Quality of Service (QoS) Provisioning Mechanisms in Fourth Generation (4G) Wireless All-IP Networks

Ahmad Belhoul

BEng in Telecommunications Engineering
MSc in Telecommunications Engineering

A thesis submitted to
The Department of Electrical and Computer Systems Engineering
Monash University, Australia
for the degree of Doctor of Philosophy

December 2007

Abstract

This thesis presents the details of two reservation models created for improved resource allocation in wireless All-IP networks in order to meet the Quality of Service (QoS) requirements of real-time applications whilst maintaining resource utilisation at high levels. The two mechanisms are proposed as extensions to the Resource Reservation Protocol (RSVP). The first was designed to exploit the future compatibility provisions built into the RSVP standard architecture and can therefore be easily installed at end systems without affecting the operation of unmodified nodes in a network. The second model provides better overall performance at the expense of a higher level of complexity, and requires changes to be made to all RSVP-capable nodes in a network.

Wireless networks are rapidly evolving into an All-IP, or Fourth Generation (4G) architecture, and are expected to deliver real-time services such as Voice-over-IP (VoIP) and Video-over-IP (VIP) seamlessly and efficiently even for mobile users. These applications impose strict QoS constraints on timely delivery of packets and packet loss. QoS guarantees for such applications require additional network resource control mechanisms to be added to the existing TCP/IP protocol stack. Firstly, a mechanism is needed to replicate the channel characteristics of Public Switched Telephone Networks (PSTNs). This is achieved using RSVP, the industry's de facto standard for wired network resource control. RSVP explicitly reserves network resources to ensure a low and fixed amount of delay with effectively no loss. Secondly, another mechanism is required to allow a node to move freely across different wireless subnets whilst maintaining its connectivity. Mobile IPv6 (MIPv6) is the standard developed by the Internet Engineering Task Force (IETF) to facilitate such seamless mobility in wireless IPv6 networks.

However, RSVP was designed for end-systems whose IP addresses do not change. Once mobility of an end-system is allowed, the dynamically changing MIPv6 address inevitably impacts on RSVP performance. The first part of this thesis aims to quantify the significance of this impact using a framework consisting of a simulation model to assess application-level performance, and a signaling cost model to measure the network-level performance. The objective of this effort is twofold: To highlight the critical issues involved in such an interaction, and to serve as a performance benchmark in the design process of a more efficient QoS scheme.

The second part of this thesis proposes the Mobility Aware Resource Reservation Protocol (MARSVP) in which mobility and QoS signaling are performed

as a single functional block. The key concept of MARSVP is to convey mobility-specific information (binding updates and their associated acknowledgments) by using newly defined RSVP objects embedded in existing RSVP messages. An appealing feature of MARSVP is that it adheres to the current RSVP standard (RFC 2205) and thus requires minimal changes at end nodes without affecting any of the conventional RSVP routers in between. MARSVP addresses several of RSVP's deficiencies by reducing the QoS re-establishment time from two Round Trip Times (RTT) to 1.5 RTT. The results of the simulation-based experiments confirm that the proposed MARSVP mechanism provides superior application-level performance during handoffs than standard RSVP. Moreover, its use reduces the network-level signaling costs accordingly.

The third part of this thesis proposes a new packet classification mechanism for RSVP (called RSVP-HoA) in which routers are configured to classify flows based on the home address option in the MIPv6 destination options header. Through this approach, intermediate RSVP routers are able to correctly identify an RSVP flow, even after a Mobile Node (MN) changes its Care-of-Address. Moreover, a crossover router (COR) using this mechanism can detect the changed portion of the end-to-end RSVP session and confine RSVP signaling to the changed nodes. As a result, the RSVP re-establishment time and network signaling costs drop substantially. Another key advantage of the proposed mechanism is that it overcomes the dual reservations issue confronted when using standard RSVP packet classification for a roaming MN.

Depending on a service provider's desired level of complexity, any of the two outlined QoS mechanisms presented in this thesis report could be implemented as extensions to the RSVP to improve the level of performance in wireless networks: While MARSVP provides a simple and efficient alternative, RSVP-HoA delivers superior application-level performance to the end user while at the same time imposing fewer signaling costs on the network. This however, is achieved at the expense of a higher level of complexity since changes are required to be made to all RSVP-capable nodes in the network.

Acknowledgments

I would like to express my deep and sincere gratitude to my supervisors, Dr. Ahmet Şekercioğlu and Dr. Nallasamy Mani for their continuous guidance and valuable comments. I would also like to thank *Emirates Telecommunications Corporation* (ETISALAT) and the *Australian Telecommunications Cooperative Research Centre* (ATcrc) for their financial support. My thanks extend to reach my colleagues at the *Centre for Telecommunications and Information Engineering* (CTIE) for their fruitful discussions, constructive arguments and uplifting coffee breaks.

I am particularly indebted to my family and friends. It would be difficult to overstate my gratitude towards my wife and my little son, on whose constant encouragement and love I relied on. This research would not have been possible without my wife's understanding, support and sacrifices she made while putting up with me as I worked on this thesis, both at home and abroad. I am also grateful to my late brother, Mohammed, who passed away before the completion of this thesis. His encouraging words during periods of desperation made me see the light at the end of the PhD journey.

Lastly and most importantly, I wish to thank my parents. They raised me, supported me, taught me, and loved me. To them I'm eternally grateful.

To my late brother, Mohammed Belhoul

Declaration

I declare that, to the best of my knowledge, the research described herein is original except where the work of others is indicated and acknowledged, and that the thesis has not, in whole or in part, been submitted for any other degree at this or any other university.

Ahmad Belhoul
Melbourne
December 2007

Contents

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Project
ARSVP	Adaptive Resource Reservation Protocol
AMPS	Analogue Mobile Phone Service
AP	Access Point
AR	Access Router
BAck	Binding Acknowledgment
BU	Binding Update
CBR	Constant Bit Rate
CDMA	Code Division Multiple Access
CN	Correspondent Node
CoA	Care-of-Address
COR	Crossover Router
DAD	Duplicate Address Detection
DAT	Dynamic Address Translation
DHCP	Dynamic Host Configuration Protocol
DiffServ	Differentiated Services
DSCP	DiffServ Code Point
EDGE	Enhanced Data rates for GSM Evolution
EV-DO	Evolution-Data Optimized
FA	Foreign Agent
FMIP	Fast Handover for Mobile Internet Protocol
FN	Foreign Network
GOP	Group Of Pictures
GPRS	General Packet Radio Services
GSM	Global System for Mobile Telecommunications
HA	Home Agent
HAck	Handover Acknowledgment
HI	Handover Initiation
HMIP	Hierarchical Mobile Internet Protocol
HN	Home Network
HoA	Home Address
HSPA	High-Speed Packet Access
ICMP	Internet Control Management Protocol
IETF	Internet Engineering Task Force

IntServ	Integrated Services
IP	Internet Protocol
ITU	International Telecommunications Union
LCoA	Local Care-of-Address
MA	Mobility Agent
MAP	Mobility Anchor Point
MBWA	Mobile Broadband Wireless Access
MIP	Mobile Internet Protocol
MN	Mobile Node
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MRSVP	Mobile Resource Reservation Protocol
MSE	Mean Square Error
nAR	Next Access Router
nCoA	Next Care-of-Address
NHOP	Next Hop
oAR	Old Access Router
PHB	Per-Hop Behaviour
PHOP	Previous Hop
PrRtAdv	Proxy Router Advertisement
PSTN	Public Switched Telephone Network
PSNR	Peak Signal-to-Noise Ratio
QCIF	Quarter Common Intermediate Format
RCoA	Regional Care-of-Address
RSVP	Resource Reservation Protocol
RSVP-MP	Resource Reservation Protocol Mobility Proxy
RtSolPr	Router Solicitation for Proxy
RTP	Real-Time Transport Protocol
RTT	Round Trip Time
RFC	Request for Comment
SIP	Session Initiation Protocol
SMRP	Sender-initiated and Mobility-support Reservation Protocol
SNR	Signal-to-Noise Ratio
TACS	Total Access Communications System
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UTP	User Datagram Protocol
VBR	Variable Bit Rate
VoIP	Voice over Internet Protocol
VIP	Video over Internet Protocol
WCDMA	Wide-band Code Division Multiple Access
WiMAX	World Interoperability for Microwave Access
WFQ	Weighted Fair Queuing

WLAN	Wireless Local Area Network
WLRP	Wireless Lightweight Reservation Protocol
QCIF	Quarter Common Intermediate Format
QoS	Quality of Service

List of Symbols

Ψ^{FMIP}	Fast Handovers for Mobile IPv6 signaling cost
$\Psi^{\text{FMIP}}_{\text{MARSVP}}$	Total signaling cost of deploying MARSVP over FMIPv6
$\Psi^{\text{FMIP}}_{\text{RSVP}}$	Total signaling cost of deploying RSVP over FMIPv6
$\Psi^{\text{FMIP}}_{\text{RSVP-HoA}}$	Total signaling cost of deploying RSVP-HoA over FMIPv6
Ψ^{FPT}	Packet Tunneling costs of the Fast Handovers mechanism
Ψ^{FR}	Registration costs of the Fast Handovers mechanism
Ψ^{HMIP}	Hierarchical Mobile IPv6 signaling cost
$\Psi^{\text{HMIP}}_{\text{MARSVP}}$	Total signaling cost of deploying MARSVP over HMIPv6
$\Psi^{\text{HMIP}}_{\text{RSVP}}$	Total signaling cost of deploying RSVP over HMIPv6
$\Psi^{\text{HMIP}}_{\text{RSVP-HoA}}$	Total signaling cost of deploying RSVP-HoA over HMIPv6
Ψ^{MIP}	Mobile IPv6 signaling cost
$\Psi^{\text{MIP}}_{\text{MARSVP}}$	Total signaling cost of deploying MARSVP over MIPv6
$\Psi^{\text{MIP}}_{\text{RSVP}}$	Total signaling cost of deploying RSVP over MIPv6
$\Psi^{\text{MIP}}_{\text{RSVP-HoA}}$	Total signaling cost of deploying RSVP-HoA over MIPv6
Ψ^{RSVP}	Total Resource Reservation Protocol signaling cost
I_d	Impairment factor for equipment delay
β	Buffering cost at nAR
γ	Processing cost at a node
δ_B	Per-hop binding update transmission cost
δ_D	Per-hop data packet transmission cost
δ_R	Per-hop RSVP message transmission cost
λ_{Adr}	Address Resolution Delay
$\lambda_{Handoff}$	Handoff Delay
λ_{L2}	Layer-2 Handoff Delay
λ_{L3}	Layer-3 Handoff Delay
λ_{RSVP}	RSVP Signaling Delay
A	Advantage Factor
I_e	Impairment factor for equipment loss
I_s	Impairment factor for signal-to-noise
l_{mc}	Average distance between the MN and CN in hops
l_{mh}	Average distance between the MN and HA in hops
l_{mm}	Average distance between the MN and oAR in hops

l_{mn}	Average distance between the MN and nAR in hops
l_{mo}	Average distance between the MN and oAR in hops
l_{on}	Average distance between the oAR and nAR in hops
M	Average number of global handoffs
MSE	Mean Square Error
N	Total number of mobile nodes
P	Average number of packets dropped during a handoff
P_{mc}	Processing costs between the MN and CN
P_{mh}	Processing costs between the MN and HA
P_{mm}	Processing costs between the MN and MAP
P_{mn}	Processing costs between the MN and nAR
P_{mo}	Processing costs between the MN and oAR
P_{on}	Processing costs between the oAR and nAR
R	R-factor
R_{mc}	Registration costs between the MN and CN
R_{mh}	Registration costs between the MN and HA
R_{mm}	Registration costs between the MN and MAP
R_{mn}	Registration costs between the MN and nAR
R_{mo}	Registration costs between the MN and oAR
R_{on}	Registration costs between the oAR and nAR
R_o	Signal-to-noise ratio
Rsv_{mc}	Resource reservation costs between MN and CN
TI_{QoS}	Total Interruption in Quality of Service
T_{mc}	Transmission costs between the MN and CN
T_{mh}	Transmission costs between the MN and HA
T_{mm}	Transmission costs between the MN and MAP
T_{mn}	Transmission costs between the MN and nAR
T_{mo}	Transmission costs between the MN and oAR
T_{on}	Transmission costs between the oAR and nAR
tr	MN residence time in a subnet
V_{Peak}	Maximum possible signal energy

Chapter 1

Introduction

1.1 Background and Motivation

Mobile phone usage began with the launch of the First Generation (1G) mobile technology in the early 1980s [FAG95]. The main design objective of 1G mobile networks was to provide a wireless architecture that allows subscribers to place mobile calls and maintain connectivity as they moved from one coverage area (cell) to another. 1G standards include the Analogue Mobile Phone Service (AMPS) in the United States, Total Access Communications System (TACS) in the United Kingdom, in addition to the C-450 in West Germany, Portugal and South Africa. The second generation (2G) on the other hand, witnessed the introduction of digital mobile communications [ZAB99]. Digitised voice signals have the advantage of being compressed and multiplexed much more efficiently than analog signals, thus resulting in a significant increase in link capacity utilisation. Although this was the main driving force behind the development of 2G, an all-digital system also has the capability of delivering some data services such as the Short Messaging System (SMS) and email. Another key advantage is that digital mobile calls are much more less to eavesdropping [SMMA06]; thereby making 2G phone communications immensely more private than their predecessors. 2G standards include GSM (Global System

for Mobile telecommunications) [GSMb], originally from Europe but used worldwide and CDMA1 (Code Division Multiple Access) [CDM] used in the Americas and parts of Asia.

With the massive success of 2G technology, the number of mobile subscribers increased from 214 million in 1997 to 1.162 billion in 2002 [ITU02]. It is predicted that this figure will continue to grow, reaching 1.7 billion by 2010 [KJC+03]. This indicates that voice-oriented mobile technology is approaching its saturation point (as depicted in Figure 1.1 [IR02]). With this notion in mind, service providers started exploring ways of creating new demand by introducing new (and more bandwidth hungry) services, including faster Internet access and the Multimedia Messaging Service (MMS) [Ope]. To explore the pickup rate of these new services and demand potential with minimal financial risks, intermediate generations were introduced as add-ons to the existing 2G infrastructure to facilitate packet-switched connections (thereby improving data transfer capability). A popular 2.5G standard is GPRS (General Packet Radio Services) offering an average data rate of 40 kbps [GSMa]. Another standard is the 2.75G Enhanced Data rates for GSM Evolution (EDGE) [Glo] capable of delivering data at a theoretical maximum rate of 384 kbps, although actual data rates average at around 100 kbps.

While 2.5G and 2.75G offered basic data services, the most significant feature offered by the third generation (3G) mobile technology is its broadband capabilities to support the increasing demand for high data rates. Marketing of 3G services often focused on video telephony, although upon roll-out, music downloading and video streaming proved to be the most popular. The two dominant 3G standards are CDMA2000 in the United States, and Wide-band CDMA (WCDMA) [UMT] in Europe and Asia.

Despite its promising potential, 3G adoption has been largely underwhelming; part of the reason is the separate standards maintained for 3G such as the 3rd Generation Partnership Project (3GPP) [Thea] and 3GPP2 [Theb]. The initial speculation was that 3G would serve as a universal standard, although from an economic point

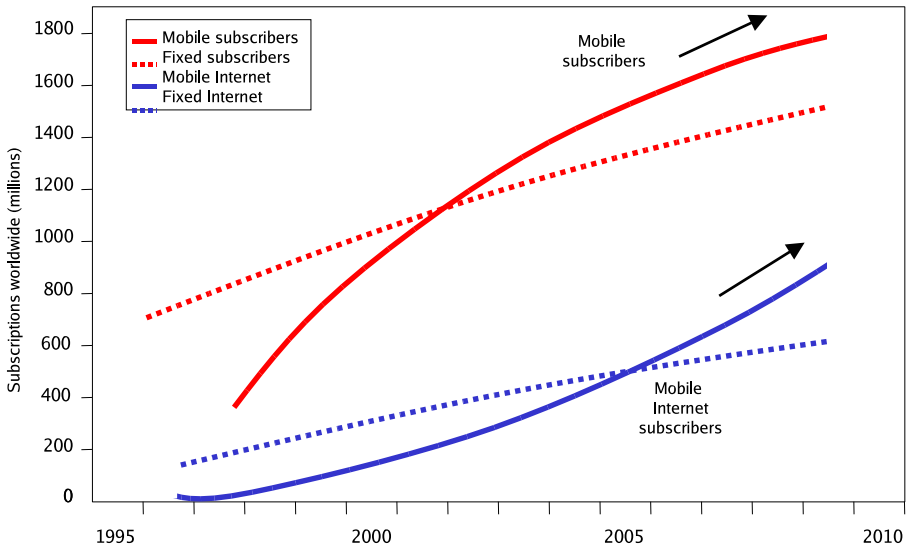

Figure 1.1: Evolution of Mobile and Fixed Subscribers [IR02].

of view, it was far more cost-effective for service providers to make 3G networks backwards compatible with their existing 2G infrastructure. This propagated the incompatibility of the competing 2G standards into 3G, and hence 3GPP was based on GSM, while 3GPP2 on CDMA1.

Another drawback is the financial costs of launching 3G systems, which include the billions of dollars invested in acquiring 3G licenses alone. As a result, the reluctant Telecom industry slowed down the roll-out of 3G networks (confining coverage to metropolitan areas) and relied on the intermediate 2.5G and 2.75G technologies to meet demand for greater bandwidth.

Even though the speculated 3G maximum data speeds of 2 Mbps were theoretically sound, they where not practically attainable in real life implementations. As a result, 3G subscribers had to temporarily settle for speeds below 384 kbps [Dur01], which made them start to question 3G network's ability to deliver genuine broadband capability. As demand for mobile broadband services increased and competition matured, another intermediate generation (3.5G) was launched. With 3.5G, users are able to reach download speeds ranging from 800 kbps to 2 Mbps depend-

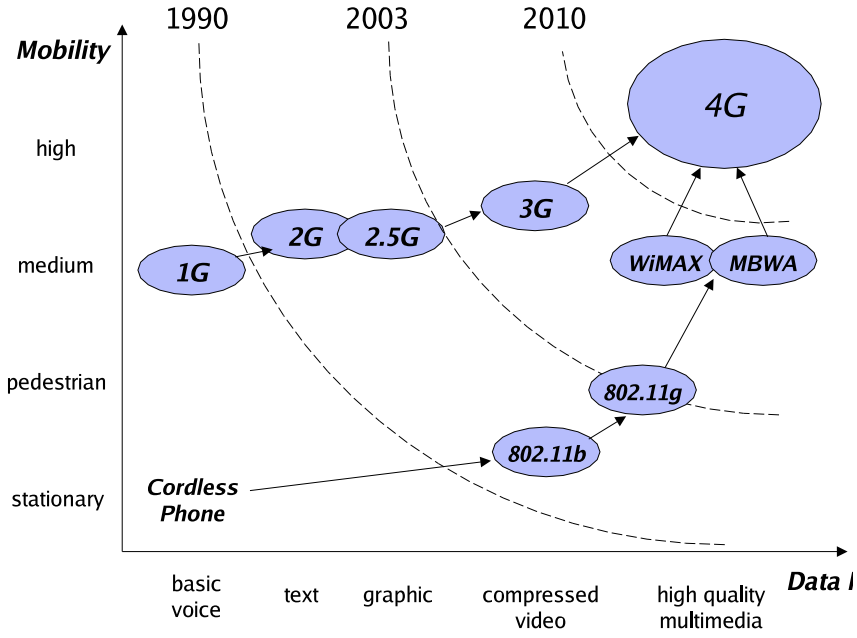

Figure 1.2: Evolution of Wireless Communications.

ing on location and network configuration. The two dominant 3.5G standards include the Evolution-Data Optimized (EV-DO) [3GP], based on CDMA2000 in the United States; and High-Speed Packet Access (HSPA) [GSMc] based on WCDMA in Europe and Asia.

