the route optimization at the LMA. The trigger packet travels from the corresponding node's MAG (CN-MAG) all the way to the LMA. The LMA then processes the trigger packet, and returns the feedback to the involved entities, such as for example, the CN-MAG. Given that the LMA may be located far away from the CN-MAG, it can take a significant delay before the routing path is optimized. Consequently, a considerable number of packets will traverse a non-optimal path through LMA. Even if the packets arriving during the route optimization procedure are buffered

at CN-MAG, as in [86][88], a large buffer needs to be deployed to store these packets. In addition, when the MN performs handover to a new MAG, there may be an extra delay to rebuild the optimized routing state because of the path that the control messages have to travel before arriving at the LMA.

This chapter presents a new scheme that extends PMIPv6 to perform routing path optimization for an MN that is roaming in a PMIPv6 domain. The scheme provides fast route optimization; and it mitigates packet loss during route optimization and handover procedures. It introduces a Coordinating Mobile Access Gateway (CMAG). The CMAG is a MAG that is strategically located at the shortest path to other MAGs in the PMIPv6 domain. It provides other MAGs in the domain with the MN's mobility information, to setup and maintain an optimized routing path. The new MAG to which the MN hands over can access the old MAG information and MN's HNP from CMAG.

The main goal behind introducing the CMAG is to shorten the path that the control messages have to travel, in order to get the MN's binding information. This design can reduce the time taken in setting up an optimized routing path, and the time required to rebuild the optimized routing path when an MN performs a handover, and thereby reduces the handover delay and the packet loss. The scheme could be used by network operators to reduce the data traffic burden to the core network; while they are preparing to adopt the distributed mobility management schemes discussed in the previous chapters.

6.2 Related Work

PMIPv6 standard [17] addresses the MIPv6 [20] limitations by removing the mobility management functions from the MN and moving them to the network. The PMIPv6 transport mechanism requires that all data traffic should go through the LMA – even though there may be a shorter path between the MAGs of the communicating nodes. This causes a data traffic burden on the LMA; and it increases the communication delay. The reason is that PMIPv6 does not address the routing path optimization in its basic standard [44].

H. Jung *et al*.[86] and J.W. Park *et al*. [88] propose schemes that extend the PMIPv6 to reduce the processing burden at LMA. These schemes establish a data tunnel between the MN's MAG and the CN-MAG. When the CN-MAG receives a packet from the MN, it starts buffering

the incoming packets and sends a query to LMA, so as to get a proxy care-of-address (CoA) for the MN. The LMA then processes the query message, and sends to CN-MAG an acknowledgement message with the MN's proxy CoA. Given that the LMA may be topologically located far away from CN-MAG, the scheme may incur a long delay before it receives the MN's proxy CoA from LMA. This can result into a considerable delay before the optimized route is established. Consequently, a significant number of packets will be buffered at the CN-MAG – before the route is optimized. Hence, a big buffer size may be needed at CN-MAG to avoid packet loss. Moreover, the schemes have not addressed the re-establishment of the optimized route, when the MN performs handover.

Q. Wu and B. Sarikaya [87], as well as S. Krishnan *et al*. [47] discuss localized routing for PMIPv6 (LR-PMIPv6). They propose mechanisms to establish a tunnel between the source and the destination MAGs in the PMIPv6 domain. The data packets are sent from the CN to the CN-MAG, and then to the LMA. When the LMA receives the packets, it makes a decision regarding localizing the routing path, and sends the localized routing initiation (LRI) message to both the CN-MAG and the destination MAG. Then, the destination MAG and the CN-MAG establish a bi-directional tunnel between them. The data traffic from the CN to MN travels through the established tunnel. Similarly, when the MN performs handover, the localized routing is re-established by LMA upon receiving the proxy binding update (PBU) from the new MAG. Although, the mechanisms establish an optimized routing path, the signalling to setup the optimized routing path may experience a long path delay towards MAGs, especially when the LMA is located far away from MAGs. Therefore, it may take a long delay before the localized routing is established; and many packets will traverse a non-optimal route via LMA. In addition, the mechanism may lead to a significant packet loss during handover, as well as to an increased handover delay due to re-establishing the localized routing after the MN has performed handover.

6.3 Architecture of PMIPv6 with CMAG

This section discusses the design of the proposed scheme for routing path optimization of PMIPv6 with a coordinating MAG (CMAG). Figure 6-1 depicts the architectural framework of the scheme. The main goal is to design a new scheme that can reduce the delay in getting the MN's binding information, so as to allow for fast routing path optimization. This design will

reduce the data packets traversing the non-optimal path via LMA, as well as the data packets being buffered during the route path optimization process. This is achieved through making the MN's mobility information available at MAG level, by introducing CMAG. Moreover, the design considers alleviating the data traffic load at LMA in PMIPv6 through building a data routing path between the MAGs under the same network service provider administrative domain.

6.3.1 Architecture Overview

PMIPv6 with CMAG (P-CMAG) is composed of a Control Function entity (CF), a coordinating MAG and the MAGs.

The control function (CF) is a modified LMA in PMIPv6 [17]. It performs all the LMA functions, in addition to the new function introduced in this design. For example, it provides the CMAG with the MN's mobility information.

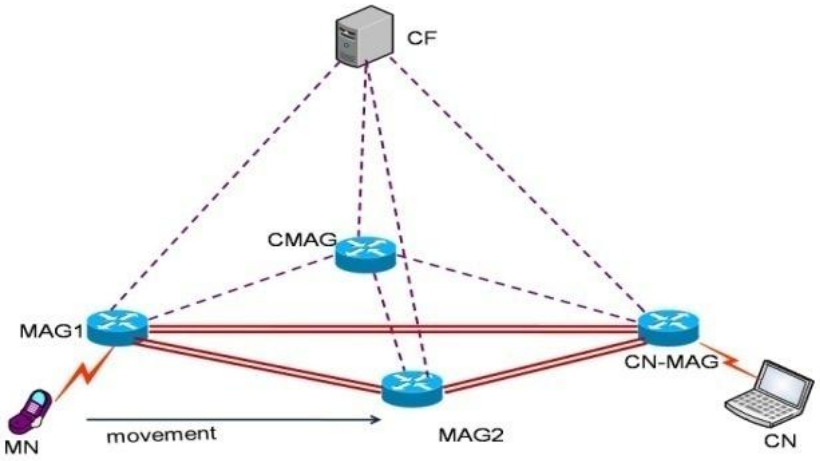


Figure 6-1 Architecture for PMIPv6 with CMAG

The coordinating MAG (CMAG) is a MAG that is located in an optimal path to all MAGs in the PMIPv6 domain. That is, CMAG is a MAG in the PMIPv6 domain, which is situated at the shortest distance from the other neighbouring MAGs. It is strategically selected by the network administrator, preferably, a MAG that is situated at the centre of the PMIPv6 domain. It manages the MN's mobility information in a newly defined list table, named the MAG List Table (MLT). The MLT maintains mobility information, such as the address of MAG currently serving the MN (proxy CoA of the MN), MN-ID, and HNP. Other MAGs in the domain can access the mobility information from CMAG, when they need such information. For

example, the CN-MAG (Correspondent Node MAG) gets MN's proxy CoA from CMAG, when it receives packets from the CN, in order to optimize the route. Similarly, the new MAG gets the MN HNP, when an MN performs a handover to its access network. It is important to note that the CMAG is consulted on the control plane only, such as during route optimization or handover signalling.

Given that CMAG is positioned on the same access network level as other MAGs in the domain, the MN's mobility information obtained from CMAG becomes quickly available to other MAGs, when compared to accessing them from LMA. This feature can reduce the latency to optimize the routing path, handover delay and packet loss.

The mobile access gateway (MAG) performs all the MAG's tasks defined in PMIPv6. However, in P-CMAG, MAGs are extended to get the proxy CoA of the MN from CMAG, when they receive packets from the CN attached to their network, so as to optimize the routing path. Moreover, the proposed scheme assumes the existence of a level of security association among MAGs, and between MAGs and CMAG.