While 3G and 3.5G mobile technologies promise to support mobile broadband access, they will soon find themselves competing with emerging wireless access technologies. Competition first appeared in the form of 802.11b wireless LAN (WLAN) [IEEc] hotspots, which provide Mbps speeds in public areas (albeit within a limited coverage area). Hotspot deployment has been expanding ever since, mainly due to the cost savings involved: WLAN access points cost a fraction of the mobile infrastructure equivalent since there are no licensing fees involved in deploying WLAN networks. Moreover, equipment costs are significantly lower due to the highly competitive nature of hardware manufacturing in the computer world (in contrast to highly monopolised telecommunications hardware manufacturing). Early applications already started to appear in the marketplace such as dual mode

mobile phones that could connect to hotspots for access to cheaper broadband services. Nonetheless, the 802.11 group of technologies do not pose a direct threat to 3G mobile networks due to their limited coverage and lack of seamless mobility (yet).

However, newly proposed wireless access technologies such as the IEEE 802.16 World Interoperability for Microwave Access (WiMAX) [IEEa] and the IEEE 802.20 Mobile Broadband Wireless Access (MBWA) [IEEb] have the potential to create a big impact on the mobile broadband market. WiMAX and MBWA not only promise continuous and ubiquitous metropolitan network coverage, but also offer high data rates surpassing those of 3G mobile networks. Not surprisingly, these advanced wireless access systems received an unwelcome response from giant 3G manufacturers[1]. This however has not hindered their development. WiMAX research, for example, is actively supported by chipset manufacturing giant Intel[2]. Moreover, the price paid per Hz for WiMAX and MBWA spectrum is approximately a thousand fold cheaper than that for 3G spectrum [Mar07]. This resulted in a high number of licensees with a total of 721 licenses worldwide for WiMAX/MBWA, compared with 106 for 3G [Mar07]. Endless debates have spurred concerning the future of mobile networks. A comprehensive 52-page report [Dat05] regarding this issue reached the following conclusions:

1. With more than 2 billion wireless subscribers globally, no single mobile technology can satisfy all market needs. A mixture of diverse networks is needed to provide optimal service both indoors and outdoors.

2. The report shows the dramatic decline in subscribers that 3G mobile networks can support as users adopt broadband services. The only way to support more broadband users is to employ more densely deployed wireless networks whether WiMAX or MBWA.

[1] In July 2005, EU frequency allocation for WiMAX was blocked by France and Finland, where manufacturers have invested heavily in 3G technology [Cia05].

[2] Intel's WiMAX chipset, Code-named Rosedale 2, is currently shipping in sample quantities to allow equipment manufacturers to develop their products [S. 06].

3. In mobile markets, WiMAX/MBWA operators must employ low-cost, high-density base station architectures to deliver superior capacity and in-building penetration. A feasible solution is to design and integrate WiMAX/MBWA into existing 2.5 and 3G base stations.

The conclusions presented in [Dat05] (and various other recommendations suggested by the research community) lay the foundations of an IP-based Fourth Generation (4G) wireless network solution in which all technologies are integrated through a single IPv6-based core [JTKK05]. An all-IP 4G network has inherent advantages over its precursors, the principal advantage being that IP supports transparency above the radio access technology (i.e. IP can operate over different underlying network platforms such as UMTS and WiMAX). This eliminates the close coupling between the core networking protocol and the link layer (radio) protocol, thereby allowing a high degree of flexibility in selecting the access network. Moreover, both layers could be developed independently of each other. Another advantage is that an open system IP network offers a high level of equipment interoperability, resulting in significant cost savings for service providers since they would not be constrained to a single vendor for the entire network system. Furthermore, because an all-IP core layer is easily scalable, it is well suited to meet the increasing demand for rich wireless multimedia content. Finally, using an IP-based 4G wireless solution, network providers can offload broadband data services from the valuable mobile network spectrum on to the less expensive WiMAX/MBWA access networks. According to [Tac03], the cost per bit should be reduced to between 10% to 1% of current 3G systems costs.

1.2 Basic Concepts

Voice and data networks serve different purposes, and hence differ in their design and application. Traditional voice networks are comprised of circuit-switched connections over Public Switched Telephone Networks (PSTN). Data networks, how-

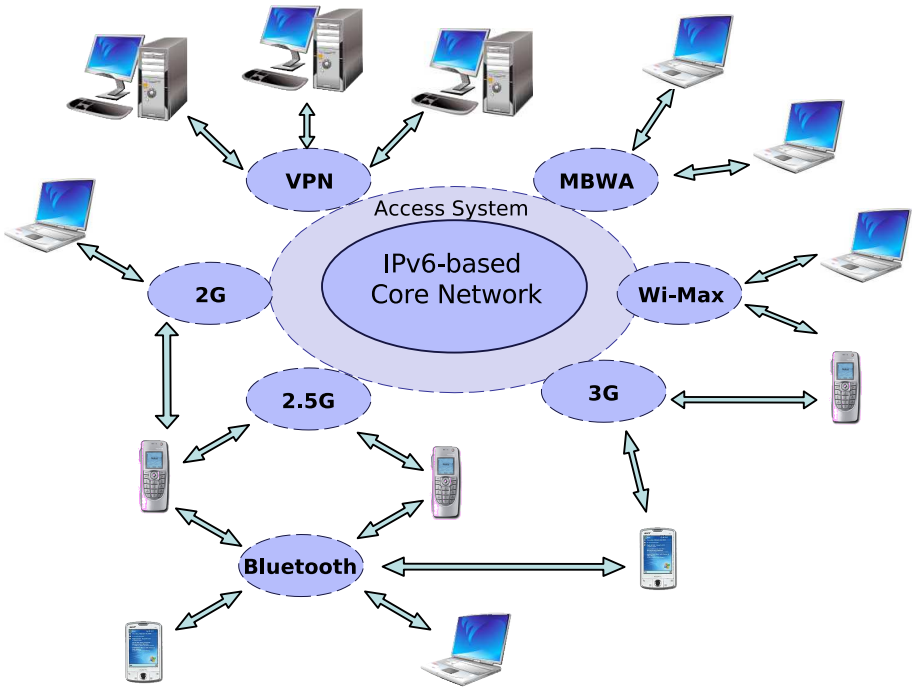

Figure 1.3: A Fourth Generation All-IP Network.

ever, rely on packet-switched datagrams over the Internet. Two different modes of transmission are used in packet-switched networks:

- Connection-oriented (virtual circuit networks): Developed to emulate circuit-switched functionality over the packet-switched Internet. It necessitates the establishment of a session between the sender and receiver. A virtual circuit network guarantees correct packet sequencing (packets routed along the same path with minimal delay variation), although it suffers from greater overhead than a connectionless one.

- Connectionless (datagram networks): Does not require session initiation, packets are delivered independently to the receiver (each may use different paths). A datagram network provides minimal services with neither guarantees to packet delivery nor correct packet sequence.

Virtual circuit networks are not widely used nowadays. The Internet's basic architecture is built on connectionless mode of transmission. Connection-oriented transport layer protocols that operate over connectionless mode of transmission such as the Transmission Control Protocol (TCP) was originally designed to reliably deliver a packet with little emphasis on the amount of time it takes to reach its destination. This type of "guaranteed delivery" transmission is adequate for delay-tolerant applications such as web browsing or file downloading. Today the convergence of voice, video and streaming multimedia results in highly diverse traffic; with each traffic type requiring different levels of bandwidth allocation, delay and tolerable packet loss. The guaranteed delivery service provided by TCP is not suitable for such time-sensitive applications (since packet retransmission goes against the nature of real-time services). Voice applications, for example, which originate from PSTN, exhibit a very deterministic behaviour. In PSTN, voice traffic experiences a low and fixed amount of delay with effectively no loss. However, when voice is transported over an IP network, a variable and unpredictable amount of delay is introduced with some voice packets being dropped during congestion periods. As a result, the traditional IP network does not provide the behaviour that the voice application requires. To address the issue, the Quality-of-Service (QoS) concept was developed with the following design objectives:

- Support dedicated bandwidth;
- Set traffic priorities across the network;
- Improve loss and delay characteristics;
- Maximise network utilisation.

By deploying QoS provisioning mechanisms in an IP network, an application could specify a set of parameters essential to guarantee its application level performance. These QoS parameters may range from bandwidth and packet loss to delay and jitter, depending on the specific traffic characteristics. The QoS mechanism would

then manage network resources by setting certain routing priorities and traffic shaping in order to ensure that the services are delivered in an acceptable form to the end user.

In the Internet research community, two main schools of thought have been developed with regards to QoS provisioning: Differentiated Services (DiffServ) [BBC+98] and the Integrated Services (IntServ) [BCS94]. The DiffServ model follows a flexible approach in classifying its datagrams. Packets are marked individually by setting specific DiffServ Code Point (DSCP, an 8-bit field in the IP packet header) values and are forwarded in a hop-by-hop (connectionless) manner by routers in the network. The way packets are forwarded by the routers is referred to as Per-Hop Behaviour (PHB). Since no end-to-end sessions are set up. DiffServ provides a high level of scalability and is therefore well suited to manage resources in core IP networks.

IntServ, on the other hand, abides by a thorough classification system: Network resources are explicitly identified and reserved, and datagrams are treated in a per-flow manner. Each particular data flow is assigned specific QoS parameters known as FlowSpec; imposing explicit reservations on the end-to-end path which works in a similar fashion to conventional circuit-switched networks.

Resource Reservation Protocol (RSVP) [BZB+97] is a robust signaling protocol developed to operate within the IntServ model. It defines how applications place reservations and how they can relinquish the reserved resources once an RSVP session is terminated. RSVP operation generally results in resources being reserved in each node along the end-to-end path, although it can function through non-RSVP routers along the way.

While QoS mechanisms ensure application level performance, IP mobility refers to mechanisms that allow a node to move freely across different subnets whilst maintaining IP connectivity. Mobile IP (MIP) [Per02] is the current standardised mobility protocol by the Internet Engineering Task Force (IETF) to facilitate mobility of end nodes mainly in a wireless environment. The limitations of traditional

IP addressing and routing are the main driving force behind developing MIP. In an IP network each host maintains at least one unique IP address. An IP address is essentially comprised of two parts: A network prefix and a host suffix. This IP address is not only used to uniquely identify the host, but also to find a path to this host from hosts in other subnets[3] across an IP network.

The close coupling of the IP address a host uses means that if it changes subnets due to mobility, its current IP address should be considered invalid as it would not reflect its new location at the newly connected subnet. MIP addresses this issue by allowing a host, called Mobile Node (MN) in MIP parlance, to have two IP addresses: a permanent Home Address (HoA) and a Care-of-Address (CoA), which is associated with the foreign subnet. Using MIP, nodes may move across different subnets whilst maintaining connectivity and transparency at the application level. With introduction of IPv6 networks, a new MIP standard named Mobile IPv6 (MIPv6) [JP04] was proposed accordingly. Since its standardisation, various deficiencies have been addressed and proposed as extensions. Among these, Hierarchical Mobile IP (HMIPv6) [Cas00] and Fast Handovers for Mobile IP (FMIPv6) [Koo05] are the most promising [PCTMH03]. The first aims to reduce the number of mobility signaling messages transmitted to the home subnet, while the latter aims to reduce the handoff delay by acting proactively once a handoff is deemed imminent.

1.3 Migrating to 4G

In 4G wireless All-IP networks, two contradicting demands exist: ubiquity and diversity. Users expect a large variety of services (with different QoS requirements) to be delivered across a diverse platform of mobile and wireless access technologies. Since 4G systems will be based on an IP core network, architectural considerations

[3] A subnet, or "subnetwork", is a logical group of connected network devices, typically within close physical proximity, sharing a common network prefix.

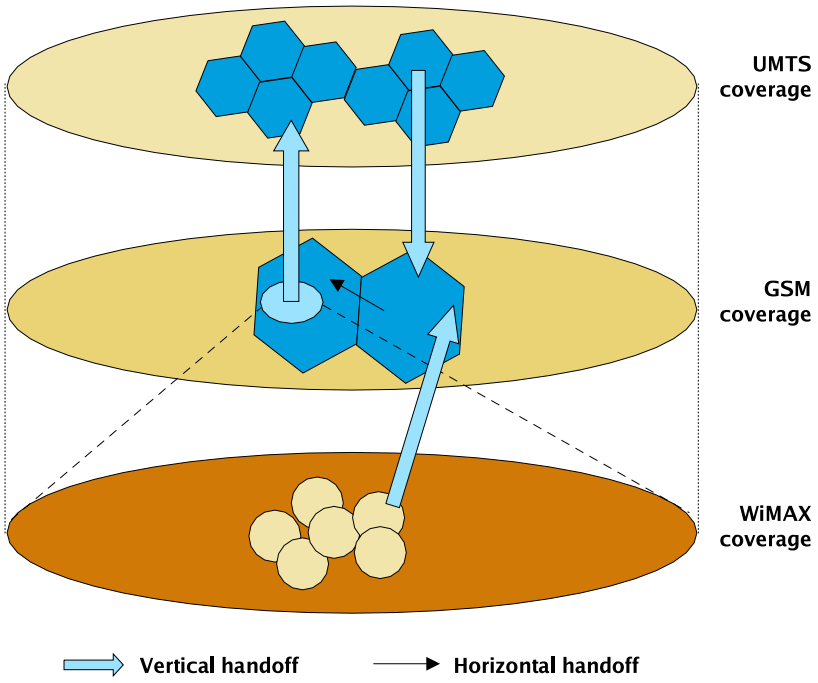

Figure 1.4: Vertical and horizontal handoffs of a mobile terminal [HY03].

in the IP layer become a key factor to 4G's success. 4G mobile terminals are expected to roam freely across different wireless systems and, in doing so, undergo handoffs both horizontally and vertically. Figure 1.4 [HY03] illustrates the concept of horizontal and vertical handoffs. A horizontal handoff occurs when a mobile terminal moves from one access point to another within the same wireless system (e.g. from one GSM cell to another). A vertical handoff, however, occurs when a mobile terminal moves from one wireless system to another (e.g. from WiMAX to GSM). The key challenges to migrate current systems to 4G have been reviewed extensively in [HY03]. In the work presented in this thesis, the focus is on two main network-level issues:

- **Seamless Mobility:** A mobile terminal should be able to roam across different access points with minimal disruption in service (i.e. handoff execution should be seamless at the application level).

- **Resource Control:** An active application should be restored to an acceptable level of performance (minimal, or no, degradation in QoS) once a handoff is completed.

Supporting QoS in 4G wireless networks poses a major challenge, mainly due to the varying characteristics of the different wireless systems involved. End-to-end QoS guarantees have to be established in the IP core layer (using RSVP). Once a mobile terminal undergoes a vertical handoff (e.g. from an IP-based WiMAX network to a UMTS network), the IP-level QoS parameters could then be mapped on to the equivalent UMTS-specific QoS parameters. Consequently, end-to-end IP-level QoS guarantees serve as a common ground in managing application-level performance for such heterogeneous wireless systems.

Internet telephony services in the UK shed some light on the future of 4G. Mobile Internet operator Truphone has launched the first mobile VoIP service for Wi-Fi enabled handsets[4]. Using the well-established Session Initiation Protocol (SIP) [RSC+02], users can place and receive VoIP calls when they are within the range of a Wi-Fi hotspot and use the licensed mobile network at other times. While the concept of Voice-over-Wi-Fi (VoWiFi) is not considered a pioneering development, a stepping milestone achieved by Truphone is its ability to facilitate seamless handovers between mobile and Wi-Fi networks. Although the precise definition of a true 4G network varies, Truphone's mobile VoIP satisfies the basic requirements of a 4G system: high data rates, all-IP infrastructure and the use of open Internet standards (e.g. SIP).

In terms of broadband capabilities, research in Japan reveals optimistic results: In controlled experiments, prototype phones were used to view 32 high definition video streams in a vehicle traveling at 2 km/h. Officials from NTT DoCoMo claim that the phones were able to receive data at 100 Mbps on the move and at up to 1 Gbps while stationary [DoC07]. This was further increased fivefold at the end

[4]Truphone (Software Cellular Network Limited) has declared itself the world's first 4G network operator with the launch of its mobile VoIP services based on Nokia's E-series handsets.

of 2006[5]. In spite of this substantial achievement, researchers argue that initial implementations may restrict data speeds to 20 Mbps due to the current high cost of the technology. This however is still considered a substantial improvement from the current 3G/3.5G speeds.

It is evident that the future of 4G looks bright. The two main motivations propelling its development are the broadband capabilities (always-on connection), and the economical incentives of a truly worldwide wireless network based on a hybrid of mobile and wireless access technologies.

1.4 Critical Issues and Research Aims

The research presented in this thesis focuses on resource control and seamless mobility in the IP layer which will be the backbone of the 4G technology [JTKK05]. Although QoS mechanisms, such as RSVP, address application-level performance efficiently, they were initially designed for land-based systems whose end nodes do not move. Once mobility of the end nodes is allowed, several issues arise which inevitably impact the RSVP performance.

The first issue relates to the failure of RSVP routers to establish reservations for a MN, such as a Truphone handset, in foreign subnets. As outlined in Section 1.2, a MN would start using a new IP address called the CoA once it moves outside the boundaries of its home subnet. In MIPv6, however, the Home Agent (HA) forwards the MN's packets to its new location using IP-in-IP encapsulation [Per96], which basically wraps packets with an external IP header. As a result, if a sender tries to start an RSVP session with the roaming MN, the RSVP messages would be first encapsulated by the HA and then forwarded to the MN at the foreign subnet. However, RSVP routers along the way would not be able to identify these "concealed" RSVP packets and hence no resources would be reserved for the MN.

[5]In December 2006 in Yokosuka, Kanagawa; Japanese mobile communications company NTT DoCoMo, claimed that it has achieved a maximum packet transmission rate of approximately 5 Gbps in the downlink using 100 MHz frequency bandwidth to a mobile station moving at 10 km/h.

Similarly, if a MN has an active RSVP communication session before it moves to another subnet, it will lose any existing reservations after a handoff occurs. This is because the MN's IP address has changed while the RSVP routers are still in the process of reserving resources using the old IP address. A MN would therefore have to initiate a new RSVP session using its new CoA. However, MIPv6 and RSVP were designed independently of each other and therefore operate as two distinct functional blocks. A MN with an existing RSVP session would therefore undergo two phases during a handoff: Mobility signaling and RSVP signaling.

The first relates to the basic MIPv6 handoff delay, which is the amount of time it takes the MN to connect to the new subnet and acquire a new CoA. During this time, any traffic sent to or from the MN is lost (i.e. destination host unreachable). The second phase refers to the amount of time it takes the MN to create a new RSVP session using its new CoA (RSVP signaling delay). During this time, communication is resumed with the MN, but receives best-effort treatment until the necessary resources are reserved on the new link.

At the application level, the above-mentioned phases would be observed first as a disruption in service (handoff delay), followed by a temporary substandard level of service (RSVP signaling delay). Since we are mostly concerned with real-time applications, we measure the Total Interruption in QoS (TI_{QoS}) as the time it takes the application to return to its previous QoS level. Therefore, TI_{QoS} is simply the sum of the handoff delay and the RSVP signaling delay. The aim of this research is to reduce this value as much as possible.

Furthermore, a handoff would most likely change only a small segment of the complete path between the sender and the receiver. However, RSVP currently has no mechanisms to identify this and has to therefore establish a completely new session along the end-to-end path. This results in double reservations on the unchanged portion of the link for the same MN (one reservation using the old IP address while the other using the new CoA). In severe congestion scenarios, a MN's own old reservation could possibly block it from acquiring a new reservation after a handoff.

With these issues in mind, a feasible reservation model should satisfy the following design requirements:

- Interoperable with Mobile IP (specifically with IP-in-IP encapsulation);
- Minimise TI_{QoS} in the event of a handoff;
- Localise RSVP signaling to the affected sections of the end-to-end path.

A key aspect in the design process of any optimisation problem is to identify the constraint. In this particular case, the optimisation constraint is to adhere to the current RSVP standard (RFC 2205) [BZB+97] and hence be backwards compatible with it. Although incorporating this constraint into the design process would limit the efficiency of the solution (since modifications to the protocol would be kept to a minimum), it has the advantage of simple integration into RSVP routers. Within the scope of this research, two reservation models are proposed. The first takes into account the optimisation constraint and hence requires minimal changes to the end nodes (without affecting other routers in the network). The second, however, reduces TI_{QoS} further, at the expense of requiring changes be made to all RSVP routers in the network. Nonetheless, one could argue that the second solution would still be considered viable since RSVP routers are not widely deployed and hence such a solution could be standardised and integrated into RSVP routers in the near future.

1.5 Thesis Overview and Contributions

The remainder of this thesis is organised as follows:

The next chapter lays the necessary background theory required for this research by presenting an overview of the Resource Reservation Protocol and the Mobile Internet Protocol. The functionalities and limitations of both protocols are discussed, in addition to a review of proposed extensions. The chapter then continues with a

critical review of the research community's related work in the area of QoS provisioning mechanisms in wireless IP networks. This literature survey was presented at the Third Workshop on the Internet, Telecommunications and Signal Processing (WITSP'04) in Adelaide, Australia, and published in the conference proceedings [BcM04].

In order to accurately measure any improvements achieved by newly proposed mechanisms, a performance benchmark needs to be established as a point of reference. Prior to this work, few speculations have been made by the research community as to how would RSVP and MIPv6 nodes most likely interact in real networks [Tho02]; although no experimental or analytical models were developed to fully comprehend the issue.

Chapter 3 addresses this requirement by introducing a performance analysis study to investigate the interaction of RSVP and Mobile IPv6 (including its extensions). The analysis framework is comprised of a simulation-based section to measure application-level performance, and a signaling cost analysis section to measure network-level performance. Preliminary results of this study were presented at the IEEE International Conference on Communications (ICC'06) in Istanbul, Turkey, and published in the conference proceedings [BcM06b].

Chapter 4 presents a mechanism for enhancing RSVP performance over Mobile IPv6 and its extensions, called Mobility Aware Resource Reservation Protocol (MARSVP). The key concept of MARSVP is to convey mobility-specific information using newly defined RSVP objects embedded in existing RSVP messages. This allows a single message exchange to establish both IP-level connectivity as well as QoS guarantees on the new link. The appealing attribute of MARSVP is that it requires minimal changes to end nodes and hence compatible with the RSVP standard RFC 2205. Preliminary results of this work were presented at the IEEE Global Conference on Telecommunications (GLOBECOM'06) in San Francisco, USA, and published in the conference proceedings [BcM06a]. Comprehensive results were submitted as a journal article to the Computer Communications journal,

published by Elsevier.

In Chapter 5, stronger emphasis is put on improving RSVP performance over Mobile IPv6 with less obligation to comply with RFC 2205. In view of this, a new packet classification mechanism for RSVP (called RSVP-HoA) is proposed in which routers are configured to classify flows based on the home address option in the MIPv6 destination options header. Through this approach, intermediate RSVP routers are able to correctly identify an RSVP flow, even after a MN changes its CoA. Moreover, a crossover router (COR) using this mechanism can detect the changed portion of the end-to-end RSVP session and confine RSVP signaling to the changed nodes. Results of this study were accepted for publication as a journal article in the Wireless Communications and Mobile Computing journal, published by Wiley InterScience.

Finally, in Chapter 6 we conclude the thesis by presenting our recommendations and identifying the research directions for future work in the field.

Chapter 2

Mobility and QoS Support in Wireless All-IP Networks

2.1 Overview

This chapter first reviews the basic principles of RSVP, MIPv4 and MIPv6 (in addition to two mobility extensions: HMIPv6 and FMIPv6). The functionalities and limitations of each protocol are discussed. Once the reader is familiarised with the basic concepts, the chapter then continues with a comprehensive literature survey of the research community's prior work in the field. The literature can be subdivided into two sections according to the approach used: RSVP-based approaches, and handoff-based approaches. The chapter then concludes with a summary of the remaining issues of RSVP and Mobile IP interaction and points out the method in which they should be addressed.

2.2 Resource Reservation Protocol

The Resource Reservation Protocol (RSVP) [BZB+97] was standardised by the Internet Engineering Task Force [IETF] to operate within the Integrated Service

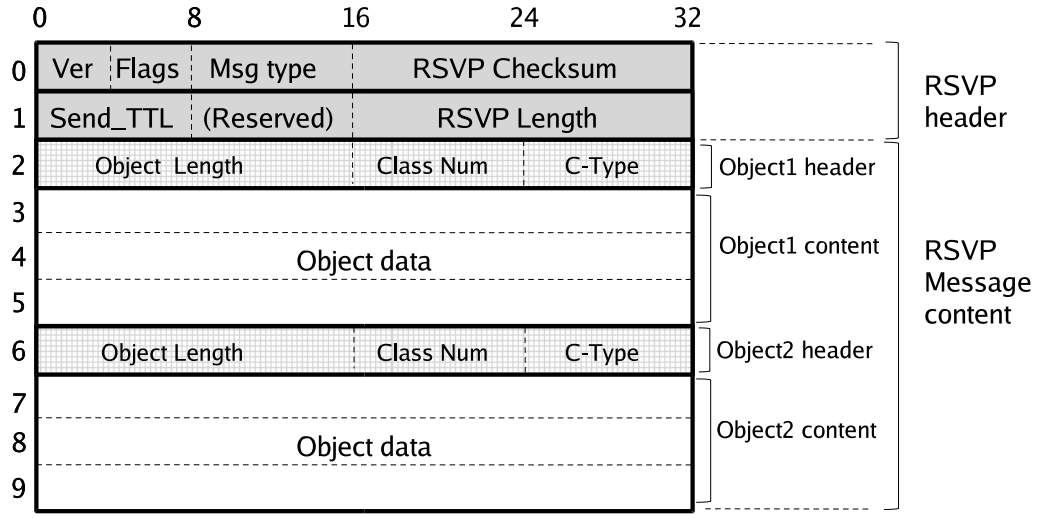

Figure 2.1: RSVP message format.

(IntServ) model [BCS94]. RSVP is essentially a network-control protocol used by hosts to guarantee certain levels of Quality-of-Service (QoS) for time-sensitive applications such as Voice-over-IP (VoIP) or Video-over-IP (VIP). RSVP is a receiver-initiated protocol in which the sender advertises the characteristics of its pending data flow (data rate, token bucket rate and token bucket size). If the receiver chooses to accept the connection, necessary resources (bandwidth required to satisfy the advertised data flow characteristics) are reserved upstream towards the sender.

The main motivation behind this receiver-initiated design is to cater for large multicast broadcasts where different receivers may request different levels of QoS. RSVP also permits merging of multiple reservations which increases its scalability for such broadcast applications. Within the scope of this thesis, however, RSVP's multicast and merging abilities are not delved into, mainly because we focus our attention to unicast communications typically used in real life mobile networks. It is also worth noting that RSVP does not perform its own routing but rather complements the underlying routing protocol by prioritising the way in which certain packets are handled by routers. RSVP is usually transported over UDP or directly over IP.

Two main message types exist in RSVP: The Path message and the Resv message. A sender establishes an RSVP session by sending a Path message which contains three important pieces of information known as Objects (Figure 2.1):

1. Sender Template, which essentially consists of information used to correctly identify packets that belong to the sender's data flow (e.g. sender's IP address and port number).

2. Sender Tspec, that specifies the characteristics of the traffic to be sent, and hence the desired level of QoS.

3. PHOP, which is used to store the IP address of the previous hop router.

As the Path message propagates downstream towards the receiver (Figure 2.2), it does not reserve any resources but rather installs what is referred to as the *Path State* in every intermediate router along the way. The Path State is used to store the IP address of the previous hop (PHOP) router. This information is used to ensure that when the receiver replies with a Resv message, it is routed hop-by-hop upstream along the reverse path of this associated Path message.

Once the Path message reaches its destination, the receiver has the choice of either accepting the connection or rejecting it. If the receiver accepts the connection, it generates a Resv message which contains the following RSVP objects:

1. FlowSpec: Defines the desired QoS level to be assigned to the data flow (may not necessarily be the same as the advertised QoS level in the Path message).

2. FilterSpec: Contains information needed to correctly identify the sender's data flow, in addition to the session information.

At every RSVP-enabled router along the way, the Resv message is processed by the router's RSVP module which first consults two decision modules (Figure 2.3):

1. Policy Control Module: To check whether or not the user has administrative permission to make the reservation.

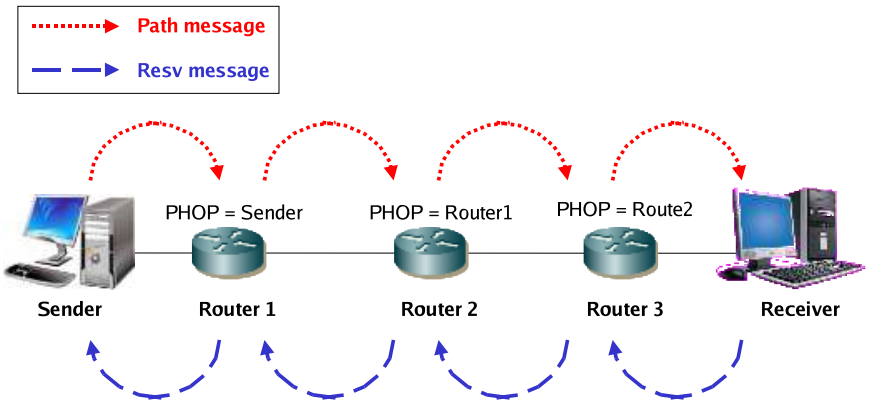

Figure 2.2: Path and Resv messages exchanged between sender and receiver to reserve resources along the way.

2. Admission Control Module: To check whether or not there are sufficient available resources to deliver the desired QoS level.

If the request successfully passes both decision modules, QoS guarantees are implemented using two modules that together constitute the Traffic Control Mechanism:

1. Packet Classifier: Used to classify packets and identify RSVP flows.

2. Packet Scheduler: Implements QoS for each flow, using one of the service models defined by the Integrated Services Work Group.

The RSVP module passes the FilterSpec object to the packet classifier and the FlowSpec object to the packet scheduler. This effectively establishes the required reservation for that particular flow and the Resv message is forwarded upstream to the next router using the address stored in PHOP. The process is repeated for every router along the way until it reaches the sender which is now ready to send its data flow. However, in the case that any router fails to reserve the necessary resources, a ResvErr message is generated and is sent downstream towards the receiver, relinquishing all preceding reservations.

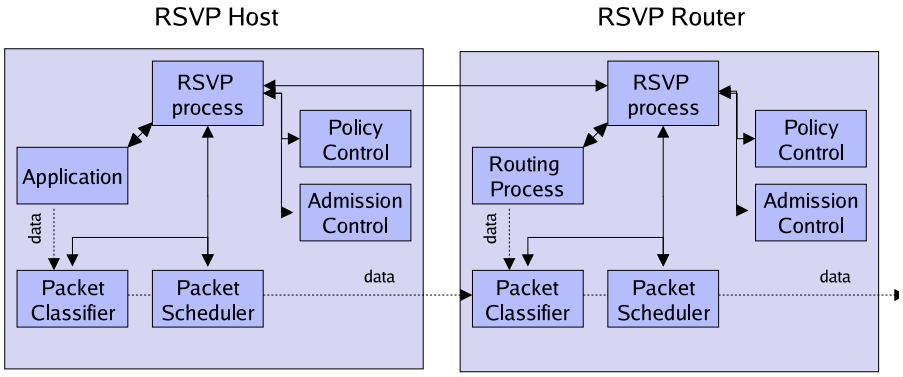

Figure 2.3: RSVP functional diagram.

If a reservation request is successful and the sender has sent all its data flow (i.e. is ready to terminate the connection), it sends a ResvTear message that traverses downstream towards the receiver, relinquishing the existing reservations and hence freeing up valuable network resources. However, the ResvTear message is not restricted to the sender: A receiver aware that the session has ended (or is no longer interested in the session) could send a ResvTear message upstream, thereby releasing the reservations.

An important characteristic of RSVP is that its reservations require periodical refresh messages to keep them alive. This is performed at set intervals where the sender generates a Path message down towards the receiver and expects a Resv message in reply. The advantage of this method is that if the transmission from the sender to the receiver ends abruptly (i.e. none of the two hosts sends a ResvTear message) the existing reservation's refresh timer (located in each RSVP router) would eventually expire after a certain period of time and hence resources would be freed regardless of the ResvTear message. The disadvantage of this approach is apparent: network resources are consumed by the periodical RSVP refresh messages. As a result, every active call will create significant signaling overhead.

Another important characteristic is that RSVP reservations are simplex. As a result, RSVP treats the sender as an exclusively independent entity from the receiver;

even though in duplex applications sender and receiver act simultaneously. Therefore for such applications, two distinct reservations would have to be made: One in the downstream direction (Caller A to Caller B) and another in the reverse direction (Caller B to Caller A).

Finally, RSVP is flexible in that it can operate through paths that cross networks which do not deploy RSVP-enabled routers. In such networks, if an RSVP message arrives at a non RSVP-enabled router, the router forwards the RSVP message without inspecting it or making any reservations. However, if this message reaches an RSVP-enabled router further down the way, reservations could still be established in that portion regardless of the preceding non-RSVP section. This proves useful in realistic implementations of reservations across the Internet. RSVP is typically deployed in access networks where bandwidth needs to be regulated and could be in scarcity; while in Internet backbones there is usually ample bandwidth and hence tight resource control may not be critical. Therefore a typical RSVP session would reserve resources at the sender's access network, followed by a non-RSVP cloud across the Internet, then once again reserve resources at the receiver's access network.

2.3 Mobile Internet Protocol

The Internet Protocol (IP) has proven to be a successful network layer protocol. It provides a host with an IP address used by other hosts to communicate with it across the Internet. IP achieves this by implementing a distinct addressing mechanism that splits an IP address into two portions: a network prefix and a host suffix. The first specifies the network in which the host is located while the latter uniquely identifies the host within that network. This addressing format allows routers to efficiently route packets across the Internet since IP packets are essentially routed according to their network prefixes. Once a packet arrives at the destination's edge router, the host suffix is used to deliver the packet to the concerned host in the destination

network.

Although this approach which couples closely a host's IP address to its home network facilitates universal packet routing, it is strictly confined to stationary nodes. Once the mobility of a host from network to network is permitted, complications start to occur: a Mobile Node's (MN's) IP address becomes invalid as it moves from one network to another since it no longer reflects the MN's current location at the foreign network (the network prefix still points to the home network). To address this issue, the IETF standard called Mobile IP (MIP) [Per02] was proposed. MIP provides an efficient mechanism that enables a MN to seamlessly roam across different subnets whilst maintaining its IP connectivity. This transparency above the IP layer allows a user, for example, to maintain an active application such as VoIP while moving freely from one WLAN to another. MIP accomplishes this by allowing a MN to have two IP addresses:

- Home Address (HoA): The conventional permanent IP address of the MN used by all nodes to communicate with it, regardless of its exact location.
- Care-of-Address (CoA): A temporary IP address assigned to the MN when it moves into a foreign subnet.

These IP addresses are managed by two MIP entities: The Home Agent (HA) and the Foreign Agent (FA). The HA is essentially a MIP-enabled gateway at the edge of the home network that stores information in its binding cache (a regularly-updated table containing information about local MNs). The FA, on the other hand, is a MIP-enabled gateway at the foreign network that assigns CoAs to visiting MNs. When a MN leaves the boundaries of its home subnet and enters a foreign subnet (Figure 2.4), it acquires a CoA from the FA and has to then notify its HA by sending it a Binding Update (BU) message which contains the MN's HoA and CoA. Upon receiving the BU, the HA creates an entry in its binding cache associating the MN's new CoA with its HoA, in addition to replying to the MN with a Binding Acknowledgment (BAck). Any packets destined to this particular MN (using the

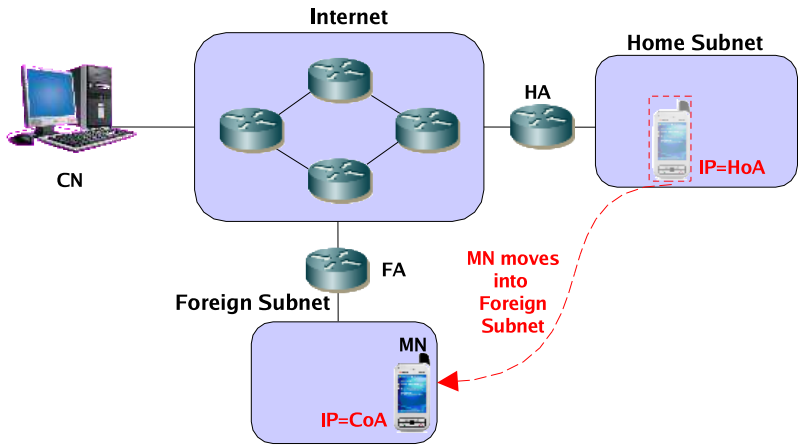

Figure 2.4: A Mobile Node undergoing a handoff in Mobile IP.

HoA) would be intercepted by the HA, encapsulated with the CoA as the destination address and forwarded to the MN (i.e. packets are tunneled from the HA to the MN). In essence, this forwarding mechanism is analogous to the conventional postal system in which a customer could move cross-country and acquire a new mailing address. The customer would then subscribe to the mail forwarding service, in which all mail destined to the old mailing address would be forwarded by the local post office to the customer's new mailing address.

As is the case in such postal systems, it is more efficient for the MN to directly receive its packets (using the CoA) rather than have them tunneled by the HA. A Correspondent Node (CN), for example, could be co-located with the MN at the visited foreign network but would still have to send packets using the HoA across the Internet to the MN's home subnet, only to have them tunneled back to the MN by the HA. This indirect routing delays the delivery of the MN's packets, in addition to placing unnecessary burden on networks and routers along the way. To overcome this, an extension called Route Optimisation was proposed [JP00] where the MN sends an additional BU to the CN. This notifies the CN of the MN's new CoA, thereby facilitating direct communication and avoiding the previously mentioned triangular routing.

The initial MIP proposal was principally designed to offer mobility to IPv4 nodes, and is hence duly referred to as MIPv4. With the advent of IPv6, Mobile IP naturally evolved from MIPv4 to MIPv6. The new MIPv6 proposal not only offers the inherit large addressing space of IPv6, but also adds several improvements [JP04]:

- MIPv6 does not require a dedicated FA to assign CoAs at the foreign network.
- Route optimisation is a fundamental part of MIPv6 and is not an added extension.
- Route optimisation operates securely without pre-arranged security associations.

In MIPv6, a MN acquires its CoA using either a stateful or stateless address configuration. Stateful address configuration utilises the Dynamic Host Configuration Protocol for IPv6 (DHCPv6) [DBV+03] and hence requires a dedicated DHCP server (located within the boundaries of the visited network) to assign and control IP addresses. This address configuration mechanism is useful if the network administrator requires tight control of addressing. Stateless auto-configuration, on the other hand, gives the MN the flexibility to configure its own CoA. This is accomplished by appending the MN's Ethernet hardware address (also known as MAC address) to the network prefix received in the advertisement generated by the visited access router. Furthermore, according to [TN98], the MN should perform Duplicate Address Detection (DAD) in order to verify the uniqueness of the generated IP address. This is done by the MN advertising the generated CoA, and if no hosts reply indicating that an address conflict exists, the CoA is considered valid and is hence assigned to the MN.