6.3.2 The Mechanism of Operation of PMIPv6 with CMAG

This sub-section discusses in detail the operation mechanism of P-CMAG. The discussion includes the procedure for the initial attachment of the MN, the establishment and maintenance of an optimized routing path, and the handover procedure.

6.3.2.1 MN Initial Attachment Procedure

When an MN enters the PMIPv6 with CMAG domain, it attaches to a MAG, for example, MAG1, as shown in Figure 6-2. The MAG1 detects the MN attachment, gets the MN's identifier (MN-ID), and performs the access authentication, as described in PMIPv6 [17]. Then, MAG1 sends to the CF the PBU, so as to register the MN. In parallel, MAG1 sends a binding request to CMAG through a new defined Binding Request (BReq) message – asking CMAG for the MN's previous binding information, such as the address of the old MAG. Upon receiving the BReq message, the CMAG performs a lookup for the previous MN's binding from its MLT. Since the MN is attaching to the domain for the first time, the CMAG finds no previous binding information for the MN. Then, the CMAG creates an entry in its MLT for the MN, and records the MN-ID and MAG1 address. Then, the CMAG responds to MAG1 with a binding-reply

(BRep) message that includes an empty address field for MN's old MAG.

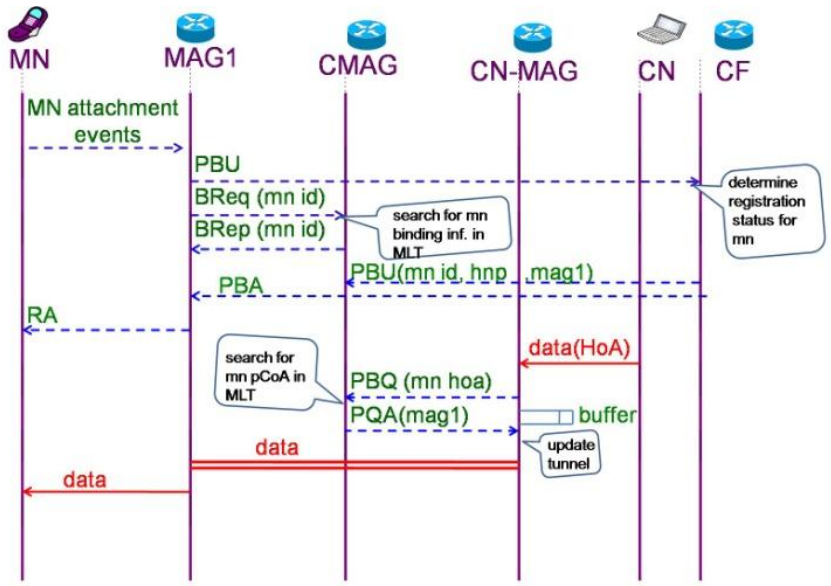


Figure 6-2 Signalling flow for MN registration and establishment of the optimized routing path for PMIPv6 with CMAG

On the other hand, when CF receives the PBU, it performs a lookup from its cache entry, in order to determine the registration status of the MN. Because the MN is attaching to the domain for the first time, the registration status shows a new MN. Hence, CF allocates an HNP to the MN, and responds to MAG1 with the proxy binding acknowledgement (PBA). In parallel, the CF notifies CMAG of the attachment of the new MN to the domain, through the modified PBU message. The modified PBU message is the newly defined message, which is delivered from CF to CMAG, so as to inform the CMAG about the attachment of the new MN in the domain. The message includes information, such as the address of the MAG to which the MN is currently attached, i.e. MAG1, HNP and MN-ID. This facilitates the availability of the binding information for the registered MNs closer to the MAGs.

When MAG1 receives the BRep message from CMAG with an empty address field for MN's old MAG, it realizes that the MN is attaching to the domain for the first time. Hence, it does not attempt to optimize the routing path. On the other hand, when MAG1 receives the PBA message from CF, it caches the MN's binding in its Binding Update List (BUL). Finally, MAG1 advertises the HNP to the MN, and the MN configures the HoA from the advertised HNP. Now the MN is ready to send/receive packets to/from the CN.

6.3.2.2 A Mechanism for Fast Routing Path Optimization

When a CN has a packet to send to the MN, it forwards the packet destined for MN's HoA (MN-HoA) to CN's MAG (CN-MAG). The CN-MAG then checks if it has a proxy CoA for the MN, using the packet destination address information. If the proxy CoA is found, it forwards the packet directly to the MN's MAG. If the CN-MAG finds no proxy CoA for the MN, it starts buffering the packets and sends to CMAG a proxy binding-query message, PBQ (MN-HoA), so as to obtain the MN's proxy CoA. This is shown at the lower part of Figure 6-2. Upon receiving the PBQ, CMAG looks up the MLT for proxy CoA of the MN using MN-HoA information.

CMAG then responds to CN-MAG with the proxy CoA (i.e., the address of MAG1). Now, the CN-MAG tunnels the buffered packets, followed by the incoming packet to MAG1. The packets are tunnelled through the optimized path between CN-MAG and MAG1. Because CN-MAG gets the proxy CoA for MN from CMAG, the routing path optimization is faster than when the proxy CoA comes from LMA or CF.

6.3.2.3 Handover Procedure and Maintenance of Optimized Routing Path on Handover

Figure 6-3 shows the signalling flow diagram when an MN performs handover from MAG1 to MAG2, and the maintenance of the optimized routing path, when the MN performs handover.

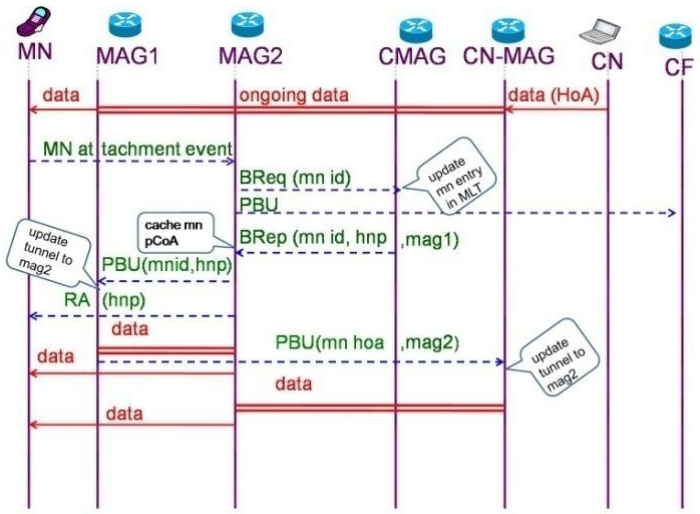


Figure 6-3 Handover procedures and optimal routing path updates when the MN performs handover signalling flow of the proposed scheme

From Figure 6-3, when the MN moves to MAG2, MAG2 detects its attachment and sends to CF the PBU, so as to update the MN's registration. In parallel, MAG2 sends to CMAG a BReq message, so as to get the IP address of MN's old MAG, and possibly, the MN HNP. Upon searching its cache entry the CF realizes that the MN is already registered in the domain. CF then updates its cache entry and leaves MAG2 to get the MN HNP from CMAG. The cached entry at CF is used to forward the data packets in a scenario where the CN is located outside the PMIPv6 with CMAG domain.

On the other hand, the CMAG updates the MAG address field to MAG2 address in its MLT, and responds to MAG2 with the address of MAG1 and MN HNP. Then MAG2 sends the router advertisement to the MN, and tells MAG1 to forward the MN's packet to it through the modified PBU message, PBU (MN-ID, HNP). Now, MAG1 tunnels packets to MAG2 and notifies CN-MAG about the address of MAG currently serving the MN, MAG2, through PBU (MN-HoA, MAG2 address), in order to update the tunnel. CN-MAG updates the tunnel to MAG2 and tunnels the incoming packets from CN to MAG2. The optimized routing path is then re-directed to MAG2. To mitigate the packet loss during handover, the scheme employs a buffer at MAG1.