2.4 Hierarchical Mobile IPv6

Although Mobile IP enables seamless mobility of end nodes, it does not take network scalability into account: Every time a MN performs a handoff, it acquires a new CoA and has to send BU messages across the Internet to its HA and CN (in addition to receiving the associated acknowledgments). When considering the growing number of mobile devices and available access networks, MIPv6's signaling overhead could impose a significant burden in such high mobility scenarios.

Hierarchical Mobile IPv6 [Cas00] was introduced as an extension to MIPv6. HMIPv6 organises a foreign network into a multi-level hierarchy architecture, isolating a MN's global mobility (movement across different subnets) from its local mobility (movement between two access points that belong to the same subnet). When a MN first enters a HMIPv6-enabled foreign network (i.e. global mobility occurs), it acquires two CoAs from the Mobility Anchor Point (MAP) which is essentially an FA at the highest level of the hierarchy (Figure 2.5):

- Regional Care-of-Address (RCoA)
- Local Care-of-Address (LCoA)

The RCoA is valid throughout the MN's duration of stay at the foreign network (regardless of any local movement), and is therefore the address used by external nodes to communicate with it. The LCoA, on the other hand, is valid only on the exact link that the MN is connected to and is only known to the MAP, in addition to any local nodes within the foreign network. The MN then notifies its HA and CN of the RCoA through a BU message. It also notifies the MAP of its exact location by sending it another BU message indicating its LCoA. As a result any packets destined to the MN are sent using the RCoA. Once a packet reaches the MAP, it checks its binding cache and retrieves the MN's LCoA, encapsulates the packet with the LCoA, and then forwards it down to the MN's exact location.

Although this multi-addressing mechanism might seem inefficient at first, it sig-

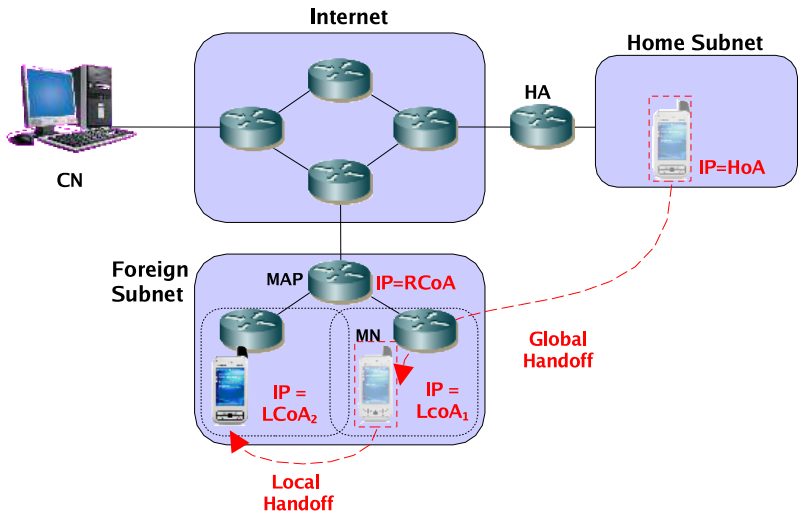

Figure 2.5: Local and Global Handoffs in Hierarchical Mobile IPv6.

nificantly reduces signaling overhead and handoff latency when considering local mobility: When the MN experiences a local handoff (i.e. connects to a different access point within the same HMIPv6 foreign network), it only acquires a new LCoA whereas the RCoA remains unchanged. The MN then notifies its MAP of its new location by sending it a BU message. This hierarchical addressing limits local mobility signaling to the edge of the foreign network, in addition to significantly reducing handoff latency since the BU only traverses to the local MAP (as opposed to across the Internet to the HA and CN). As a result, the MN's local movement at the foreign network is completely transparent to the HA and CN.

Due to its efficiency in managing local mobility, HMIPv6 is often referred to as a Micro-Mobility protocol and is used within access networks. MIPv6 on the other hand, is used for global mobility (Macro-Mobility protocol). Another advantage of HMIPv6 is that it is backwards compatible with MIPv6: In the case that a MN's home network deploys only MIPv6, the MN could still use HMIPv6 at the foreign network while using the RCoA as the *conventional* CoA to communicate with its HA.

2.5 Fast Handovers for Mobile IPv6

Fast Handovers for Mobile IPv6 (FMIPv6) [Koo05] was developed to reduce the latency perceived by a MN during a handoff, without much emphasis on mobility signaling load. The main goal of FMIPv6 is to allow a MN to pre-configure its CoA before it moves into the new subnet, and be able to immediately use it once it gets connected to the new Access Router (nAR). Another key advantage of FMIPv6 is that it also establishes a temporary forwarding tunnel between the old Access Router (oAR) and the nAR during handoff execution. The idea is that once the new CoA (nCoA) has been negotiated and the MN is about to disconnect from its existing subnet, the oAR is requested to forward any incoming data for the MN to the nAR. This temporary tunnel ensures that no packets are dropped during handoff execution since the nAR would buffer them temporarily and once the MN reconnects at the nAR, the nAR would deliver them using the nCoA.

A handoff is essentially comprised of two sub-processes: the Layer-2 handoff and the Layer-3 handoff. The first consists of the physical process of the MN connecting to the new access router, while the latter consists of network-level signaling required to gain a new IP address and hence resume IP-level connectivity. A key requirement of FMIPv6 involves the anticipation of the MN's movement.

This requires the involved nodes to gather information from lower layers (e.g. signal power measurements) to inform the Layer-3 mechanism (FMIPv6) that a handoff is about to occur. This ensures that the Layer-3 handoff commences before the one at Layer-2 (i.e. *make-before-break*). However, in the case that FMIPv6 fails to anticipate a handoff, a MN could still perform a conventional handover as outlined in the MIPv6 standard. Four message types are introduced in FMIPv6:

- Router Solicitation for Proxy *(RtSolPr)*;
- Proxy Router Advertisement *(PrRtAdv)*;
- Handover Initiation *(HI)*;

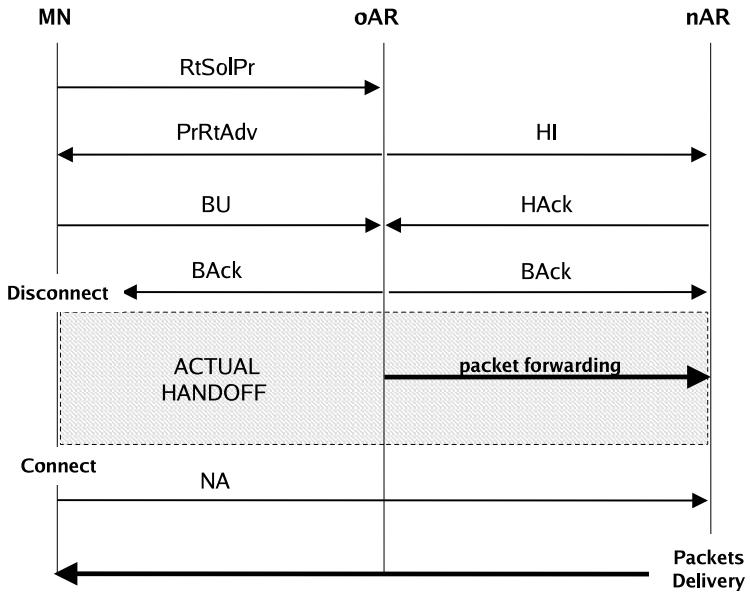

Figure 2.6: FMIPv6 Signaling.

- Handover Acknowledgment *(HAck)*.

FMIPv6 is executed in two phases, the first is *pre-handoff* and involves nCoA configuration while the second is executed during the actual handoff and involves triggering the forwarding mechanism:

Pre-handoff

When a MN receives information about an anticipated handover (through Layer-2 triggers), it sends a RtSolPr to its oAR containing the IP address of the nAR to which it wishes to attach to (Figure 2.6). The oAR uses this information to configure the nCoA and sends this information back to the MN in a PrRtAdv message. The oAR also sends a HI message to the nAR, informing it of the MN's current CoA and the proposed nCoA. The nAR checks the nCoA and, if valid, will reply with a HAck to the oAR; the two ARs then stand by for the actual handoff.

Handoff Execution

When the Layer-2 handoff is imminent, the MN sending a BU to the oAR just before it disconnects. Upon receiving the BU, the oAR sends two BAck messages, one to the MN (to inform it that the binding was successful) and another to the nAR to notify it of the incoming forwarded packets for the MN. During the actual handoff, the oAR forwards all the MN's packets to the nAR which temporarily buffers them. Once the MN enters the nAR's subnet, it notifies the nAR by sending it a Neighbour Advertisement (NA), to start receiving the forwarded packets. The MN continues receiving its forwarded packets (using the oCoA) until it notifies the CN and the HA of the nCoA through two BU messages. Once the CN and the HA are notified, packets could then be delivered directly to the MN using its nCoA.

2.6 QoS Provisioning Mechanisms in Wireless Networks

2.6.1 The Two Approaches for QoS Provisioning in Wireless Networks

IP protocols have been designed to operate in wired networks using fixed IP addresses. The behaviour of these protocols can be affected (sometimes considerably, as in the case of RSVP) once mobility is allowed and IP addresses are dynamically assigned and changed constantly. This is largely due to the fact that protocols supporting mobility and QoS have been developed independently of each other.

To resolve this matter, two main approaches have been undertaken by the research community (Figure 2.7): The RSVP-based approach, and the Handoff-based approach. The first aims to modify default RSVP in order to make it more efficient and feasible in wireless scenarios, while the latter modifies the mobility signaling mechanism to incorporate QoS signaling. The following sections review several

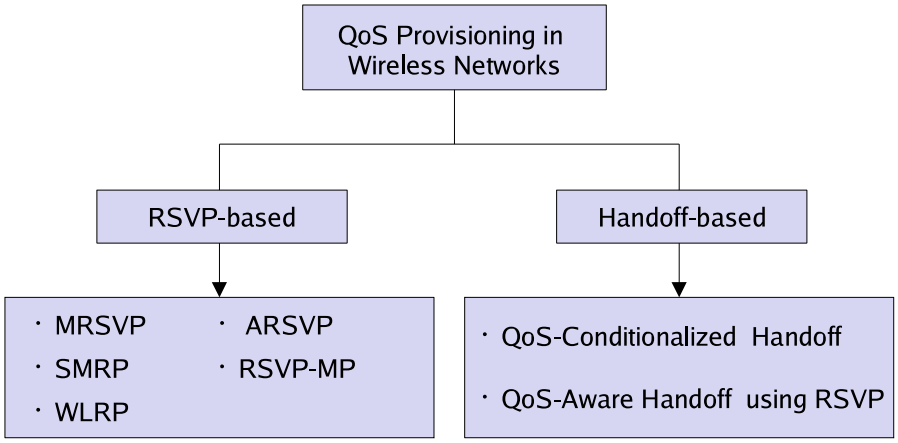

Figure 2.7: QoS Provisioning Mechanisms in Wireless Networks.

proposals that fall within these two approaches, outlining the contributions and limitations of each.

2.6.2 RSVP-based Approaches

Mobile Resource Reservation Protocol

Talukdar *et al.* [TB01, TB99] proposed one of the first extensions to RSVP, called Mobile RSVP (MRSVP) in order make it feasible for deployment in wireless networks. MRSVP relies primarily on advance resource reservations in neighbouring subnets made by the Mobile Agent (MA) on behalf of the MN. In order to achieve this, however, a MN must provide a message called *MSpec* which contains information about the MN's movement and the prospective subnet to be visited. MRSVP also supports two type of reservations: *Active reservations* (currently used by the MN) and *Passive reservations* (advance reservations for the MN but not currently allocated to it). Three new message types are also introduced by MRSVP:

1. MSpec message.

2. Passive Path message.

3. Passive Resv message.

A typical MRSVP session commences in a similar manner to that of standard RSVP: A sender sends a Path message to the receiver (installing Path state in routers along the way), and the receiver (MN) replies with a Resv message to activate the soft-state reservations along the reverse path. MRSVP, on the other hand, adds the following procedure: The MN also sends an MSpec message to its MA which establishes passive reservations (using the Passive Path and Passive Resv messages) on behalf of the MN at the nominated subnets indicated in the MSpec message. Once the MN moves into one of these subnets, the passive reservation in that particular subnet is activated while the active reservation in the old subnet becomes a passive one (Figure 2.8). Although this approach reduces the time required to re-establish an RSVP session after a handoff, it suffers from several drawbacks:

- It assumes a MN's movement to be deterministic, which is not necessarily the case in real life scenarios.
- It forms a possibly large set of passive reservations which affects the network's bandwidth utilization.
- A MN may have to suffer from a long waiting time before all the passive reservations are completed in order to start receiving its MRSVP data flow.
- In order to make the passive reservations, a communication protocol needs to be implemented between the mobility proxies; which increases the complexity of the network.

To improve network utilisation, passive reservations in neighbouring subnets could either be assigned to other data flows requiring weaker QoS levels or used for best-effort services. However, when the MN moves into the subnet and activates the reservations, theses flows may be affected. Mahmoodian *et al.* [MH99] proposed a *Progressive Resource Registration* mechanism as an extension to MRSVP in order

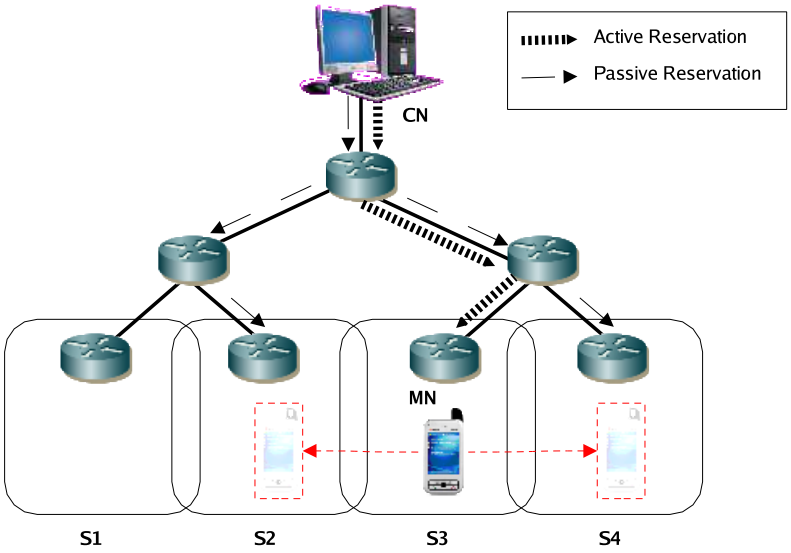

Figure 2.8: MRSVP active and passive reservations.

to address the first issue mentioned above: The MA acts as an RSVP sender and distributes a Path message to all neighbouring mobile proxies (rather than use MSpec to send a Passive Resv message to nominated mobile proxies). This method checks for the available resources at router-level along the paths to all surrounding MAs. Each MA could then either reply with a Resv message (invoking passive reservations) or reject the reservation with a ResvErr message. Although this procedure eliminates the need for the MSpec message, it still suffers from the same disadvantage of poor resource utilization, as well as additional traffic overhead (since all surrounding cells are invoked).

Sender-initiated and Mobility-support Reservation Protocol

Shangguan *et al.* [SSK00] proposed Sender-initiated and Mobility-support Reservation Protocol (SMRP) as a fundamental modification to RSVP, rather than a complementing extension. The authors argue that RSVP was designed specifically for multicast groups and hence incurs additional processing and storage overheads on the network routers. This is mainly due to RSVP's receiver initiated approach:

Each receiver makes its own reservation based on the information advertised by the sender. This detachment of the path-finding process and reservation-setup is reflected in RSVP's implementation (Sender sends Path, receiver replies with Resv)

SMRP is essentially a sender-initiated protocol, adapted for unicast communication. The path selection and resource reservation is combined into one process. An SMRP sender initiates a session by sending a *Request* message to the receiver. This message is processed by every intermediate router along the way, checking for the available resources. Every router stores the reservation request as a success or a failure, modifies the Request message accordingly and forwards the message downstream to the receiver. The receiver then simply returns these results through an *Echo* message sent back to the sender. This technique cuts the processing time in half since all processing is done in the Request message while the echo message simply informs the sender of the results and is not processed by any routers.

Another advantage of SMRP is the improved soft-state reservation mechanism: A sender does not have to periodically send Request messages in order to refresh the reservations as along as data is being transmitted. Request messages are only required in idle mode (no data is being sent to the receiver).

Given the above mentioned advantages, SMRP still suffers from one setback: An SMRP Agent needs to be installed in all nodes along the end-to-end path between the sender and receiver in order to successfully deploy SMRP. Although SMRP could function in non-SMRP clouds, it would not be able to function if any RSVP routers exist at any point down the line. Therefore from a practical point of view, SMRP would not be commercially appealing since it is not backwards compatible with RSVP.

Wireless Lightweight Reservation Protocol

The Wireless Lightweight Reservation Protocol (WLRP) [Par03] proposed by Parameswaran, incorporates loss tolerance into the reservation process. A MN not only specifies the required QoS level, but also the degradation level it is willing to toler-

ate. The author argues that this approach increases the probability of an on-going application to establish a successful reservation in the visited subnet. Moreover, WLRP utilises passive reservations in neighbouring subnets and implements hard-state for its active reservations in the existing subnet. A WLRP-enabled MN periodically sends out two messages: A *Mobility Profile* (anticipated motion path), and an *Application Profile*, which consists of the following:

- *br*: Application data rate.
- *LNEG*: Loss negotiability, which is a negative degradation of the application's acceptable loss in data rate (from 0 to 1) that an application is willing to accept in order to avoid rejection under overload.
- *LProfile*: Loss Profile, an application can choose either a distributed or a bursty loss.
- *QHO*: Levels of service an application expects during a handoff.

The Mobility Agent (MA) monitors the periodically received Mobility Profile (to predict the MN's anticipated subnets to be visited) and Application Profiles (to send passive reservations to the nominated subnets). The success or failure of these passive reservations is then fed back to the MN, thereby providing the user with a QoS forecast of the surrounding subnets. In the case that the passive reservations fail in a particular subnet, the user has the choice of either changing the route to another subnet or staying in the current subnet.

Unlike RSVP, WLRP requires no refresh messages since its active reservations are hard-state. An active reservation expires by an explicit Teardown message sent to the old subnet. Passive reservations, on the other hand, are soft-state and can be used to support best-effort traffic in neighbouring subnets until they becomes active when the MN enters the subnet. Since these passive reservations are soft-state, they automatically expire after a network-tunable duration of time.

When reviewing WLRP, the true benefit of utilising hard-state reservations seems particularly questionable. In hostile environments of wireless communications, the probability of an active session being ended abruptly is significant. For example, a session could be terminated due to low signal levels (call drop), or a mobile device's battery running out. In such conceivable situations, an active WLRP reservation would theoretically be held for an indefinite amount of time since the MN would not have sent a Teardown message to the subnet.

Another questionable attribute of WLRP is its tolerance for a lower level of QoS. From a commercial point of view, a service provider is held accountable for the level of service it provides. Whereas in WLRP, there is no specific guarantee to deliver the desired QoS. Some applications, such as Voice and Video over IP, explicitly specify a set of QoS parameters necessary to ensure an acceptable level application performance. In WLRP however, there is always a chance that a user will be assigned the lower-bound QoS level. With a growing number of occurrences, this could affect the user's experience and dissatisfaction with the service provider.

Adaptive Resource Reservation Protocol

When a handoff occurs, only a specific portion of the entire end-to-end path between the sender and receiver is affected. Nonetheless, RSVP treats this session as a completely new one and hence renews the entire end-to-end path. Adaptive RSVP (ARSVP) [MTT+03] was proposed to confine the RSVP re-establishment process to the changed portion of the link. Routers are required to record the next hop (NHOP) router address, in addition to the standard PHOP (as specified in the RSVP specification). ARSVP introduces a new message, called *Search*, used to identify the changed nodes in the event of a handoff. When a MN with an existing RSVP session is about to disconnect from its subnet, the following procedure is followed (Figure 2.9):

1. A MN sends a Search message from its old Access Router (oAR) to the new

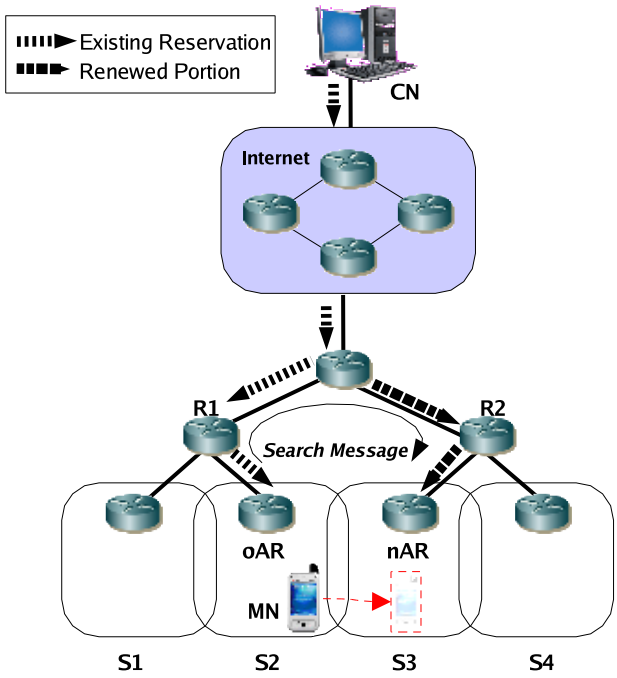

Figure 2.9: ARSVP existing and renewed reservations.

Access Router (nAR).

2. Since this message travels from the oAR to the nAR, it effectively passes through the nodes that will change in the new RSVP connection.

3. As routers receive and process this Search message, each records the IP address of the PHOP router and the NHOP router as $PHOP_{SEARCH}$ and $NHOP_{SEARCH}$ respectively.

4. The $PHOP_{SEARCH}$ and $NHOP_{SEARCH}$ are compared against the original RSVP session's PHOP and NHOP stored in each RSVP router. This effectively identifies the changed nodes in the link, the routers then update their respective Path states with the IP address of the new nodes.

5. Once the MN completes the handoff, it receives this Search message from the nAR, and replies with a Resv message to the sender.

6. When a router receives this Resv message, it refers to its Path state entries and either reserves the new link (if it is part of the changed portion) or maintains the current link reservation and forwards the Resv message upstream (if it hasn't changed).

In this process, only the reservations in the changed routers are renewed while the routers that are common between the old and new path are left unaltered. An important observation is that ARSVP assumes that, if a router is common between the old and new RSVP session, it would not be renewed and the Resv message would simply be forwarded upstream towards the sender.

The RSVP Packet Classifier, however (Figure 2.2), classifies RSVP packets according to the IP addresses and port numbers of the sender and receiver. Since the MN would change its IP address after a handoff, the reservations would therefore not be applicable to it. This implies that ARSVP assumes that a MN's IP address doesn't change after the handoff and is therefore suitable only for simple WLAN scenarios that deploy several access points serving a single access router. In such architectures, users are assigned unique static IP addresses (e.g. a university campus). Hence for true seamless mobility, a feasible solution should cater not only for changed nodes, but changed IP addresses as well.

RSVP Mobility Proxy

The key idea behind RSVP Mobility Proxy (RSVP-MP) [PKZM02] is to enhance QoS and mobility functions while at the same time minimising the modifications required to the existing infrastructure and protocols. RSVP-MP relies on a hierarchical mobility architecture (such as HMIPv6) since it isolates signaling within the access network. In the same manner that a MAP controls mobility signaling at the edge of the access network, RSVP-MP is responsible for RSVP message handling. RSVP-MP could be implemented at the MAP, although it is not a crucial requirement.

As outlined in section 2.4, HMIPv6 allows a MN to have two IP addresses: the RCoA for communication with external nodes, and the LCoA for internal communication within the access network. If default RSVP is deployed, the IP-in-IP encapsulation performed at the MAP would effectively conceal the RSVP messages' identity from routers, and would thus be forwarded as normal data packets (i.e. no reservations are made). This is because the IP-in-IP encapsulation assigns an outer packet header with a protocol ID number of 4 while RSVP uses 46, in addition to concealing the router alert option (RSVP uses this to inform routers that the packet needs to be processed).

RSVP-MP modifies the content of inbound and outbound RSVP messages, swapping the MN's RCoA and LCoA (depending on the direction of the packet). This method ensures that reservations are successful regardless of the HMIPv6 architecture since RSVP-MP avoids the IP-in-IP encapsulation of RSVP messages at the MAP. This is performed using Dynamic Address Translation (DAT). For an RSVP message originating from a MN inside the network to a CN outside the network, RSVP-MP swaps the LCoA with RCoA (in the source address field). Similarly, any packet destined to a MN residing in the network is intercepted by RSVP-MP and the RCoA is swapped with the LCoA (in the destination field).

In the event of a local handoff where the MN is the receiver, a MN sends a BU (with the new LCoA) to the MAP which notifies the RSVP-MP. Since the MN is the receiver and therefore can not renew its RSVP session using a Resv (it needs to reply to a Path message), RSVP-MP sends a Path message to the MN on behalf of the CN (using the CN's IP address). This triggers the MN which replies with a Resv message; RSVP-MP then intercepts this message and swaps the LCoA with the RCoA. Although this Resv message does not create any reservations beyond the MAP (the RCoA hasn't changed), it is forwarded as a periodical RSVP refresh message.

Similarly if the MN is a sender and undergoes a local handoff, it issues a Path message towards the CN in order to re-establish the RSVP session. The RSVP-MP

intercepts this message and replies with a Resv message to the MN (using the CN's address). The RSVP-MP also swaps the LCoA with the RCoA in the Path message and forwards it to the CN (as a refresh message).

As can be observed, RSVP-MP not only enables reservations in a HMIPv6 network (avoiding IP-in-IP encapsulation), but also significantly reduces the RSVP re-establishment time after a handoff (RSVP signaling is exchanged between the MN and RSVP-MP rather than the CN). RSVP-MP is an efficient protocol that moves the complexity away from the end nodes, and performs central packet processing at the edge of the access network.

Nonetheless, RSVP Operation over IP Tunnels [TKWZ00] seems to achieve the same goal with less complexity (no RCoA and LCoA swapping), in addition to not being restricted to a HMIPv6 network (could be applied to any IP tunnel). RSVP Operation over IP tunnels, recursively sends a Path message at the tunnel entry point and creates a separate RSVP session over the tunnel. Therefore, when this mechanism is used with HMIPv6 it effectively establishes two independent RSVP sessions (the external one and the local one). This enables the MN to re-establish the local session in the event of a local handoff without modifying the external session.

2.6.3 Handoff-based Approaches

The previously mentioned approaches are all considered extensions of RSVP and hence implement two independent functional blocks: The handoff procedure and the RSVP re-establishment procedure. The following approaches aim to integrate the two procedures into one, thereby reducing the signaling load and handoff latency experienced by a MN.

QoS-Conditionalised Handoff

Qos-Conditionalised Handoff [SK03] is designed over a HMIPv6 infrastructure in order to utilize its inherited advantages. A QoS option is added to the IP header of

the Binding Update message (BU) to include QoS-related data, and is consequently called a *BU+QoS* message. This approach allows the MN to perform a one-pass check for resource availability during the handoff procedure. Furthermore, the proposed approach assumes that the coverage areas of the wireless subnets overlap in order to provide the MN with a choice during handoff execution.

When a handoff is about to take place, the MN sends a BU+QoS message to the MAP via the new Access Router (nAR). As the message propagates upstream, each router passes the QoS parameters (stored in the QoS option) into its internal QoS mechanism. Resources are checked for availability and are reserved if available. The message is then forwarded to the next hop. If resources are not available, a negative Binding Acknowledgment (BA+QoS) is sent downstream to the MN (releasing any prior reservations made in previous hops). The MN then has a choice of either retrying with lower QoS requirements or choosing a different access router. However, if all resources are available (BU+QoS message reaches the MAP) a positive BA+QoS is returned to the MN and the MAP's binding cache is updated to reflect the new LCoA. Finally, the MAP sends a Teardown message towards the old Access Router (oAR) to release the reservations in the old path.

The advantage of this proposal is apparent in the shorter time delay due to the merging of the handoff process and the RSVP re-establishment. On the other hand, the fundamental drawback of the proposal is that it explicitly requires the MN to send the BU+QoS message through the new access router *before* the actual handoff occurs. This puts a strict requirement on highly overlapping coverage areas, in which the MN could access both the old and the new access routers simultaneously during the execution of the proposed mechanism: Maintain the connection with the oAR to continue receiving data packets while resource checking is done through the nAR. Another point to consider is that the proposed mechanism requires all nodes involved to be modified, which is not highly appealing from a commercial viewpoint.

QoS-Aware Handoff using RSVP

QoS-Aware Handoff using RSVP [BA03] works in a preemptive manner and focuses on reducing the latency of handoff call admission control. Handoff call admission control mandates that, if a connection requires certain levels of QoS, it should not be accepted during a handoff unless resources are available in the visited subnet (i.e. call is dropped if resources can not be allocated). This puts a stringent obligation on the network to either completely fulfill the QoS requirements or decline the connection. Even though this approach could increase the call block probability, it would provide the network administrator accurate feedback regarding the level of service provided and hence assist in more accurate network planning and improve the overall performance of the network.

The aim of the proposed mechanism is to reduce the latency of handoff call admission for QoS traffic by performing the resource availability check well before the handoff occurs. This results in a faster handoff response since result of the QoS check is readily available in a continuously updated database. This is achieved through the decomposition of the conventional handoff message into two:

1. **Pre-Handoff Info**: Is a non-real time component containing handoff-related information such as: Mobile ID, Traffic Specification, QoS Expectations and Resource Demands.

2. **Handoff Request**: Real time component containing minimal information (Mobile ID).

The proposed mechanism also introduces two new RSVP messages: *PathQuery* and *ResvQuery*, used to check for resource availability without making any reservations. Moreover, Expected Visitor List (EVL) processors need to be installed at all surrounding subnets. An EVL processor maintains records of the candidate MNs in EVL entries indicating the Pre-Handoff Information along with two dynamic fields: Decision (*Accept* or *Reject*) and Validity (*Valid* or *Invalid*).

The Pre-Handoff information message is sent from all MNs to their respective local EVL processors. The EVL processors then update their EVL entries with the gathered information (the Decision and Validity fields however are not yet filled). Next, all EVL processors exchange the generated information amongst themselves and hence every EVL processor gains the handoff information of the neighbouring MNs. A resource availability check is made on behalf of these candidate MNs by the EVL processors of surrounding subnets. This is done by sending *PathQuery* messages to the respective CNs. Unlike the conventional RSVP Path message, a PathQuery message only checks for resource availability without making any reservations.

The CN then either replies with a *ResvQuery* message indicating that the required resources are available (Decision field marked as *Accept*) or a notification of failure is sent back and the Decision field is marked as *Reject*. Regardless of the Decision result, the Validity field is marked as *Valid* initially. However, if any of the resource status in the network are changed (capacity increase or decrease), it is reflected in the EVL entry as follows: An increase in capacity would change a *Reject* decision's Validity field from *Valid* to *Invalid*, while an *Accept* decision remains *Valid* (since more resources are available). In contrast if the capacity decreases, an *Accept* decision would become *Invalid* while a *Reject* would remain unmodified.

The method proposed reduces the handoff admission control for QoS traffic flows by preemptively checking for resources availability. Although this process helps reduce the decision-making process of accepting handoffs, it does not affect the actual handoff execution. This means that if a handoff is accepted, the MN would still have to perform mobility signaling and the RSVP signaling independently. Another point to consider the signaling load on the network. The first concern since the exchange of PathQuery and ResvQuery messages between EVL processors and the respective CNs of surrounding MNs. The second concern is the synchronisation of EVL entries amongst the different EVL processors. In a highly congested network, this could affect the network's performance.

Table 2.1: Summary of Surveyed Proposals.

Proposal	*Mobility Prediction*	*New Messages*	*Modifications*	*New Nodes*
MRSVP	Y	9	MN and CN	Proxy Agents
SMRP	N	3	All network entities	-
WLRP	Y	3	MN and MA	-
ARSVP	N	1	MN and all Internal Routers	-
RSVP-MP	N	0	DAT on all RSVP packets	Mobility Proxy
QoS-Conditionalised Handoff	N	0	MN and all Internal Routers	-
QoS-Aware Handoff	Y	2	MN and all Internal Routers	EVL Processors

2.7 Conclusion

This chapter outlined the necessary concepts required for the remainder of the thesis. RSVP's functionality, in addition to Mobile IP and two of its enhancements, was explored. Although RSVP is considered a mature protocol in fixed networks, offering adequate levels of quality, various issues arise when considering deployment in wireless and mobile environments. This is largely due to the independent and non-synchronised standardisation of RSVP and Mobile IP, which did not take into account the joint performance of the two protocols.

The chapter also presented a literature survey of the research community's re-

lated work in the field (summarised in Table 2.1). All of the examined proposals reflected the diversity in the approaches used, whether by extending the standard RSVP or by embedding QoS functionalities in existing handoff protocols. At the present stage no flawless solution exists, but rather compensations that have to be made. Each proposal tries to improve a certain aspect of the mechanism (e.g. signaling overhead, latency or complexity of the signaling protocol) at the expense of another. For example, MRSVP provides a higher level of transparency and shorter latency at the expense of excessive number of advance resource reservations. The two handoff-based approaches, on the other hand, impose a higher level of complexity on the wired-network in order to alleviate the load on the wireless nodes by integrating QoS signaling with handoff signaling.

Based on this literature survey, more efficient wireless QoS solutions were investigated within the scope of this research. First, a performance benchmark and methodology were established in order to serve as a reference point and to accurately measure any improvements achieved. The next Chapter provides the details by introducing a performance analysis study comprised of a simulation-based section to measure application-level performance, and a signaling cost analysis section to measure network-level performance.

Chapter 3

A Framework for Analysing RSVP Performance over Wireless Networks

3.1 Overview

The growing popularity of advanced multimedia services has highlighted the importance of Quality of Service as a key network ingredient to allocate resources and ensure acceptable levels of application performance to the end users. However, the introduction of node mobility makes QoS provisioning an even more challenging task. Since RSVP and Mobile IP were developed independently of each other, they can work quite efficiently when deployed separately. Yet if their functionalities are combined, several inefficiencies arise in terms of QoS deterioration and underutilisation of network resources. The main reason being that RSVP was designed for end-systems whose IP addresses do not change. Once mobility of an end-system is allowed, the dynamically changing Mobile IP address inevitably impacts RSVP performance.

Various speculations have been made by the research community as to how would an RSVP and a Mobile IP entity most likely interact in real networks, although no experimental or analytical models have been developed to date for a

full comprehension of the issue. The framework presented in this chapter aims to address this issue by quantifying the impact of mobility protocols on RSVP performance. It is comprised of two components: A simulation model to assess the application-level performance as perceived by the end user, and an analytical model to examine the signaling costs incurred at the network-level. The results presented in this framework serve as a performance benchmark and a reference point when proposing a more efficient QoS solution.

The simulation model is particularly focused on the effects of end-to-end packet delay and packet loss on a Voice-over-IP (*VoIP*) or Video-over-IP (*VIP*) session under different congestion scenarios. The E-Model (ITU-T Recommendation G.107 [IT00]) is used to assess the performance metrics of the VoIP experiments, while the Peak-Signal-to-Noise-Ratio (PSNR) is used to assess the performance of VIP experiments. Using these assessment methods, the simulation results are mapped onto realistic network figures, namely the R-factor and the Mean Opinion Score (MOS), commonly used in real-life network planning.

The analytical model, on the other hand, examines the signaling costs involved during the execution of a handoff. The total signaling cost is calculated as the sum of the mobility signaling cost (binding updates and the associated acknowledgments), and the QoS signaling cost (re-establishing the reservations on the new link). The costs are further subclassified into transmission costs, processing costs, and buffering costs. The effect of the number of mobile nodes is examined, in addition to the effect of the average residence time in a subnet.

Although other models exist (such as network queuing models with blocking), the simulation and analytical models were selected for this research as they provide an overall indication of application and network-level performance. Network queuing models, on the other hand, are used to analyse the internal mechanics of packet scheduler implemented in routers. Therefore such models are more appropriate for research focusing on network queuing algorithms (e.g. Low Latency Queueing or Weighted Fair Queuing), as opposed to network control protocols (e.g. Point-to-

Point Protocol or Resource Reservation Protocol).

The remainder of this chapter is organised in the following manner: The next section outlines the methodology used to investigate the application-level performance of RSVP, including simulation environment, network topology and traffic characteristics. This is followed by a description of the experimental procedures used to quantitatively measure the performance of the two types of traffic examined (Section 3.2.2). The results are then presented in Section 3.2.3, followed by analysis and discussion.

In order to investigate the network-level performance, signaling cost models are formulated in Section 3.3. Numerical results are obtained by assigning parameter values to the derived models and are presented in Section 3.3.3, followed by a brief discussion. The chapter then concludes with a comprehensive comparison and analysis of the results obtained from both application- and network-level perspectives.

3.2 Application-Level Performance

3.2.1 Methodology and Implementation

The aim of the simulation model is to model RSVP behaviour over wireless networks as accurately as possible. This is accomplished by examining the application's performance as observed by a single mobile node experiencing different congestion levels in the network. With this in mind, the network topology shown in Figure 3.1 was chosen. This topology depicts a typical Mobile IPv6 deployment configuration in a simplified form and has been used extensively in various earlier studies [HS02, SCMB05].

The Network Simulator 2 [UCB] (version ns-2.26 patched with RSVP, HMIPv6 and FMIPv6 extensions [Mur, Wid, Hsi]) was used for the experiments. The RSVP model was further extended to implement reservations in wireless scenarios. For HMIPv6, RSVP Operation over IP Tunnels [TKWZ00] was used to establish a local

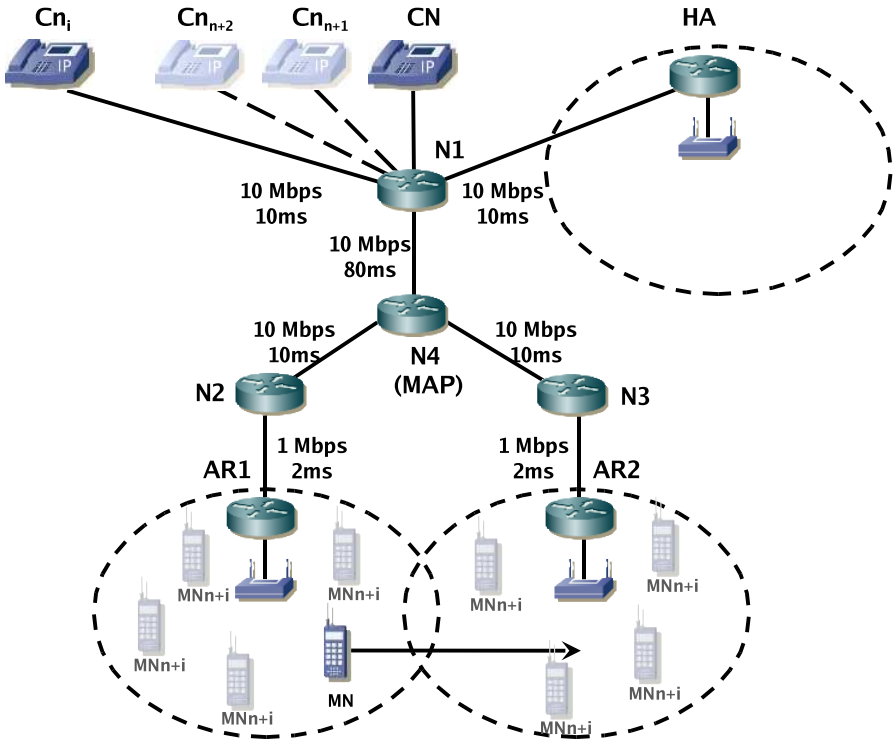

Figure 3.1: Simulation Model.