6.4 Performance Evaluation

In this section, the performance evaluation of P-CMAG, using ns-2 (release2.29) [66] (briefly discussed in Section 3.5.1) with NIST mobility package [67], is conducted. The simulation environment extends the implementation of PMIPv6 in [67] to implement P-CMAG, signal-driven PMIP (S-PMIP) [86], and LR-PMIPv6 [47].

6.4.1 Simulation Scenario in ns-2

Figure 6-4 shows the simulated topology, in which CF (in P-CMAG) and LMA (in S-PMIP and LR-PMIPv6) are co-located. The CMAG is implemented on the access network level with other MAGs in the PMIPv6 domain.

In many operators' networks, the LMA is located at the gateway within the core network [44]; and the topological distance between LMA and MAG is longer than that between neighbouring MAGs. Hence, the simulation considers the distance between MAGs and LMA to

be significantly larger than the distance between MAGs.

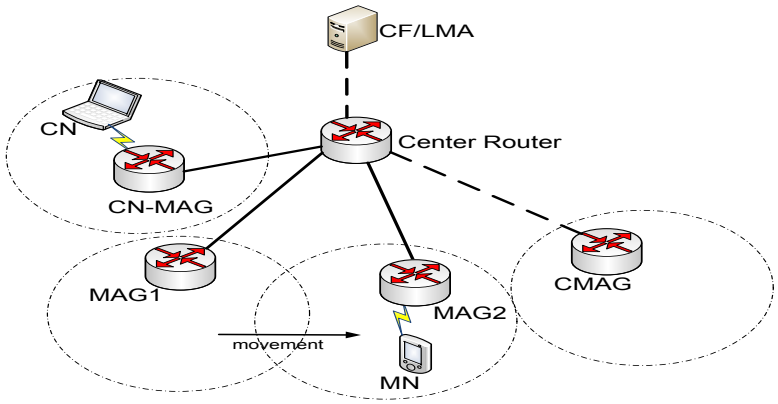


Figure 6-4 Network topology for simulation

In the simulation, a mobile CN is attached to CN-MAG inside the PMIPv6 domain. The mobile CN transmits CBR UDP traffic with a packet size of 1000bytes to the MN. The interval between successive packets is 0.005seconds. All wired links are configured with a bandwidth of 100Mbps. The data rate for the wireless link between MAGs and CN/MN is configured as 11Mbps. The MN starts from MAG1 and moves to MAG2 with an arbitrary velocity of 30m/s, while communicating with its CN.

P-CMAG deploys a buffer at CN-MAG (as discussed earlier) to mitigate packet loss during the route optimization process. Also, a buffer is deployed at MAG1 to mitigate the packet loss during handover. The evaluation investigates these two scenarios – both when the buffer is deployed at MAG1, and when it is not.

The design of P-CMAG focuses on mitigating the delay incurred in setting up and maintaining the optimized routing path, as well as the packet loss in the PMIPv6 domain. Hence, the performance evaluation considers the total delay incurred to establish an optimized routing path (i.e., route optimization-establishment latency), the number of delayed packets due to either buffering at CN-MAG or traversing LMA during the route optimization operation, handover delay, packet loss, and packet delivery latency as the performance metrics. The delay incurred in establishing the optimized routing path is defined as the time elapsed from when the CN-MAG receives the first packet from CN to when it gets the proxy CoA of the MN. Other evaluation parameters are defined, as in the previous chapters.

The performance evaluation compares P-CMAG with S-PMIP [86] and LR-PMIPv6 [47]

in terms of the delay incurred to establish an optimized path, the number of delayed data packets, and packet delivery latency. Since S-PMIP discusses only the establishment of the optimized route in PMIPv6, and does not consider a scenario, where the MN undergoes handover, the evaluation has compared P-CMAG with standard PMIPv6 [17] and LR-PMIPv6 for handover delay and packet loss performances. LR-PMIPv6 [47] is the current standardized route optimization for PMIPv6, as given by IETF. It is, therefore, worthwhile to compare the improvement gain of P-CMAG with LR-PMIPv6.

6.4.2 Simulation Results and Analysis

This sub-section contains the simulation results, which show the performance of the proposed scheme in comparison with other schemes reviewed under the related work.

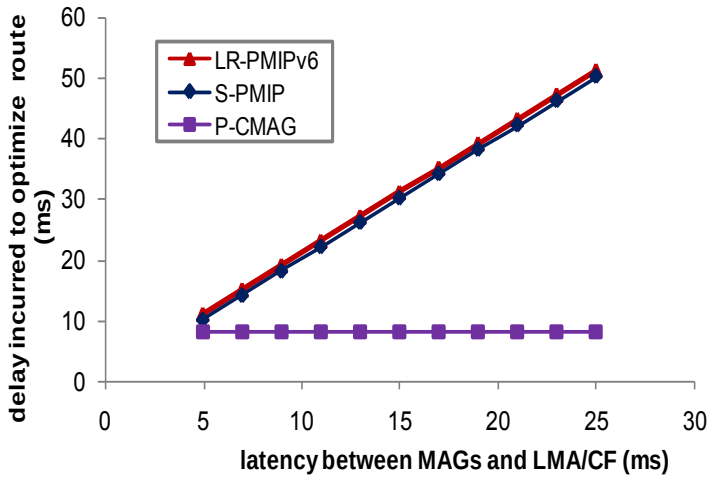
6.4.2.1 Route Optimization-Establishment Latency and Delayed Data Packets

Figure 6-5 depicts the impact of the latency between MAGs and LMA on the delay incurred to setup an optimized routing path, as well as the average number of delayed data packets between P-CMAG, S-PMIP and LR-PMIPv6 during the route optimization operation. The simulation results show the advantage of introducing CMAG in the route optimization process when the latency between MAGs and LMA/CF is varied from 5ms to 25ms, and the latency between MAGs is maintained at 4ms. The data rate and the network load for all schemes are kept the same.

From Figure 6-5a and Figure 6-5b, it is observed that the route optimization-establishment latency and the number of the delayed data packets during the route optimization operation for S-PMIP and LR-PMIPv6 increase as the distance between the LMA and the MAGs increases. The reason is that both S-PMIP and LR-PMIPv6 rely on LMA to optimize the route. Hence, when the LMA is far away from MAGs, the round trip time for the route optimization signalling messages significantly increases. This causes an increase in route optimization-establishment latency. As a result, a huge number of packets arriving at the CN-MAG before completion of the route optimization operation are delayed: either due to buffering at CN-MAG (e.g., in S-PMIP), or to traversing the LMA (e.g., in LR-PMIPv6). So, S-PMIP may need to deploy a large buffer size when LMA is placed far away from MAGs, in order to handle the arriving packets during the route optimization operation (thereby avoiding packet loss). In

contrast, P-CMAG mitigates these effects by utilizing the CMAG in optimizing the route, which is on the access network level with other MAGs (the latency between them is low).

a. Impacts of various delays between MAGs and LMA on route optimization-establishment latency



b. Impacts of various delays between MAGs and LMA on the delayed data packets

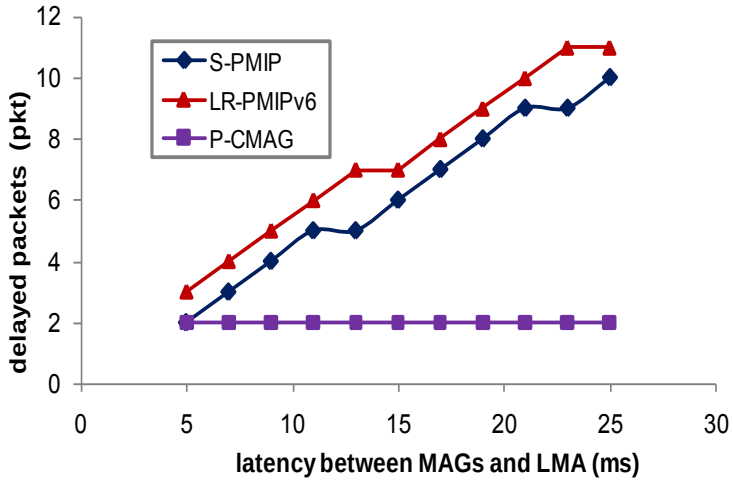


Figure 6-5 The impacts of latencies between MAGs and LMA over the route optimization-establishment latency and the amount of delayed data packets

6.4.2.2 Data Packet Delivery Latency

Figure 6-6 shows the influence of the latency between MAGs and LMA on data packet delivery latency. This influence is compared among the candidate schemes by varying latency between MAGs and LMA, as explained above – without altering the latency between the MAGs.