RSVP session across the tunnel between the Mobile Node (MN) and its Mobility Anchor Point (MAP). The Weighted Fair Queuing (WFQ) mechanism was used to manage all packet queues. Route Optimization was also implemented to avoid triangular routing of packets from the Correspondent Node (CN) to the MN via the Home Agent (HA).

The simulation scenario is comprised of a CN and a HA connected to a central switch node (N_1). The link from N_1 to the N_4 models an Internet backbone connection. In the case of HMIPv6, the MAP functionality is implemented at N_4 to manage local mobility signaling in the foreign network which consists of Access Points 1 and 2 connected to the N_4 via nodes N_2 and N_3 respectively.

As can be observed in Figure 3.1, the bottlenecks occur in the “last mile” at the foreign network (links N_2-AR_1 and N_3-AR_2) where the link capacity is 1 Mbps. Therefore, in order to increase the network load, contention should be created at

these specific links. For example, for a 10% network load, background traffic of 100 kbps is created at both N_2-AR_1 and N_3-AR_2 (to ensure that the target MN experiences the same congestion level as it handoffs from one access point to the other). This is done by adding a pair of CNs (CN_{n+1}, CN_{n+2}) which transmit data at 100 kbps to their corresponding MNs (MN_{n+1}, MN_{n+2}); located at AR_1 and AR_2 respectively. Similarly, for every increase of 10% network load, another pair of CNs/MNs is added until there are finally 20 CN/MN pairs for both AR1 and AR2, resulting in a network load of 200%. This implementation provides a more realistic approach than using a single CN/MN pair in which the data rate is simply increased from 100 kbps to 2 Mbps in steps of 100 kbps. Moreover, the background traffic is split into a combination of 70% VoIP traffic and 30% VIP traffic. According to [FCX+03] the average number of voice calls per day for the mass market segment is 1.768, compared with 0.679 for video calls which results in an average mixture of 70% voice traffic and 30% video traffic.

Table 1 summarizes the simulation parameters used in addition to the specific traffic parameters. For the wireless nodes, a 914 MHz Lucent WaveLan DSSS Card running the Wireless LAN 802.11 protocol was simulated with a transmission range of 100 m. Two types of traffic are considered: Voice-over-IP (VoIP) and Video-over-IP (VIP). The VoIP source is modeled as a 2-state "on-off" Markov chain. The alternate periods of activity (on) and silence (off) are exponentially distributed with average durations of 1.004 s and 1.587 s respectively. As recommended by the ITU-T recommendation P.59 [IT03] for conversational speech, the average activity cycle can be modeled as 38.53%. During talk spurts Constant Bit Rate (CBR) data stream is generated with a packet of size 220 bytes (Figure 3.2a), which is transmitted using RTP (Real-time Transport Protocol) over UDP. For the VoIP model, a G.711 codec is assumed, which requires a total bandwidth of 88 kbps. The VoIP packet format consists of a payload size of 160 bytes and a total header of 60 bytes (12 bytes RTP + 8 UDP + 40 bytes IPv6 header). Therefore the VoIP packet size is 220 bytes total.

For the VIP traffic source, an MPEG-4 encoded video sequence in Quarter Com-

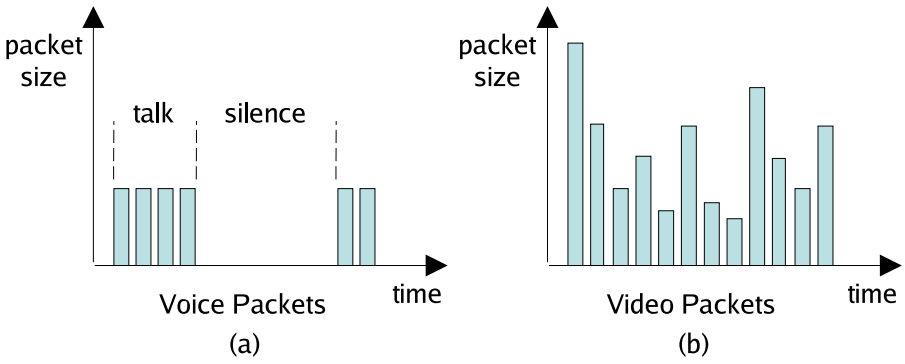

Figure 3.2: Voice and video packet generation sequence.

mon Intermediate Format (QCIF) was used. QCIF is a videoconferencing format that specifies data rates of 30 frames per second (fps) with each frame containing 144 lines and 176 pixels per line. QCIF was chosen because its support is required by the ITU-T H.263 [IT05] videoconferencing standard and is widely used by the 3rd Generation Partnership Project (3GPP) for video-capable mobile devices. MPEG-4 on the other hand, is a widespread video coding technique developed to target the low bit rates of Internet video. It achieves this through intra-frame and inter-frame compression. Intra-frame compression is done strictly within the same single video frame, while the latter compresses the temporal redundancies that typically occur between successive video frames. An MPEG sequence consists of three kinds of frames: I, P, and B-frames. An I-frame (intra-frame) is an intra-coded compression of a single frame and can be reconstructed without any reference to other frames.

A P-frame (predictive) is a frame constructed using information from the previous I-frame, while a B-frame is a “bidirectional predicted” frame depending on the previous or following I- or P-frames. Furthermore, MPEG-4 frames are arranged in Group of Pictures (GOP). A GOP consists of exactly one I-frame and some related P-frames and optionally some B-frames between these I- and P-frames (see Figure 3.3). Given the fact that an I-frame contains the most information, losing it would cause a “ripple-effect” distortion of all the following frames in a GOP. A P-frame

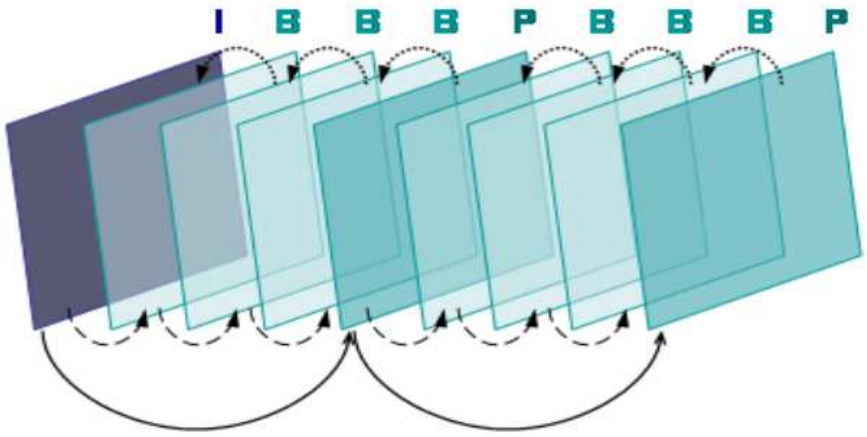

Figure 3.3: MPEG-4 GOP structure and its dependencies.

loss however, would only influence the adjoining B-frames while the loss of a B-frame would not influence any other frame. For the presented simulation study, the NCTU MPEG-4 codec [Nat]was used. Each VIP packet has a maximum payload of 1000 bytes with a 60 byte total header (12 RTP + 8 UDP + 40 IPv6) resulting in a maximum packet size of 1060 bytes. It is assumed that a dataflow already exists between the different CNs and their corresponding MNs. As for the target MN, an active RSVP session is assumed to be already in place before the handoff occurs. Finally, all reported results are based on averages taken from 20 simulation runs initiated with different random seeds.

3.2.2 Experimental Procedures for Methodology Assessment

Assessment of VoIP Quality Using the E-Model

Since voice quality is typically assessed through a listener's subjective perception, the Mean Opinion Score (MOS) [IT96] test has become the de facto standard of analysing the performance of VoIP systems. In these tests, listeners grade the perceived quality where "excellent" quality is given a score of 5, "good" a 4, "fair" a 3, "poor" a 2, and "bad" a 1. An arithmetic average is then computed to produce a final

Table 3.1: R-Factor to MOS Mapping.

$R-factor$	$QoSAssessment$	MOS
$90 < R < 100$	Excellent	4.34 - 4.5
$80 < R < 90$	Very Good	4.03 - 4.34
$70 < R < 80$	Good	3.60 - 4.03
$60 < R < 70$	Fair	3.10 - 3.60
$50 < R < 60$	Poor	2.58 - 3.10
$0 < R < 50$	Bad	1.0 - 2.58

MOS score. As one can observe, the MOS test is both time consuming and expensive. Instead, a computational model, called the E-Model ITU-T Recommendation G.107 [IT00] can be utilized to estimate MOS value. The E-model was designed to be used for transmission and QoS planning, and its output is calculated as a single quantitative figure, called the "R factor," using various impairment factors including packet delay and loss. Once the value of R- factor value is derived, it can be mapped to an estimated MOS value (Table 3.1). This method is often used in planning and predicting the performance of VoIP systems. The R-factor is expressed as:

$$R = R_o - I_s - I_d - I_e + A, \tag{3.1}$$

where R_o represents the signal-to-noise ratio (SNR), and I_d and I_e represent equipment delay and packet loss impairment factors, respectively. The value of I_d is codec dependent while I_e depends on loss patterns such as random or bursty. The Advantage Factor (A) is used to represent the convenience to the user of being able to make the phone call, for example a mobile phone is convenient to use therefore people are more forgiving on quality. Since it is a very subjective term, quantifying the advantage factor is a non-trivial task. Although ITU-T Recommendation G.701 recommends a value of 0 for conventional PSTN telephones and a value of 5 for cellular networks, no such agreement has been reached with regards to

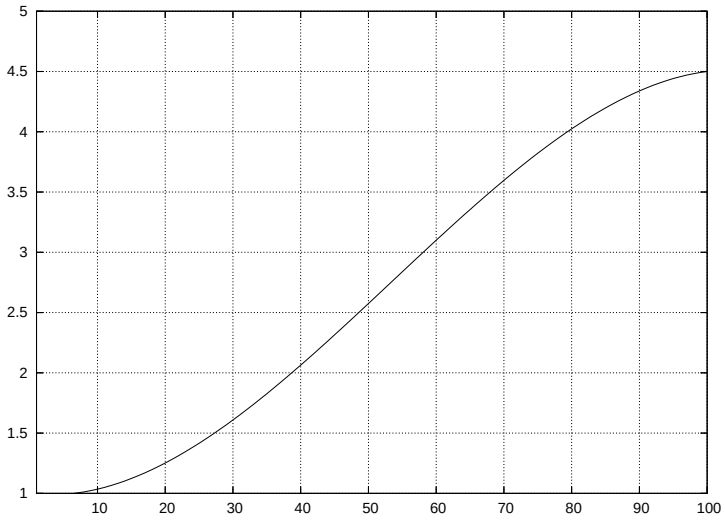

Figure 3.4: MOS as a function of R-factor.

wireless VoIP services. One could argue that there is little convenience in using wireless VoIP due to its limited coverage area compared to conventional cellular networks, which implies that there exists a similar geographical limitation to that of conventional PSTN telephones (especially when considering cordless phones). Consequently, the advantage factor is set to the default value of 0 for our calculations. I_s, on the other hand, represents the signal-to-noise impairment factor and is a function of several parameters which are independent of the underlying transport protocol. Since equipment performance is not the focus of the study but rather the transport protocol, the set of default values for these parameters as recommended by ITU-T Recommendation G.107 have been used. Choosing these default values, the R-factor (Equation 3.1) can be simplified to:

$$R = 94.2 - (I_d + I_e). \tag{3.2}$$

Calculating I_d and I_e is a lengthy process which involves a range of quantities [IT00]. For the simulation analysis, a method suggested by [CR01] was used to estimate I_d and I_e values using packet latency and loss. Once the R-factor is computed

using Equation 3.2, the corresponding MOS value can then be derived as follows [IT00]:

$$MOS = 1 + 0.035R + R(R-60)(100-R)7.10^{-6}. \tag{3.3}$$

This relationship between MOS and the R-factor is depicted in Figure 3.4. As can be observed, the minimum value for MOS is 1 and the maximum is 4.5 (due to codec imperfections). Once the values of R-factor and MOS have been computed, Table 3.1 can then be used to map them to the matching subjective QoS assessment as perceived by a human listener.

Assessment of VIP Quality using Peak Signal-to-Noise Ratio

As is the case with VoIP, various objective metrics [WP02] have been developed to estimate the MOS of VIP systems. The most widespread method is the calculation of the peak signal-to-noise ratio (PSNR) of individual frames. PSNR is a derivative of the well-known SNR [HCS01]. The definition of the PSNR of a source image (s) and a distorted image (d) is given as:

$$\mathrm{PSNR}(s,d) = 20\log\frac{V_{peak}}{\mathrm{MSE}(s,d)}[dB], \tag{3.4}$$

where V_{peak} is the maximum possible signal energy. In the case of a video transmission, this equates to, (2*k* - 1) where *k* is the bit colour depth. In the simulations presented, an 8-bit colour depth was used and therefore V_{peak} is equal to 255. $\mathrm{MSE}(s,d)$ is the Mean Square Error of (s) and (d):

$$\mathrm{MSE}(s,d) = \sqrt{\frac{1}{N_{col}N_{row}}\sum_{i=0}^{N}\sum_{j=0}^{N}[S(n,i,j)-D(n,i,j)]^2}. \tag{3.5}$$

The PSNR values of all individual images are then averaged to produce the mean PSNR of the complete video sequence; which is then mapped to the corresponding MOS value by using Table 3.2.

Table 3.2: PSNR to MOS Mapping.

$PSNR[dB]$	MOS
> 37	5 (Excellent)
$31 - 37$	4 (Good)
$25 - 31$	3 (Fair)
$20 - 25$	2 (Poor)
< 20	1 (Bad)

Mobility Impacts

An RSVP flow is identified by the 5-tuple (Source IP Address, Destination IP Address, Protocol ID, Source Port Number and Destination Port Number) [BZB+97]. Since a MN changes its IP address after a handoff, the data flow is no longer identified correctly by the intermediate RSVP-enabled routers existing reservations become invalid. As a result, a new reservation needs to be established. In the case that the MN is the sender, it would have to send a new Path message (with the new IP address) to the CN which will reply with a Resv message to activate the new reservation. However, if the MN is the receiver, it can not simply send a new Resv message since there would be no corresponding Path-state established for it at the intermediate routers (the existing path-state would still be pointing to the old CoA). If default RSVP behaviour is used, a MN would typically have to wait anywhere up to 30 seconds before receiving the Path message periodically sent from the CN (default value to refresh an RSVP session is 30 seconds). This is clearly unacceptable and thus it is assumed that a signal generated by the MIP module triggers the RSVP module at the CN in the following manner: Upon receiving a BU, a CN will use this information to immediately send a new Path message to the MN's new CoA. As one can observe, the total service disruption is at least two round trips (one for exchanging BU/Back and another for Path/Resv message pairs).

Another key issue is the behaviour of the sender during this disruption in ser-

Figure 3.5: VIP transmission using RSVP (left image) and best-effort (right).

vice. Once again, if the default RSVP behaviour is used, the sender would treat the reservation as a new RSVP session and would have to wait to receive the Resv message from the receiver before transmitting data. Two cases are considered: One where the sender follows default RSVP behaviour and stops transmitting data while the reservation is being re-established; and another where the sender immediately resumes sending data upon receiving a BU. This requires some degree of correlation between RSVP and Mobile IP which is feasible since it only requires minimal change at the end node (triggering a Path message upon reception of a BU from the mobility routing protocol). For analysis purposes, the first is called *"RSVP Default Flow"* (RSVP-DF) and the later *"RSVP Continuous Flow"* (RSVP-CF). The data flow of RSVP-CF is treated as Best-Effort (BE) traffic by the intermediate routers until the RSVP session is re-established.

3.2.3 Results and Observations

RSVP performance for fixed nodes

The performance of RSVP over MIPv6, HMIPv6 and FMIPv6 was tested under varying congestion levels while monitoring the effect on VoIP and VIP application-level performance. In order to ensure that RSVP performs accurately, the packet

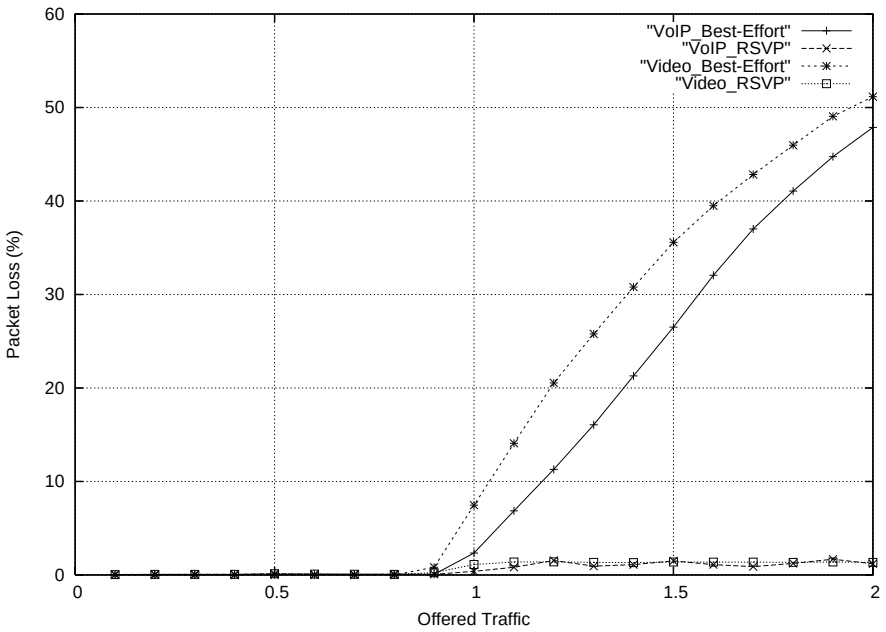

Figure 3.6: The effect of RSVP on packet loss. The results were obtained by changing the background traffic density to increase the offered link load from 0% to 200% of the link capacity.

loss of RSVP-enabled traffic flows was measured and compared against best-effort traffic. As observed in Figure 3.6, when the offered traffic is less than 90% of the link capacity, the packet loss for both RSVP and best-effort is zero (whether using VoIP or VIP traffic). Beyond this point however, the prioritization of RSVP starts to take effect, confining the RSVP-VoIP and RSVP-VIP packet loss to about 1.3%.