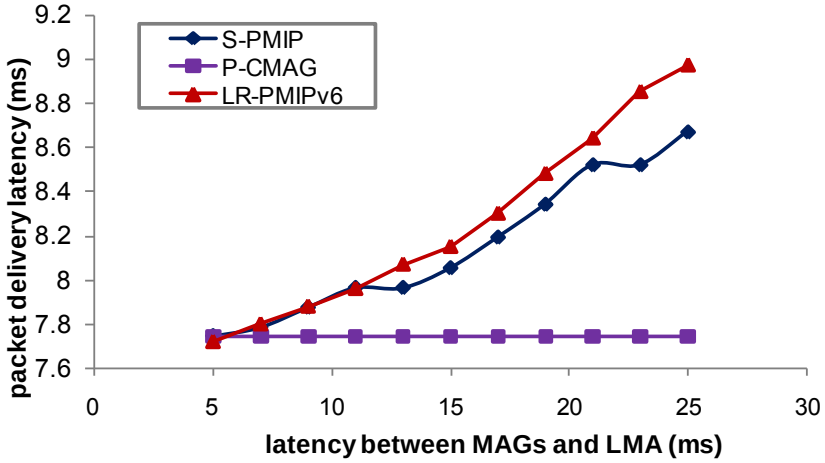


Figure 6-6 The impacts of various delays between MAGs and LMA on average end-to-end delay

From the figure, all schemes have almost similar average packet delivery latency at low latency between MAGs and LMA. When this latency is increased, P-CMAG maintains the same average packet delivery latency; whereas other schemes respond with a positive increase in average packet delivery latency. This is due to the fact that the increase in latency between MAGs and LMA in S-PMIP and LR-PMIPv6 causes many delayed packets, as depicted in Figure 6-5. Consequently, the average packet delivery latency becomes large. Since the delayed packets in P-CMAG are independent of this latency, the P-CMAG maintains constant packet delivery latency.

6.4.2.3 Impact of Traffic Transmission Rate on Delayed Data Packets

The result in Figure 6-7 compares the impacts of various CBR UDP traffic transmission rates on the number of delayed packets before the route is optimized between P-CMAG, S-PMIP, and LR-PMIPv6. In this scenario, the latency between neighbouring MAGs is set at 4ms; the latency between the MAGs and the LMA is set at 25ms; and the data traffic rate is varied from 200packets/second to 1600packets/second. The results show that the amount of delayed packets before the route is optimized increases as the data traffic rate increases for all schemes.

Nevertheless, P-CMAG has a lower number of delayed packets than other schemes at various data traffic rates. This is because P-CMAG shortens the route optimization-establishment latency by utilizing the CMAG.

However, when the data traffic rate exceeds the saturation throughput, there is a decrease in the number of delayed packets before the route is optimized in all schemes. This is due to the increase in end-to-end delay, which is caused by the queuing effect at the CN. The results indicate that at a higher data transmission rate the route optimization mechanism of LR-PMIPv6 will cause overhead at LMA, due to the high number of first data packets going through a non-optimal path via LMA.

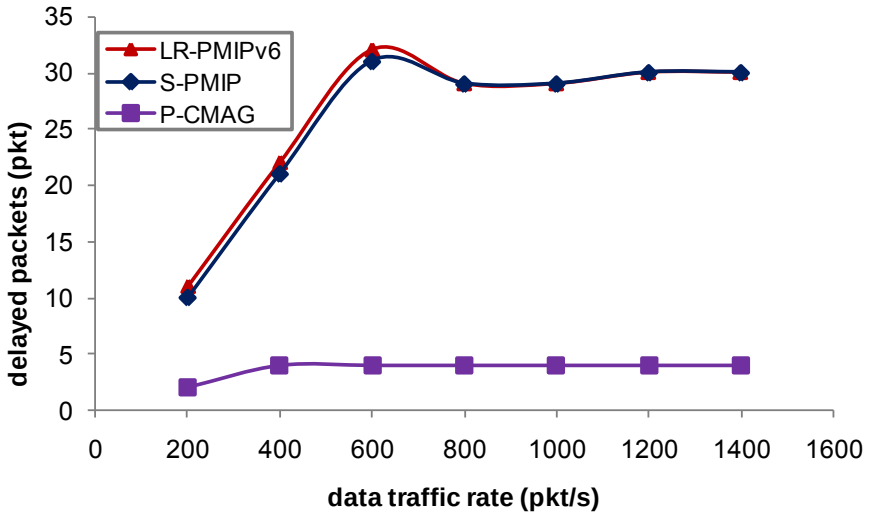
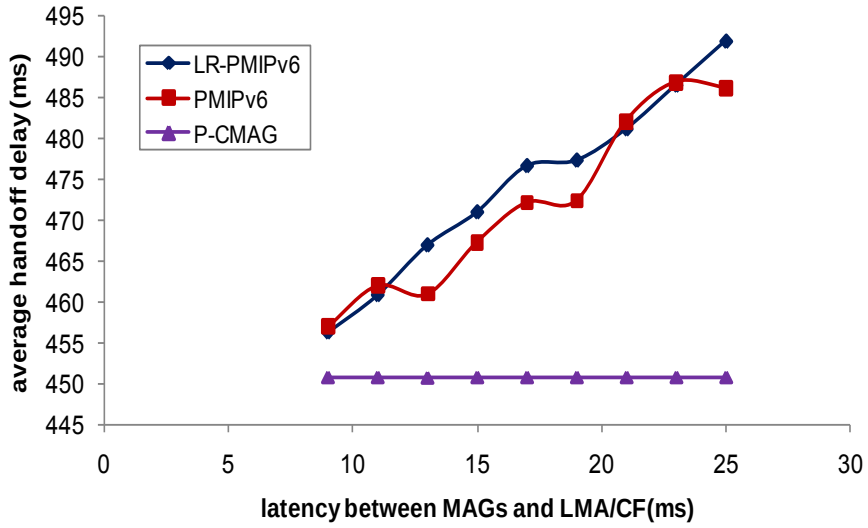


Figure 6-7 Influence of data traffic transmission rates over average number of delayed data packets during the routing path optimization procedure

6.4.2.4 Impact on Handover Delay and Packet Loss due to Latency between MAGs and LMA

Figure 6-8 compares the impacts of the latency between the MAGs and the LMA on handover delay and packet loss performance of P-CMAG, LR-PMIPv6, and PMIPv6.

a. Impacts of latencies between MAGs and LMA on handover delay



b. Impacts of latencies between MAGs and LMA on packet loss

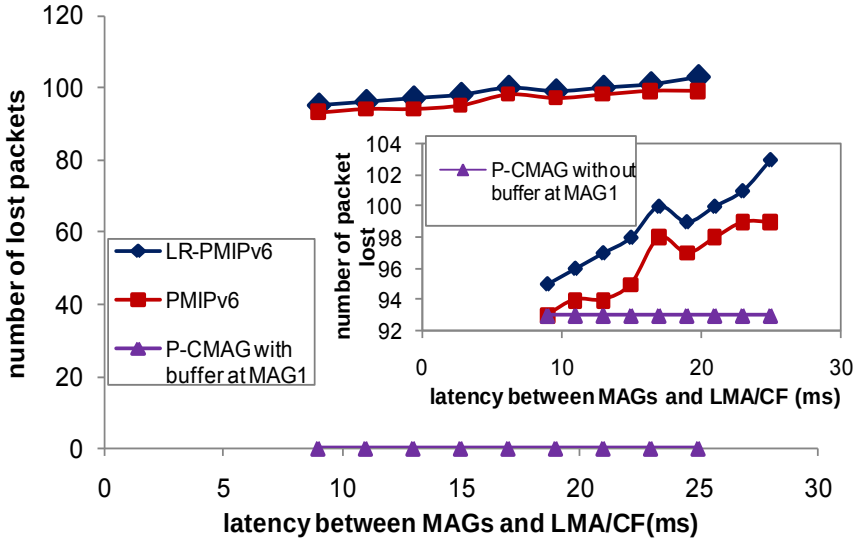


Figure 6-8 The impacts of various delays between MAGs and LMA on handover delay and packet loss

From Figure 6-8a and Figure 6-8b, the handover delay and the packet loss of P-CMAG are not influenced by the latency between the MAGs and the LMA. However, the LR-PMIPv6 and PMIPv6 schemes are highly affected; and the impact becomes severe, as the latency between the MAGs and the LMA gets longer. This is due to the fact that whenever the MN performs a

handover; layer 3 mobility signalling is exchanged between the new MAG and the LMA, in order to update the route. Moreover, the LMA in LR-PMIPv6 needs to re-establish the optimized route. The time to complete mobility signalling between MAG and LMA in PMIPv6 and LR-PMIPv6 depends on the latency of signalling message exchange between LMA and MAG. Whenever this latency increases, the delay in completing the mobility signalling process increases as well. Consequently, an increase in handover delay and packet loss is experienced.

On the other hand, in P-CMAG, the MAG that the MN hands over to gets the MN's binding information from the CMAG, which is situated at the shortest distance to other MAGs. Consequently, it allows the MN to resume its communication faster. This further reduces the handover delay and packet loss. Moreover, the buffer introduced at MAG1 mitigates the packets loss to zero, as is shown in Figure 6-8b. Figure 6-8b also shows the packet loss performance improvement of the proposed scheme – without using a buffer at MAG1. The proposed scheme reduces the packet loss in comparison with PMIPv6 and LR-PMIPv6 for various latencies between the MAGs and the LMA.

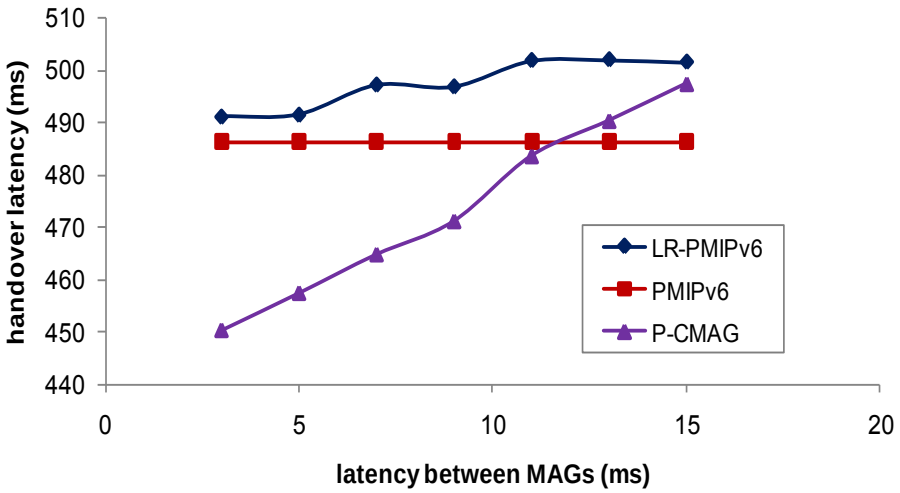
6.4.2.5 Impact on Handover Delay and Packet Loss due to Latency between MAGs

The graphs in Figure 6-9 show the impact of various latencies between MAGs on handover latency and packet loss of the proposed schemes, LR-PMIPv6 and PMIPv6. To investigate these impacts, the latency between the MAGs and the LMA is fixed at 25ms, and the latency between MAGs isvaried from 3ms to 15ms. From Figure 6-9a and Figure 6-9b, at low latencies between MAGs, P-CMAG shows a better performance in terms of handover latency and packet loss than LR-PMIPv6 and PMIPv6. However, at large latencies between MAGs, the performance of PMIPv6 gets better than those of P-CMAG and LR-PMIPv6. This is because the new MAG in P-CMAG relies on CMAG to get the MN's binding information (such as the MN's old proxy CoA), which allows the MN to receive packets at the new MAG after the handover, and then to optimize the route.

So, as the distance between MAGs increases, the delay in completing the hand-over procedure increases as well, resulting in an increase in hand-over delay and packet loss. On the other hand, in LR-PMIPv6, the new MAG relies on LMA to update the optimized route at CN-MAG (i.e. the route that allows for the forwarding of packets from the CN-MAG to the new

MAG). Hence, the latency between the MAGs and the LMA, and the increased latency between MAGs, cause the increase in packet loss and handover delay.

a. Impact of delays between MAGs on handover delay



b. Impact of delays between MAGs on number of lost packets

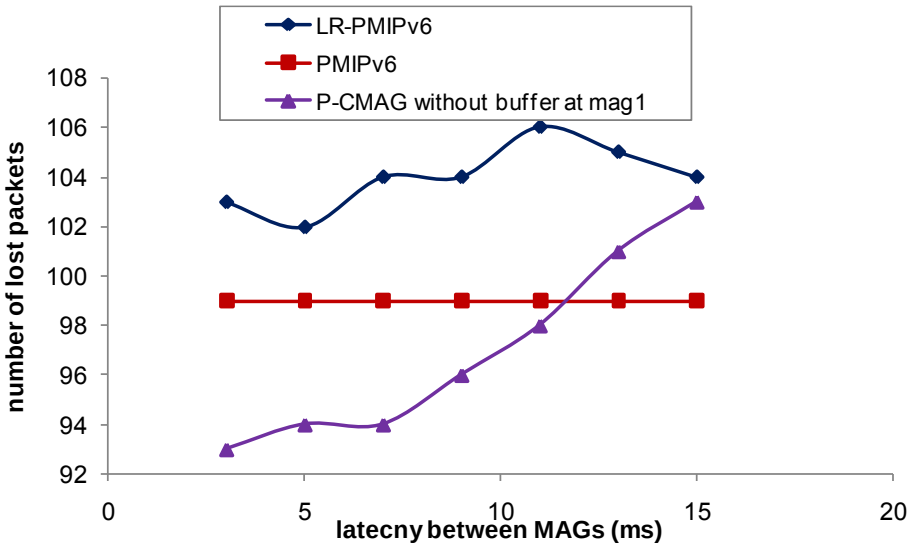


Figure 6-9 The impact of various latencies between MAGs on handover delay and packet loss

Typically, the average latency between MAGs is considerably less than the latency between MAG and LMA [44]. So, P-CMAG can achieve a smaller handover delay and less

packet loss. In addition, at large latencies between MAGs, P-CMAG gains the advantage of removing the data traffic load and the processing effect on the LMA. However, CMAG causes some overhead, such as additional messages when the MN performs handover, and needs to maintain MLT for the MN's binding information, which includes the address of the old MAG.

6.5 Summary

This chapter has presented a route optimization scheme for PMIPv6 with coordinating MAG (CMAG), P-CMAG. The CMAG shortens the path the control signal messages have to traverse compared with the path towards the LMA. The shortened path reduces the delay in the setting up of the optimized route. Moreover, the CMAG provides the new MAG where the MN is likely to attach with MN's mobility information, which enables quick establishment of the tunnel between the new MAG and the old MAG. This quick establishment reduces in-flight packet loss. The simulation demonstrated that P-CMAG provides fast route optimization and reduces average packet delivery latency compared with S-PMIP and RL-PMIPv6.

Also, the results show that as the LMA is moved away from the MAGs, P-CMAG reduces the number of packets buffered at the CN-MAG before the route is optimized. Moreover, P-CMAG mitigates handover delay and packet loss, compared to standard PMIPv6 and LR-PMIPv6.

Chapter 7 Conclusion and Future Work

This chapter concludes the thesis and gives some recommendations for future work.