In contrast, best-effort traffic packet losses increase in proportion to the increased link load and roughly 50% of packets are discarded when the offered link load reaches 200% of the capacity. Interestingly, in high levels of network congestion Best-Effort VIP traffic suffers higher packet losses than Best-Effort VoIP. This is due to two reasons: (i) VIP packets have a larger packet size of 1060 bytes, as opposed to 220 bytes for VoIP, (ii) the WFQ buffer size is "per byte" rather than "per packet" for more accurate operation of the fair queuing algorithm. Moreover, during high congestion the routers buffer incoming packets into their WFQ queues

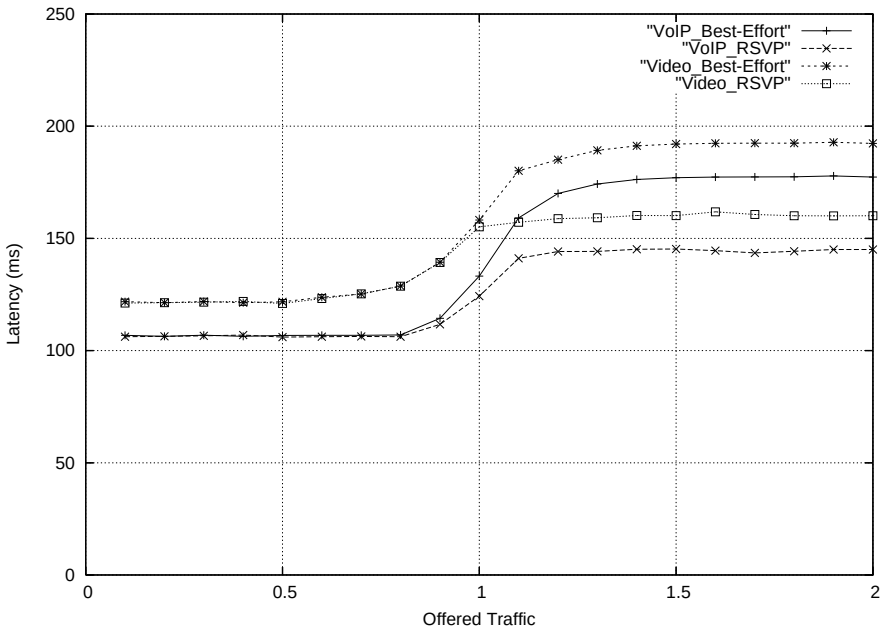

Figure 3.7: Effect of RSVP on latency under varying network loads.

and as more packets arrive, the remaining available buffer space is reduced. At the critical point where the remaining buffer size is less than 1060 bytes, an arriving VIP packet would be discarded whereas a VoIP packet would be admitted into the queue. As a result, the large VIP packet has a higher chance of being dropped than a VoIP packet.

In Figure 3.7, a similar effect on end-to-end packet latency is noted: In low congestion levels, latency of VoIP packets is around 106 ms while VIP packet latency is higher at 121 ms (due to the higher packet size). In high congestion levels, RSVP-VoIP maintains packet latency at 147 ms while best effort traffic reaches 177 ms. Similarly, RSVP-VIP maintains a packet latency of 160 ms while best effort VIP traffic reaches 192 ms. RSVP's ability to maintain shorter packet latency during high congestion is a direct consequence of the WFQ mechanism which schedules RSVP traffic to the front of the queue in order to reduce the queuing time.

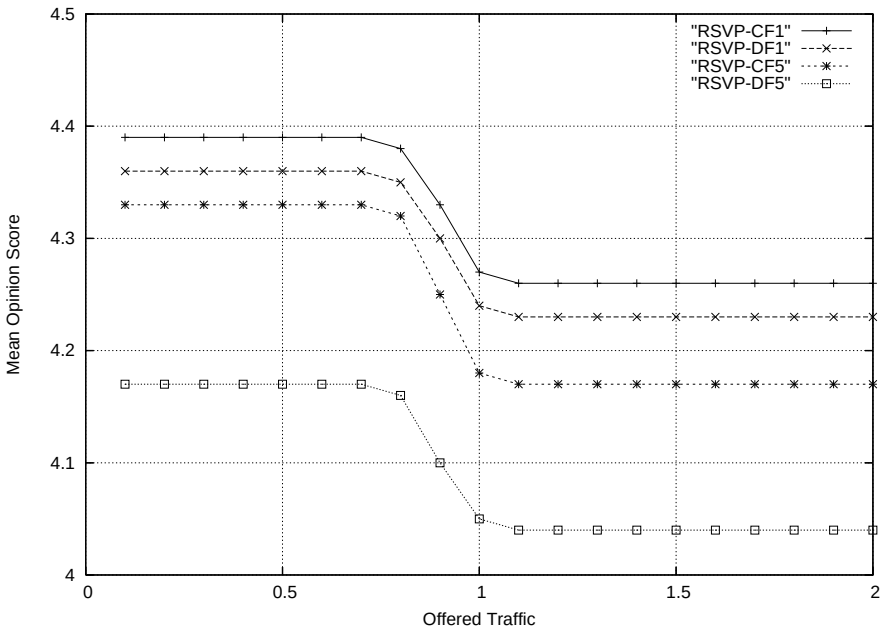

Figure 3.8: Impact of the number of handoffs.

Impact of RSVP behaviour during handoffs

Figure 3.8 presents the impact of number of handoffs on the Mean Opinion Score (MOS) of *"RSVP Default Flow"* (RSVP-DF) and *"RSVP Continuous Flow"* (RSVP-CF). When only one handoff occurs, RSVP-CF1 slightly outperforms RSVP-DF1. This is expected since even though both experience an equal amount of disruption time due to the MIPv6 handoff latency, they behave differently once the handoff is completed: During the reservation re-establishment time (2RTT + processing time at every router) RSVP-CF continues to send data (receiving best effort treatment) while RSVP-DF does not. As the number of handoffs increases to 5, the quality of RSVP-DF5 degrades even further (compared to RSVP-CF5). From a user's perspective, it is more favourable to temporarily receive slightly lower quality of service than undergo abrupt silence periods during handoffs.

Impact of Voice-over-IP Traffic

This subsection presents a study of the performance of RSVP-enabled VoIP traffic flows in wireless IP networks using different mobility protocols (MIPv6, HMIPv6 and FMIPv6). Handoffs are introduced and the offered link traffic is varied to examine the effect on voice quality. When considering a handoff in HMIPv6, mobility signaling is exchanged with the MAP (node N4), rather than the CN which is several hops away. Moreover, because RSVP Operation over IP Tunnels is implemented, only the local RSVP session between the MN and its MAP is re-established. This means that HMIPv6 generally yields better performance than MIPv6 in almost all applications.

One exception, however, is observed in this study: When the offered traffic is increased beyond 80% of link capacity, HMIPv6 performs slightly worse than MIPv6 (Figure 3.9). Given that a typical VoIP packet is relatively small in size, the IP-in-IP encapsulation introduced by HMIPv6 at the MAP adds an extra 40 byte IPv6 header. This increases the packet size to a total of 260 bytes which translates to an 18% increase. Moreover, the "per byte" treatment of the WFQ buffer results in slightly higher packet loss and end-to-end packet latency than MIPv6. Since the E-Model computation depends primarily on packet delay and loss, the R-factor decreases and so does the corresponding mapped MOS value. However, as the number of handoffs increases, HMIPv6 starts to outperform MIPv6 (Figure 3.10). This is largely attributed to the reduced handoff latency of HMIPv6 which results in fewer lost packets and hence a higher MOS value.

FMIPv6 performance on the other hand, surpasses both MIPv6 and HMIPv6 (Figures 3.9 and 3.10). The underlying reason for this is twofold: FMIPv6 anticipates a handoff through Layer-2 triggers and prepares for it in advance. This reduces the handoff latency since no address resolution time is required (the new CoA has already been negotiated between oAR and nAR). The other reason relates to the Layer 3 registration process (when the MN sends a BU to the HA and another

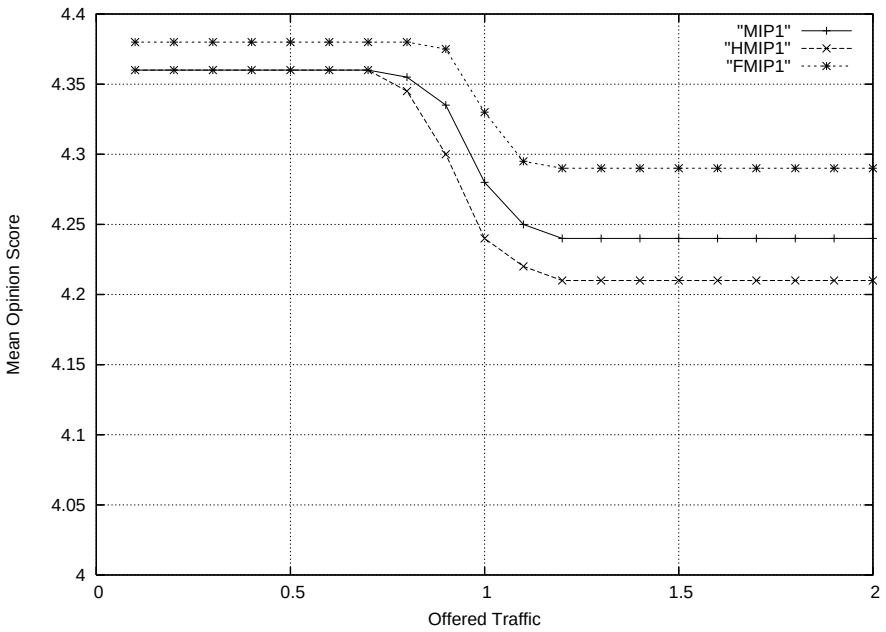

Figure 3.9: VoIP performance at 1 handoff under varying offered link loads.

to the CN). Any data destined to the MN using the old CoA is forwarded by the oAR to the MN's new location. This technique proves to be very effective in terms of packet loss. Since FMIPv6 cuts down both disruption time and packet loss, it performs best at frequent handoffs (Figure 3.10).

Impact of Video-over-IP Traffic

In this subsection the results of the experiments are presented which are similar to the ones discussed in the previous section but were repeated using a VIP traffic source instead. As can be observed by comparing Figures 3.11 and 3.12 against Figures 3.9 and 3.10, the performance of VIP is worse overall than that of VoIP. To help understand this, a closer analysis of the different phases that a MN goes through during a handoff is required: The MN has to first establish a connection on the new link (mobility signaling), and then reserve the necessary resources (RSVP signaling). During the first phase, the MN is unreachable and any packets destined

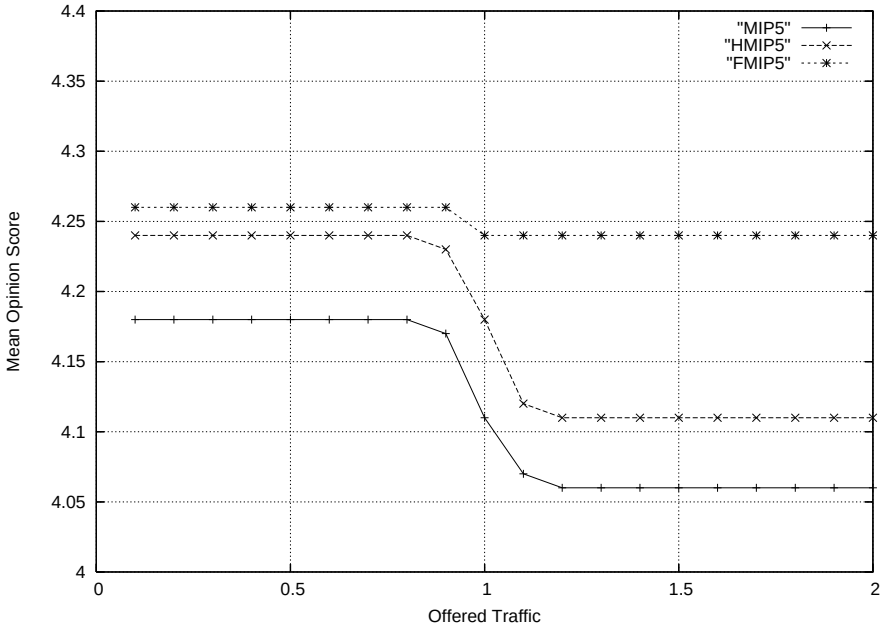

Figure 3.10: VoIP performance at 5 handoffs under varying offered link loads.

to it are dropped. Once mobility signaling is completed (i.e. a new CoA has been acquired and the HA/CN updated accordingly), packets can be delivered to the MN at its new location. However, since the MN is now using a new CoA, the old reservations (on the unchanged portion of the end-to-end path) are no longer valid for the data flow. As a result, any packets destined to the MN receive best-effort treatment until RSVP signaling is completed. An important distinction between the two phases is that during the first, packets are lost regardless of network congestion (MN is actually disconnected) while in the latter, packet loss depends on congestion level and traffic type due to best- effort treatment (Figure 3.6). With this in mind, the following observations are noted:

Single handoff, in periods of high congestion: VoIP MOS drops to 4.22 while VIP MOS drops to 3.92 (Figures 3.9 and 3.11): The primary reason for this is the best-effort treatment received during RSVP signaling after a handoff. In high congestion, VIP packets suffer a higher packet loss and latency than VoIP (Figures 3.6 and 3.7). This is reflected on the calculated PSNR values, thereby resulting in

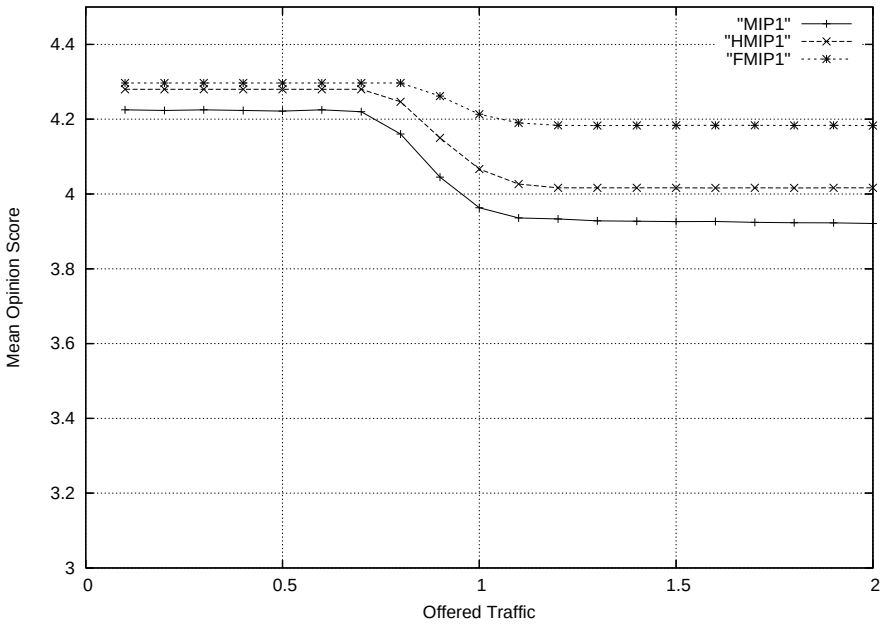

Figure 3.11: VIP performance at 1 handoff under varying offered link loads.

the observed lower VIP MOS.

Five handoffs, in periods of low congestion: VoIP MOS is 4.26 while VIP MOS is 4.05 (Figures 3.10 and 3.12): Since this occurs when congestion level is low and ample bandwidth is available, the best-effort treatment received during RSVP signaling does not have an effect (packet loss for best-effort VIP is approximately 1.3% in periods of low congestion). RSVP signaling is therefore ruled out as a contributing factor while mobility signaling becomes the dominant one. Since the VIP traffic source adheres to H.263 QCIF standards, it produces a frame rate of 30 fps. This translates to an average data rate of approximately 200 kbps. When compared to VoIP (88 kbps for our G.711 codec), VIP traffic would lose a larger amount of data than VoIP for the same period of time. Moreover, the VoIP model is bursty with ”on” and ”off” cycles while the VIP traffic source is constantly ”on”. Naturally there is a smaller probability of a VoIP packet being dropped than a VIP packet.

Five handoffs, in periods of high congestion: VIP MOS for MIPv6 and HMIPv6

drops at a steeper rate (Figure 3.12) than FMIPv6: When compared against Figure 3.11, it is apparent that MIPv6 and HMIPv6 have reached some degree of a break point where VIP performance deteriorates sharply in periods of high congestion. Since this slump in MOS occurs only during high congestion, it is safe to assume that it is a consequence of best-effort treatment after a handoff. To understand why this happens, the structure and dependencies of an MPEG-4 GOP have to be taken into account. As discussed in Section 3.2.2, an I-frame loss would cause a "ripple effect" loss on the following frames of a GOP (see Figure 3.3). In the simulation study, an I-frame was constructed typically every 300 ms, while RSVP signaling took one round trip between the MN and CN. This results in data traffic undergoing best-effort treatment for an average time of 204 ms for every handoff. Therefore, the probability of exactly one I-frame being treated as best-effort traffic is 68% (204 ms / 300 ms). Even if only one I-frame was treated as best- effort traffic, the probability of dropping it depends on the congestion level (for example, packet loss is 14% when offered traffic is 110% and 51% when it reaches double the link capacity). As a result, the total probability of losing an I-frame during one handoff when offered traffic is 9.52% (0.68 x 0.14) when offered traffic is 110% of the link capacity, and 34.68% (0.68 x 0.51) for an offered traffic of double link capacity. As the number of handoffs is increased to five, this probability increases dramatically and losing an I-frame becomes practically inevitable. An I-frame loss results in additional losses of the subsequent frames in the GOP, hence resulting in the observed slump in MOS.

VIP traffic and FMIPv6: FMIPv6 performs best with VIP traffic (Figures 3.11 and 3.12): When a handoff occurs in the particular implementation of FMIPv6 used in this study, the two ARs establish a local RSVP session over the temporary forwarding tunnel between them. This technique not only reduces packet loss through tunnelling but also maintains the QoS level. This arrangement results in FMIPv6's superior performance in all network conditions.

VIP traffic and HMIPv6: HMIPv6 outperforms MIPv6 (Figure 3.9): In Section 3.2.3 at one handoff using VoIP, MIPv6 was found to be surprisingly a better can-

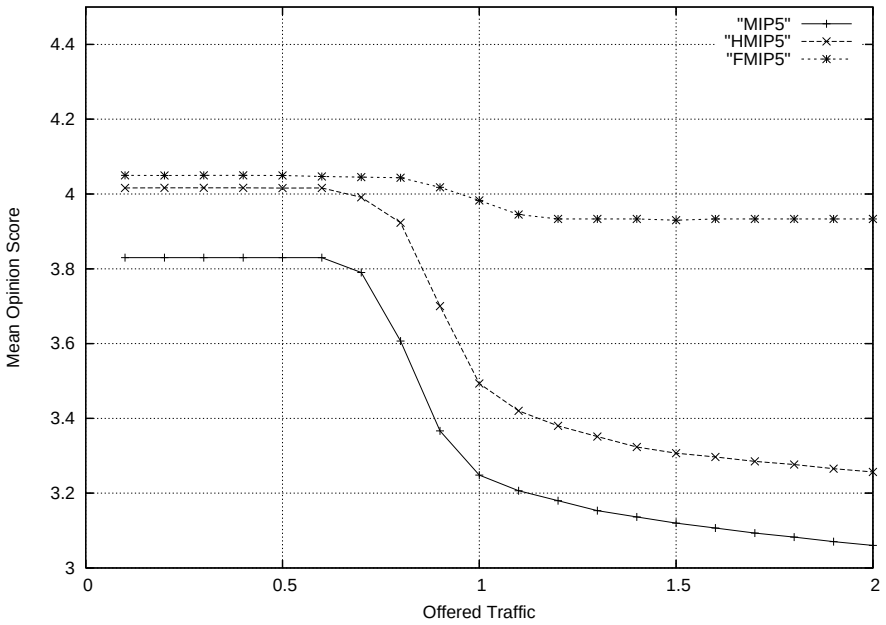

Figure 3.12: VIP performance at 5 handoffs under varying offered link loads.

didate than HMIPv6 (Figure 3.9). With video traffic however, this is not the case. Due to the larger size of a VIP packet (maximum of 1060 bytes), the added 40 bytes IP-in-IP encapsulation of the MAP in HMIPv6 results in a 3.8% increase (40/1060) as opposed to 18% (40/220) for VoIP traffic. This slight increase in packet size does not create a significant effect on VIP performance. Moreover, such a minor hindrance is easily subdued by HMIPv6's shorter handoff latency. Hence HMIPv6 can be asserted as a better candidate for video traffic than MIPv6.