7.1 Conclusion

Distributed Mobility Management (DMM) is a new paradigm for developing efficient mobility management schemes, in order to address the limitations of centralized mobility management schemes. The limitations of the centralised mobility management are more prominent, because of the current rapid increase in the number of mobile users and data traffic volumes, along with the evolution in the mobile networks towards flat network architecture. Thus, the centralized mobility management schemes are not designed to cope with the increase in traffic volume and the current network evolution, in a cost effective manner.

To promote the DMM approach, IETF has formed a DMM working group and proposed the design requirements. One of the key requirements is the need to re-use the existing IP mobility protocols and their extensions in developing DMM schemes. Therefore, it is vital to enhance the existing IP mobility management with a DMM approach. A number of IP mobility management schemes based on the DMM approach have been proposed. However, many of these schemes are at a preliminary stage, and still have some limitations. Moreover, many of these schemes still need to be analysed, in order to determine their feasibility.

In this thesis, an in-depth literature review of the existing centralized IP mobility management and recent distributed mobility management schemes has been conducted. The review has highlighted some of the limitations of the existing schemes regarding coping with the rapid increase in the number of mobile users and data traffic volumes. In addition, the review has showed the trends and limitations of the existing DMM schemes. These motivated the research carried out in this thesis.

The thesis has developed three novel network-based distributed mobility management schemes, which are based on the DMM approach. In addition, one new route optimization scheme for PMIPv6 has been developed. The following section summarizes the contributions under the four schemes developed by this thesis.

In Chapter 3, a network-based DMM scheme named DM-RMG is developed for

enhancing PMIPv6 to operate in a distributed manner. The scheme releases the load burden of the LMA in PMIPv6 by splitting its logical mobility management functions into internetwork location management (LM), home network prefix (HNP) allocation, and routing management (RM) functions; and subsequently, distributing the routing management function to the gateways of different networks. Ultimately, the data-plane routing function for the MNs is locally served by the closest routing management function at the network gateway. The scheme incorporates a routing path optimization mechanism that optimizes data path for the handover traffic of the MNs. Moreover, the scheme introduces a seamless handover mechanism, named tracking MAG (TMAG) to reduce the inter-network handover delay.

Analytical and simulation results show that DM-RMG reduces the long packet delivery latency due to triangle routing, when compared to the centralized IP mobility schemes and the previous DMM schemes that employ static traffic anchoring. In addition, the simulation results show that the handover mechanism that utilizes the TMAG has shorter, negligible, inter-network handover delay and small packet loss, when compared to one that does not use the TMAG.

In Chapter 4, a network-based DMM scheme, called NDM-RMG is proposed to offer mobility support for both non-nested and nested NEMO scenarios. The NDM-RMG scheme is designed along principles similar to those of DM-RMG. The objectives of the scheme are to reduce the pinball routing problem and the high packet overhead in NEMO, especially in complex nested NEMO scenario. The performance of the proposed NDM-RMG scheme is compared with the performance of the NEMO Basic Support protocol and the N-DMM scheme. The analytical results show that NDM-RMG reduces packet delivery latency, binding update cost and packet delivery cost when compared with the NEMO Basic Support protocol.

The results also show that NDM-RMG and N-DMM have comparable performances in terms of packet delivery latency, binding update cost and packet delivery cost, in cases when the mobile networks are close to the anchoring networks. However, NDM-RMG outperforms N-DMM when the mobile network moves far away from the anchoring point. In addition, the ns-2 simulation results show that the packet delivery latency and the packet header overhead for NDM-RMG are negligibly affected by the level of nesting.

In Chapter 5, a DM-RMA scheme is proposed to remove the need for tunnelling between RMs (co-located at the gateways) and the access routers of DM-RMG and NDM-RMG schemes.

Additionally, the proposed scheme solves the problems of a single point of failure, traffic bottlenecks, and triangular routing – that are present in centralized IP mobility management. The scheme splits the logical mobility function of LMA in PMIPv6 in a similar way, as does DM-RMG; but it co-locates the RM and HNP allocation functions to the distributed access routers with mobility client function.

The DM-RMA scheme provides the optimal path for both handover traffic and new traffic of the mobile nodes; and it releases the load burden from the centralised mobility anchor. The performance of the proposed DM-RMA scheme is compared with the performance of DP-PMIP and DF-PMIP schemes. The numerical results show that the DM-RMA scheme outperforms DP-PMIP and DF-PMIP in terms of packet delivery cost, tunnelling cost and total costs. In addition, ns-2 simulation results show that DM-RMA reduces the location update latency at the location managers, packet delivery latency, and intra-domain handover delay, as well a packet loss, when compared with DP-PMIP.

Finally, in Chapter 6, a new route optimization scheme for PMIPv6 is proposed. The proposed scheme extends PMIPv6 to perform route optimization for an MN that is roaming in a PMIPv6 domain. The goals of the scheme are to resolve the sub-optimal triangle route in PMIPv6, and to reduce the route optimization-establishment latency, delayed packets, and packet loss. The performance of schemes is compared with LR-PMIP, S-PMIP, and PMIPv6. The ns-2 simulation results show that the proposed scheme reduces the route optimization-establishment latency, the delayed data packets during the route optimization operation, and the average packet end-to-end delay, when compared with S-PMIP and LR-PMIPv6. The results also show that the proposed scheme reduces the handover delay and the packet loss when compared with the standard PMIPv6 and LR-PMIPv6.

7.2 Future Work

Following the development of different schemes during the research for this thesis, there are some interesting issues that need further investigation, in order to improve the design, the applicability and the performance of the proposed schemes. Further research in the following directions is therefore recommended:

(a) The proposed DMM schemes consider a large domain partitioned into sub-networks

under a common administrative domain. The schemes also offer DMM support to the mobile nodes moving between networks under different administrative domains having certain levels of trust relationship. It is possible that the mobile nodes' movement can involve networks under different administrative domains, with different policies, and security mechanisms. However, this creates security issues. Further studies can address the security issues under this scenario – by proposing an appropriate security mechanism. Moreover, the proposed route optimization mechanism may have administrative policy issues, such as charging. Future studies can focus on charging-related mechanisms for the proposed schemes.

(b) The performance evaluation and analysis of the proposed schemes have been carried out using analytical modelling and simulation. It will be interesting to validate the schemes with a real-life implementation or test-bed, so as to reveal the real-world performance of the schemes.

(c) It will be worthwhile to further improve DM-RMA, in order to support NEMO. This can be achieved by borrowing ideas from ND-RMG, with a few modifications. Also, it will be interesting to compare ND-RMG with N-DMM by means of simulations.

(d) In the design of the route optimization for PMIPv6, the location of the coordinating MAG (CMAG) is externally decided by the network administrator. A further study can be carried out to develop an algorithm that can dynamically ensure that the CMAG is appropriately positioned. Moreover, future research can focus on the validation of the scheme by way of a simulation in bigger scenarios.

(e) The mobile nodes in the proposed distributed mobility schemes configure and use IP addresses obtained from different networks. A mechanism to allow an MN to choose a particular IP address for a specific application may well be investigated in future work.

References

[1] K. Pentikousis, Q. Zhou and H. Wang, "Design considerations for mobility management in future infrastructure networks," in Proc. Telecom World (ITU WT), Technical Symposium at ITU, 2011, pp. 87-92.

[2] L. Bokor, Z. Faigl and S. Imre, "Flat architectures: Towards scalable future internet mobility," Future Internet Assembly, LNCS 6656, 2011,pp.35-50.

[3] C. J. Bernardos, J. C. Zunniga and A. Reznik, "Towards flat and distributed mobility management: A 3GPP evolved network design," in Proc. IEEE ICC, June 10-15, 2012, pp.6855-6861.

[4] Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017. White Paper, February 6, 2013. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html.

[5] P. Bertin, S. Bonjour and J.-M. Bonnin, "Distributed or centralized mobility?" in Proc. IEEE GLOBECOM, 2009, pp.1-6.

[6] V. Chandrasekhar, J. G. Andrews and A. Gatherer, "Femtocell networks: a survey," IEEE Communications Magazine, vol. 46, no. 9,September 2008, pp. 59-67.

[7] J. C. Zuniga, C. J. Bernardos, A. de la Oliva, T. Melia, R. Costa and A. Reznik, "Distributed mobility management: a standards landscape," IEEE Communications Magazine, vol. 51, no.3, March 2013, pp. 80-87.

[8] K. Lee, J. Lee, Y. Yi, I. Rhee and S. Chong, "Mobile Data Offloading: How Much Can WiFi Deliver?" IEEE/ACM Transactions on Networking, vol. 21, no.2, April 2013, pp.536-550.

[9] A. De La Oliva, C. J. Bernardos, M. Calderon, T. Melia and J. C. Zuniga, "IP flow mobility: smart traffic offload for future wireless networks," IEEE Communications Magazine, vol. 49, no.10, October 2011, pp.124-132.

[10] 3GPP TS 23.829, "Local IP access and selected IP traffic offload (LIPA-SIPTO)," V10.0.1, October 2011.

[11] D.-H. Shin, D. Moses, M. Venkatachalam and S. Bagchi, "Distributed mobility management for efficient video delivery over all-IP mobile networks: Competing approaches," IEEE Network, vol. 27, no. 2, April 2013, pp.28-33.

[12] H. A. Chan, H. Yokota, J. Xie, P. Seite and D. Liu, "Distributed and Dynamic Mobility Management in Mobile Internet: Current Approaches and Issues," Journal of

Communications, vol. 6, no. 1, February 2010, pp. 4-15.

[13] 3GPPTS 23.002,"Network architecture," V10.1.1, release 10, January 2011.

[14] H. Chan, D. Liu, P. Seite, H. Yokota and J. Korhonen, "Requirements for distributed mobility management," IETF draft-ietf-dmm-requirements-11 (work in progress), November 2013.

[15] 3GPP. http://www.3gpp.org/.

[16] WiMAX forum. htpp://www.wimaxforum.org/.

[17] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury and B. Patil, "Proxy mobileIPv6," IETF RFC 5213, August 2008.

[18] H. Soliman, editor, "Mobile IPv6 support for dual stack hosts and routers," IETF RFC 5555, June 2009.

[19] The Internet Engineering Task Force (IETF). http://www.ietf.org/.

[20] C. Perkins, D. Jonson and J. Arkko, "Mobility Support in IPV6," IETF RFC 6275, July 2011.

[21] IETF Distributed Mobility Management (DMM) WG. http://datatracker.ietf.org/wg/dmm/.

[22] P. Bertin, S. Bonjour and J.-M. Bonnin, "An evaluation of dynamic mobility anchoring," in Proc. IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall), 2009, pp.1-5.

[23] P. P. Ernest, H. A. Chan and O. E. Falowo, "Distributed mobility management scheme with mobility routing function at the gateways," in Proc. IEEE GLOBECOM, December 2012, pp. 5254-5259.

[24] V. Devarapalli, R. Wakikawa, A. Petrescu and P. Thubert, "Network mobility (NEMO) basic support protocol," IETF RFC 3963, January 2005.

[25] P. P. Ernest, O. E. Falowo, H. A. Chan and L. A. Magagula, "Fast route optimization considering mitigating packet loss for proxy MIPv6 with coordinating MAG" in Proc. Southern Africa Telecommunication Networks and Applications Conference, Stellenbosch, South Africa, September 1-4, 2013.

[26] D. Le, X. Fu and D. Hogrefe, "A review of mobility support paradigms for the internet," IEEE Communications Surveys & Tutorials, vol. 8, no.1, 2006, pp.38-51.

[27] T. R. Henderson, "Host mobility for IP networks: a comparison," IEEE Network, vol. 17,

no.6, 2003, pp.18-26.

[28] Y.W. Chung, M. Chung and D. K. Sung, "Effect of personal mobility management in mobile communication networks," IEEE Transactions on Vehicular Technology, vol. 52, no.5, September 2003, pp.1254-1269.

[29] I. F. Akyildiz, J. Xie and S. Mohanty, "A survey of mobility management in next-generation all-IP-based wireless systems," IEEE Wireless Communications, vol. 11, no.4, August 2004, pp.16-28.

[30] F. M. Chiussi, D. A. Khotimsky and S. Krishnan, "Mobility management in third-generation all-IP networks," IEEE Communications Magazine, vol. 40, no. 9, September 2002, pp.124-135.

[31] P. Bhagwat, C. Perkins and S. K. Tripathi, "Network layer mobility: an architecture and survey," IEEE Personal Communications, vol.3, no.3, June 1996, pp. 54-64.

[32] R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang and V. Paxson, "Stream Control Transmission Protocol," IETF RFC 2960, October 2000.

[33] D. A. Maltz and P. Bhagwat, "MSOCKS: An architecture for transport layer mobility," in Proc. IEEE INFOCOM Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies, 1998, pp.1037-1045.

[34] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley and E. Schooler, "SIP: Session initiation protocol," IETF RFC 3261, June 2002.

[35] R. Ramjee, K. Varadhan, L. Salgarelli, S. R. Thuel, S.-Y. Wang and T. La Porta, "HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks," IEEE/ACM Transactions on Networking, vol. 10, no.3, June 2002, pp.396-410.

[36] A. T. Campbell, J. Gomez, S. Kim and A. G. Valk'o, "Design, implementation and evaluation of celluar IP," IEEE Personal Communications Magazine, vol.7, August 2000, pp.42-49.

[37] R. Moskowitz, P. Nikander, P. Jokela and T. Henderson, "Host Identity Protocol ," IETF RFC 5201, April 2008.

[38] R. Koodli, editor, "Mobile IPv6 fast handovers," IETF RFC 5568, July 2009.

[39] H. Soliman, C. Castelluccia, K. ElMalki and L. Bellier, "Hierarchical mobile IPv6 (HMIPv6) mobility management," IETF RFC 5380, October 2008.

[40] J. Arkko, C. Vogt and W. Haddad, "Enhanced Route Optimization for Mobile IPv6,"

IETF RFC 4866, May 2007.

[41] R. Koodli, "IP address location privacy and mobile IPv6: Problem statement," IETF RFC 4882, May 2007.

[42] C. M. Mueller and O. Blume, "Network-based mobility with proxy mobile IPv6," in Proc. IEEE PIMRC, September 2007, pp. 1-5.

[43] L.A. Magagula, O. E. Falowo and H. A. Chan, "Enhancing PMIPv6 for Better Handover performance among Heterogeneous Wireless Networks in a Micromobility Domain," EURASIP Journal on Wireless Communications and Networking, 2010, 2010:274935 doi:10.1155/2010/274935.

[44] M. Liebsch, S. Jeong and Q. Wu, "Proxy mobile IPv6 (PMIPv6) localized routing problem statement," IETF RFC 6279, June 2011.

[45] Network-based mobility extensions (netext). http://datatracker.ietf.org/wg/netext/.

[46] J.-H. Lee and T.-M. Chung, "How much do we gain by introducing route optimization in Proxy Mobile IPv6 networks?" Annals of Telecommunications, vol. 65, no.5-6, June 2010, pp.233-246.

[47] S. Krishnan, R. Koodli, P. Loureiro, Q. Wu and A. Dutta, "Localized routing for proxy mobile IPv6," IETF RFC 6705, September 2012.

[48] J. Guan, I. You, C. Xu, H. Zhou and H. Zhang, "Survey on route optimization schemes for proxy mobile IPv6," in Proc. Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), July 2012, pp.541-546.

[49] H. A. Chan, "Proxy mobile IP with distributed mobility anchors," in Proc. IEEE GLOBECOM Workshops (GC Wkshps), 2010, pp. 16-20.

[50] H. A. Chan, "Distributed mobility management with mobile IP," in Proc. IEEE ICC, June 2012, pp. 6850-6854.

[51] P. P. Ernest, O. E. Falowo and H. A. Chan, "Network-based distributed mobility management: Design and analysis," in Proc. 9th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Lyon, France, October 7-9, 2013, pp. 516-523.

[52] P. Bertin, S. Bonjour and J.-M. Bonnin, "A distributed dynamic mobility management scheme designed for flat IP architectures," in Proc. 3rd International Conference on New Technologies, Mobility and Security(NTMS),November 2008, pp. 1-5.

[53] R. Wakikawa, G. Valadon and J. Murai, "Migrating Home Agents Towards Internet-Scale Mobility Deployments," in Proc. ACM Conference on Future Networking

Technologies (CoNEXT'06), December 2006.

[54] L. Yu, Z. Zhijun, L. Tao and T. Hui, "Distributed mobility management based on flat network architecture," in Proc. 5th Annual ICST, Wireless Internet Conference (WICON), March 2010, pp. 1-6.

[55] F. Giust, A. de la Oliva and C. J. Bernardos, "Flat access and mobility architecture: An IPv6 distributed client mobility management solution," in Proc. IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), April 2011, pp. 361-366.

[56] J. H. Lee, J. Bonnin and X. Lagrange, "Host-based distributed mobility management support protocol for IPv6 mobile networks," in Proc. 8th IEEE International Conference on Wireless and Mobile Computing, Networking and Communication (WiMob), Barcelona, Spain, October 2012, pp. 61-68.

[57] F. Giust, A. De La Oliva, C. J. Bernardos and R. P. F. Da Costa, "A network-based localized mobility solution for distributed mobility management," in Proc. 14th International Symposium on Wireless Personal Multimedia Communications (WPMC), October 2011, pp. 1-5.

[58] J.-I. Kim, H. Jung and S. J. Koh, "Distributed mobility control for mobile-oriented future internet environments," in Proc. International Conference on ICT Convergence (ICTC), September 2011, pp. 342-347.

[59] S. Yan, C. Jiayin and C. Shanzhi, "A Mobile IPv6 based Distributed Mobility Management Mechanism of Mobile Internet," Physics Procedia, vol. 25, 2012, pp. 2249-2256.

[60] A. Nascimento, R. Sofia, T. Condeixa and S. Sargento, "A decoupling approach for distributed mobility management," in Proc. 21st International Conference on Computer Communications and Networks (ICCCN), 2012, pp. 1-6.

[61] M. Fischer, F.-U. Andersen, A. Kopsel, G. Schafer and M. Schlager, "A distributed IP mobility approach for 3G SAE," in Proc. IEEE PIMRC, September 2008, pp. 1-6.

[62] H. Ali-Ahmad, M. Ouzzif, P. Bertin and X. Lagrange, "Distributed dynamic mobile IPv6: Design and evaluation," in Proc. IEEE WCNC, April 7-10, 2013, pp. 2166-2171.

[63] L. Yi, H. Zhou, D. Huang and H. Zhang, "D-PMIPv6: A distributed mobility management scheme supported by data and control plane separation," Mathematical and Computer Modelling, vol. 58, no. 5-6, September 2013, pp. 1415-1426.

[64] R. Costa, T. Melia, D. Munaretto and M. Zorzi, "When mobile networks meet content delivery networks: Challenges and opportunities," in Proc. Seventh ACM International Workshop on Mobility in the Evolving Internet Architecture, 2012, pp. 11-16.

[65] S. Thomson, T. Narten and T. Jinmei, "IPv6 stateless autoconfiguration," IETF RFC 4862, September 2007.

[66] NS-2 Network Simulator. Available: http://www.isi.edu/nsnam/ns.

[67] PMIPv6 for NS-2 patch and required packages. Available: http://commani.net/pmip6ns/download.html.

[68] N. Cranley and M. Davis, "Video frame differentiation for streamed multimedia over heavily loaded IEEE 802.11e WLAN using TXOP," in Proc. IEEE PIMRC, September 2007, pp. 1-5.

[69] T. Ernst and H. Lach, "Network mobility support terminology," IETF RFC 4885, IETF RFC 4885, July 2007.

[70] O. Troan and R. Droms, "IPv6 prefix options for dynamic host configuration protocol (DHCP) version 6," IETF RFC 3633, December 2003.

[71] H.-J. Lim, M. Kim, J.-H. Lee and T. M. Chung, "Route Optimization in Nested NEMO: Classification, Evaluation, and Analysis from NEMO Fringe Stub Perspective," IEEE Transactions on Mobile Computing, vol. 8, November 2009, no.11, pp. 1554-1572.

[72] P. Sornlertlamvanich, S. Kamolphiwong, R. Elz and P. Pongpaibool, "NEMO-based distributed mobility management," in Proc. 26th International Conference on Advanced Information Networking and Applications Workshops (WAINA), March 26-29,2012, pp. 645-650.

[73] T.-X. Do and Y. Kim, "Distributed network mobility management," in Proc. International Conference on Advanced Technologies for Communications (ATC), October 10-12, 2012, pp. 319-322.

[74] R. Droms et al., "Dynamic host configuration protocol for IPv6 (DHCPv6)," IETF RFC 3315, July 2003.

[75] M. Chuang and J. Lee, "DRO: domain-based route optimization scheme for nested mobile networks," EURASIP Journal on Wireless Communications and Networking, pp. 1-19, 2011, 2011:70 doi: 10.1186/1687-1499-2011-70.

[76] J. Lei and X. Fu, "Evaluating the benefits of introducing PMIPv6 for localized mobility management," in Proc. Wireless Communications and Mobile Computing Conference (IWCMC), August 2008, pp. 74-80.

[77] T. T. Nguyen and C. Bonnet, "DMM-based inter-domain mobility support for proxy mobile IPv6," in Proc. IEEE WCNC, Shangai, China, April 2013, pp. 1998-2003.

[78] M. Crawford and B. Haberman, "IPv6 node information queries," IETF RFC 4620,

August 2006.

[79] J.-H. Lee, S. Gundavelli and T.-M. Chung, "A performance analysis on route optimization for proxy mobile IPv6," in Proc. IEEE ICC, June 2009, pp.1-6.

[80] J.-H Lee, Y.-H. Han, S. Gundavelli and T.-M. Chung, "A comparative performance analysis on Hierarchical Mobile IPv6, and Proxy Mobile IPv6," Telecommunication Systems, vol. 41, May 2009, pp.279-292.

[81] I. F. Akyildiz and W. Wang, "A dynamic location management scheme for next-generation multitier PCS systems," IEEE Transactions on Wireless Communications, vol. 1, no.1, January 2002, pp. 178-189.

[82] C. Makaya and S. Pierre, "An Analytical Framework for Performance Evaluation of IPv6-Based mobility Management Protocols," IEEE Transactions on Wireless Communications, vol. 7, no.3, March 2008, pp. 972-983.

[83] J. Xie and I. F. Akyildiz, "A novel distributed dynamic location management scheme for minimizing signalling costs in Mobile IP," IEEE Transactions on Mobile Computing, vol. 1, no.3, 2002, pp. 163-175.

[84] J.-H. Lee, T. Ernst and T.-M. Chung, "Cost analysis of IP mobility management protocols for consumer mobile devices," IEEE Transactions on Consumer Electronics, vol. 56, no.2, 2010, pp. 1010-1017.

[85] H.-N. Nguyen and C. Bonnet, "Proxy mobile IPv6 for cluster based heterogeneous wireless mesh networks," in Proc. 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems(MASS), 2008, pp. 617-622.

[86] H. Jung, M. Gohar, J.I. Kim and S.J. Koh, "Distributed Mobility Control in Proxy Mobile IPv6 Networks," IEICE Transactions on Communications, vol. E94-B, no. 8, August 2011, pp. 2216-2224.

[87] Q. Wu and B. Sarikaya, "An extension to proxy mobile IPv6 for local routing optimization," IETF draft-wu-netext-local-ro-05 , February 2010.

[88] J. W. Park, J. I. Kim and S. J. Koh, "Q-PMIP: Query-based proxy mobile IPv6," in Proc. 13th International Conference on Advanced Communication Technology (ICACT), February 2011, pp. 742-745.