3.2.4 Analysis and Discussion

A set of simulation based experiments were used to assess the application-level performance of RSVP in wireless IP networks using real-time traffic. VIP traffic was found to be more susceptible to network congestion than what had been previously speculated. Although VIP's larger packet size relates to its reduced performance, the frame dependencies of MPEG-4 remain the pivotal factor in truly

realising higher quality video communications over wireless IP networks. In high congestion and high mobility scenarios, the I-frame dependencies of a GOP could inflict a significant deterioration in VIP performance (thereby dropping the perceived QoS assessment from "Excellent" to a borderline "Fair"). A viable solution would be to incorporate a selective dropping mechanism in RSVP-enabled routers that distinguishes the different frame types of an MPEG-4 data stream and assigns levels of priority accordingly: An I-frame would be given the highest priority while a B-frame the lowest. This prioritised treatment of MPEG-4 data streams should not only be applicable to RSVP flows but more importantly to best-effort flows (since RSVP flows inevitably go through a best-effort treatment phase after a handoff).

In terms of mobility protocols, however, FMIPv6 performs best under all network circumstances. Having said that, the following consideration needs to be taken into account: In the presented study the two ARs establish a local RSVP session over the temporary tunnel between them, hence maintaining a stable QoS level during the execution of a Layer 3 handoff. This did not have a notable impact on the other MNs since they generated best-effort traffic to create contention at the access network. Therefore, although FMIPv6 is beneficial from the target MN's point-of-view, it could have a substantial effect on other MNs utilizing RSVP flows in the same network. In severe cases (high congestion and mobility) these temporarily established reservations can block new RSVP sessions from initiating (resulting in a higher call blocking probability).

An interesting behaviour was observed for HMIPv6: In the case of VoIP traffic, MIPv6 surprisingly proves to be a better choice than HMIPv6 (owing to its smaller packet size). Nonetheless, this advantage is quickly surpassed by HMIPv6 in high mobility scenarios due to its smaller handoff latency which results in fewer packets being dropped. In the case of VIP traffic, HMIPv6 proves to be a better choice than MIPv6 in all network circumstances. A feasible solution to further enhance the HMIPv6 performance would be to implement IP header compression. This will resolve HMIPv6's degradation in quality due to the increased packet size while at

the same time benefiting from its lower handoff latency.

3.3 Network-Level Signaling Cost

This Section complements the investigation of the application-level performance presented in Section 3.2. Here, an analytical model devised to assess the signaling costs incurred on the network is presented. The analysis using this model focuses on handoff scenarios and takes into account the mobility signaling required to re-connect a session after a handoff, in addition to the RSVP signaling required to re-establish the necessary QoS levels in the upstream and downstream directions.

Although the analysis presented in this section follows a logic which is consistent with previous work [FAML05, XA02], it differs in the following manner:

1. The study focuses on signaling invoked due to a handoff and is therefore not concerned with data packet delivery cost.

2. It is assumed that an active session exists between the MN and its CN when the handoff occurs.

3. The CN is assumed to have a direct path of communication with its MN (Route Optimisation).

4. Processing costs at the HA, MAP and other nodes have been normalised by using a single universal processing cost (γ).

3.3.1 Signaling Cost Analysis

In this subsection, the individual signaling costs of the following protocols are derived: MIPv6, HMIPv6, FMIPv6 and RSVP.

Mobile IPv6

The following notations [FAML05, XA02], are used to derive the signaling cost of Mobile IPv6:

Ψ^{MIP}	Mobile IPv6 signaling cost.
R_{mh}, T_{mh}, P_{mh}	Registration, Transmission and Processing costs (MN $\leftrightarrow$ HA).
R_{mc}, T_{mc}, P_{mc}	Registration, Transmission and Processing costs (MN $\leftrightarrow$ CN).
l_{mc}	Average distance between MN and CN in hops.
l_{mh}	Average distance between MN and HA in hops.
δ_{B}	Per-hop binding update transmission cost.
γ	Processing cost at a node.
N	Total number of mobile nodes.
tr	MN residence time in a subnet.

At every subnet crossing (occurring at tr), a MN will register its new CoA at the HA and hence incur a registration cost (R_{mh}). Moreover, since a data session is already in progress with the CN when the handoff occurred, the MN has to also register its new CoA with the CN (R_{mc}). Therefore the average MIPv6 signaling cost can be estimated as the number of MNs multiplied by the registration costs with the HA and MN, divided by the average subnet residence time:

$$\Psi^{MIP} = N\frac{R_{mh} + R_{mc}}{tr}. \tag{3.6}$$

The registration cost (R) can be further broken down into the transmission cost (T) and the processing cost (P) as follows:

$$\begin{aligned} R_{mh} &= T_{mh} + P_{mh}, \\ R_{mc} &= T_{mc} + P_{mc}. \end{aligned} \tag{3.7}$$

Rather than calculate the transmission cost as simply the number of hops (l_{mh}) multiplied by the per-hop BU transmission cost (δ_{B}) , the following is taken into account: Due to the nature of wireless links (MAC contentions and frame retransmissions), the transmission cost of a wireless hop is naturally higher than that of a wired hop. As a result, a proportionality constant (ρ) is used to denote this effect.

Furthermore, when considering network-level performance, the processing costs at the end points (MN and CN) should not be included in the analysis as they are incurred on user terminals and do not directly contribute to the overall network load. Therefore P_{mh} results in a single processing cost (γ) at the HA, while P_{mc} effectively equates to zero:

$$\begin{aligned} T_{mh} &= 2(l_{mh} - 1 + \rho)\delta_B, & P_{mh} &= \gamma, \\ T_{mc} &= 2(l_{mc} - 1 + \rho)\delta_B, & P_{mc} &= 0, \end{aligned}$$

where $(l_{mh} - 1)$ and $(l_{mc} - 1)$ represent the number of wired hops from the MN to the HA and CN respectively. Note that the transmission cost equation is multiplied by a factor of two to represent the BU/BAck message pair. Substituting these values into Equations 3.7 and 3.6 yields the following:

$$\begin{aligned} \Psi^{MIP} &= N[\frac{2(l_{mh} - 1 + \rho)\delta_B + \gamma + 2(l_{mc} - 1 + \rho)\delta_B}{tr}] \\ &= N[\frac{2(l_{mh} + l_{mc} - 2 + 2\rho)\delta_B + \gamma}{tr}]. \end{aligned} \tag{3.8}$$

Hierarchical Mobile IPv6

Ψ^{HMIP}	Hierarchical Mobile IPv6 signaling cost.
R_{mm}, T_{mm}, P_{mm}	Registration, Transmission and Processing costs between the MN and MAP (MN $\leftrightarrow$ MAP).
l_{mm}	Average distance between MN and MAP in hops.
M	Average number of global handoffs (outside a MAP's domain).

In HMIPv6, a MN can either move into a subnet that lies within the domain of its current MAP (local handoff), or to a subnet serviced by another MAP (global handoff). A local handoff occurs every tr, and requires the MN to register its new Local CoA (LCoA) with the MAP (R_{mm}). Global handoffs, however, occur every ($M \times tr$) seconds, and the MN is required to register the new Regional CoA (RCoA) with its HA (R_{mh}) and its associated CN (R_{mc}). To calculate the total signaling cost of HMIPv6, the registration costs of local and global handoffs are combined as follows:

$$\Psi^{HMIP} = N[\frac{R_{mm}}{tr} + \frac{R_{mh} + R_{mc}}{M \times tr}]. \tag{3.9}$$

For the registration cost to the MAP (R_{mm}), a single processing cost is incurred at the MAP ($P_{mm} = \gamma$) in order to process the local BU. For the registration cost with the HA (R_{mh}), however, there are three processing costs incurred: one at the HA to process the RCoA, and two at the MAP to process the RCoA and a new LCoA ($P_{mh} = 3\gamma$). For the registration cost to the CN (R_{mc}), the processing cost at the CN is neglected as it does not directly contribute to the actual network cost. Therefore, for R_{mc}, only the transmission cost is taken into account whereas the processing cost is effectively zero:

$$\begin{aligned} T_{mm} &= 2(l_{mm} - 1 + \rho)\delta_B, & P_{mm} &= \gamma, \\ T_{mh} &= 2(l_{mh} - 1 + \rho)\delta_B, & P_{mh} &= 3\gamma, \\ T_{mc} &= 2(l_{mc} - 1 + \rho)\delta_B, & P_{mc} &= 0. \end{aligned}$$

Substituting these values into Equation 3.9 yields the following:

$$\begin{aligned} \Psi^{HMIP} &= N \times [\frac{2(l_{mm} - 1 + \rho)\delta_B + \gamma}{tr} + \frac{2(l_{mh} - 1 + \rho)\delta_B + 3\gamma \ + \ 2(l_{mc} - 1 + \rho)\delta_B}{M \times tr}] \\ &= N \times [\frac{2(l_{mm} - 1 + \rho)\delta_B + \gamma}{tr} + \frac{2(l_{mh} + l_{mc} - 2 + 2\rho)\delta_B + 3\gamma}{M \times tr}]. \end{aligned} \tag{3.10}$$

Fast Handovers for Mobile IPv6

Ψ^{FMIP}	Fast Handovers for Mobile IPv6 signaling cost.
Ψ^{FR}	Registration costs of the Fast Handovers mechanism.
Ψ^{FPT}	Packet Tunneling costs of the Fast Handovers mechanism.
R_{mo}, T_{mo}, P_{mo}	Registration, Transmission and Processing costs (MN $\leftrightarrow$ oAR).
R_{on}, T_{on}, P_{on}	Registration, Transmission and Processing costs (oAR $\leftrightarrow$ nAR).
R_{mn}, T_{mn}, P_{mn}	Registration, Transmission and Processing costs (MN $\leftrightarrow$ nAR).
δ_{D}	Per-hop data packet transmission cost.
β	Buffering cost at nAR.
P	Average number of packets dropped during a handoff.

FMIPv6's total signaling cost (Ψ^{FMIP}) can be broken down into three basic components as follows:

$$\Psi^{FMIP} = \Psi^{FR} + \Psi^{FPT} + \Psi^{MIP}, \tag{3.11}$$

where Ψ^{MIP} is simply the basic MIPv6 signaling cost (derived earlier in Equation 3.8). Ψ^{FR} on the other hand, is the additional registration costs specific to the Fast Handovers mechanism. By referring to Figure 2.6, it can be observed that FMIPv6 signaling occurs across three nodes: The MN, the old Access Router (oAR), and the new access Router (nAR). Therefore, Ψ^{FR} is accordingly comprised of three registration costs: registration cost between the MN and oAR (R_{mo}), MN and nAR (R_{mn}), and between the two access routers themselves (R_{on}). Since this occurs every tr, Ψ^{FR} is formulated as follows:

$$\Psi^{FR} = N[\frac{R_{mo} + R_{on} + R_{mn}}{tr}]. \tag{3.12}$$

FMIPv6 introduces four message types, and performs it signaling as illustrated in Figure 2.6. For (R_{mo}), four messages (RtSolPr, PrRtAdv, BU and BAck) are exchanged across a single wireless hop ($T_{mo} = 4\rho\ \delta_B$). The oAR incurs two process-

ing costs, one for the RtSolPr message and another for the BU message ($P_{mo} = 2\gamma$). Note that the processing cost of the PrRtAdv at the MN is ignored as explained earlier.

In a similar fashion, R_{on} includes the cost of transmitting three messages across wired hops between the oAR and the nAR ($T_{on} = 3l_{on}\ \delta_B$) and incurs two processing costs, one to process the HI message at the nAR and another to process the HAck message at the oAR ($P_{on} = 2\gamma$). Finally, when the MN moves into the new subnet (R_{mn}), it send an NA message to the nAR ($T_{mn} = \rho\ \delta_B$) and incurs a processing cost at the nAR ($P_{on} = \gamma$). The registration costs can therefore be summarized as follows:

$$\begin{aligned} T_{mo} &= 4\rho\ \delta_B, & P_{mo} &= 2\gamma, \\ T_{on} &= 3l_{on}\ \delta_B, & P_{on} &= 2\gamma, \\ T_{mn} &= \rho\ \delta_B, & P_{mn} &= \gamma. \end{aligned}$$

Substituting these values into Equation 3.12 yields the following:

$$\Psi^{FR} = N[\frac{(5\rho + 3l_{on})\delta_B + 5\gamma}{tr}]. \tag{3.13}$$

The packet tunneling cost (Ψ^{FPT}) however, includes the cost of transmitting the data packets from the oAR to the MN, via the nAR, which is multiplied by the Per-hop data packet transmission cost ($(l_{on} + \rho)\delta_D$). Two processing costs are incurred, one for the encapsulation at the oAR and another for the decapsulation at the nAR (2γ), in addition to the buffering cost at the nAR (β). This expression is multiplied by the average number of packets (P) that are received at the oAR during the MIPv6 handover. Therefore, Ψ^{FPT} can be formulated as follows:

$$\Psi^{FPT} = N[\frac{P[(l_{on} + \rho)\delta_D + 2\gamma + \beta]}{tr}]. \tag{3.14}$$

To obtain the total signaling cost of FMIPv6 (Ψ^{FMIP}), Equations 3.14, 3.13 and 3.8 are substituted into Equation 3.11 as follows:

$$\begin{aligned}\Psi^{FMIP} &= N[\frac{(5\rho + 3l_{on})\delta_B + 5\gamma}{tr}] + N[\frac{P[(l_{on} + \rho)\delta_D + 2\gamma + \beta]}{tr}] \\ &\quad + N[\frac{2(l_{mh} + l_{mc} - 2 + 2\rho)\delta_B + \gamma}{tr}], \end{aligned} \tag{3.15}$$

$$\begin{aligned}\Psi^{FMIP} &= N[\frac{2(l_{mh} + l_{mc} + 1.5l_{on} - 2 + 4.5\rho)\delta_B + 6\gamma}{tr}] \\ &\quad + N[\frac{P((l_{on} + \rho)\delta_D + 2\gamma + \beta)}{tr}]. \end{aligned} \tag{3.16}$$

Resource Reservation Protocol

Ψ^{RSVP}	Total RSVP signaling cost.
Rsv_{mc}	Resource reservation costs (MN $\leftrightarrow$ CN).
δ_R	Per-hop RSVP message transmission cost.

Similar to MIPv6, subnet crossings occur every tr and therefore reservations should be re-established between the MN and CN (Rsv_{mc}). Since we are considering full-duplex communication, Rsv_{mc} is multiplied by two to include the upstream and downstream RSVP sessions:

$$\Psi^{RSVP} = N\frac{2 \times Rsv_{mc}}{tr}, \tag{3.17}$$

where $Rsv_{mc} = T_{mc} + P_{mc}$.

Since an RSVP message is processed by all nodes along the way, it is necessary to find the exact number of nodes included in the session. For the (MN $\leftrightarrow$ CN) session there are l_{mc} hops in between (including wired and wireless hops). However, as discussed earlier, the processing cost at the end points (MN and CN) should not be included and therefore the processing cost for a single RSVP message becomes

[$(l_{mc} - 1)\ \gamma$]. This is simply multiplied by two for the Path/Resv message pair. The transmission on the other hand is the number of hops between the MN and CN, multiplied by the per-hop RSVP message transmission cost:

$$T_{mc} \quad = \quad 2(l_{mc} - 1 + \rho)\delta_R, \qquad\qquad P_{mc} \quad = \quad 2(l_{mc} - 1)\gamma.$$

Substituting the above into equation 3.17 yields the following:

$$\Psi^{RSVP} \quad = \quad N[\frac{4[(l_{mc} - 1 + \rho)\delta_R + (l_{mc} - 1)\gamma]}{tr}]. \tag{3.18}$$

3.3.2 Total Signaling Cost using RSVP

Using the equations from Section 3.3.1, we derive the total signaling cost of deploying RSVP over the different mobility protocols.

RSVP over Mobile IPv6

The signaling cost of deploying RSVP over Mobile IPv6 ($\Psi^{\text{MIP}}_{\text{RSVP}}$) is the sum of the Mobile IPv6 signaling cost (Ψ^{MIP}) and the Resource Reservation Protocol signaling cost (Ψ^{RSVP}):

$$\Psi^{MIP}_{RSVP} \quad = \quad \Psi^{MIP} \ + \ \Psi^{RSVP}, \tag{3.19}$$

where $\Psi^{\text{MIP}}$and $\Psi^{\text{RSVP}}$are formulated in Section 3.3.1 and Section 3.3.1 respectively. Substituting equations 3.8 and 3.18 into the above equation yields the following:

$$\begin{aligned} \Psi^{MIP}_{RSVP} \quad = \quad & N[\frac{2(l_{mh} + l_{mc} - 2 + 2\rho)\delta_B + \gamma}{tr}] \\ & + N[\frac{4[(l_{mc} - 1 + \rho)\delta_R + (l_{mc} - 1)\gamma]}{tr}]. \end{aligned} \tag{3.20}$$

RSVP over Hierarchical Mobile IPv6

Similar to the analysis presented in Section 3.3.2, we define $\Psi_{\text{RSVP}}^{\text{HMIP}}$ as the signaling cost of deploying RSVP over Hierarchical Mobile IPv6. The Hierarchical Mobile IPv6 signaling cost (Ψ^{HMIP}) is formulated in Section 3.3.1. However, our study focuses on local handoffs only (See Figure 3.1). Therefore, the second portion of equation 3.10 is omitted and hence Ψ^{HMIP} can be simplified to:

$$\Psi^{HMIP} = N[\frac{2(l_{mm} - 1 + \rho)\delta_B + \gamma}{tr}]. \tag{3.21}$$

Therefore, $\Psi_{\text{RSVP}}^{\text{HMIP}}$ is the sum of Equation 3.21 above and Equation 3.18:

$$\begin{aligned} \Psi_{RSVP}^{HMIP} &= \Psi^{HMIP} + \Psi^{RSVP} \\ &= N[\frac{2(l_{mm} - 1 + \rho)\delta_B + \gamma}{tr}] \\ &\quad + N[\frac{4[(l_{mm} - 1 + \rho)\delta_R + (l_{mm} - 1)\gamma]}{tr}]. \end{aligned} \tag{3.22}$$

RSVP over Fast Handovers for Mobile IPv6

To formulate the total signaling cost of deploying RSVP over FMIPv6 ($\Psi_{\text{RSVP}}^{\text{FMIP}}$), we simply add the FMIPv6 and RSVP signaling costs:

$$\Psi_{RSVP}^{FMIP} = \Psi^{FMIP} + \Psi^{RSVP}. \tag{3.23}$$

However, another signaling cost has to be taken into consideration, which is the temporary RSVP tunnel established between the oAR and nAR (as outlined in Section 3.2.4). Since this signaling occurs on the wired portion of the network, all transmission and processing costs involved are taken into account and hence the

additional RSVP signaling cost over the tunnel is given by:

$$\Psi_{tunnel}^{RSVP} = N[\frac{4l_{on}(\delta_R + \gamma)}{tr}]. \quad (3.24)$$

Adding equation 3.24 to 3.23 yields the following:

$$\begin{aligned}\Psi_{RSVP}^{FMIP} &= \Psi^{FMIP} + \Psi^{RSVP} + \Psi_{tunnel}^{RSVP} \\ &= N[\frac{2(l_{mh} + l_{mc} + 1.5l_{on} - 2 + 4.5\rho)\delta_B + 6\gamma}{tr}] \\ &\quad + N[\frac{P((l_{on} + \rho)\delta_D + 2\gamma + \beta)}{tr}] \\ &\quad + N[\frac{4[(l_{mc} + l_{on} - 1 + \rho)\delta_R + (l_{mc} + l_{on} - 1)\gamma]}{tr}].\end{aligned}$$

3.3.3 Results and Observations

Numerical results were obtained by using the parameter values presented in Table 3.3. The network topology assumes that the MN is 10 hops away from its CN *(*$l_{mc} = 10$*)* and is 3 hops away from its MAP *(*$l_{mm} = 3$*)*. The oAR and nAR are 2 hops away from each other and are connected to the same MAP *(*$l_{on} = 2$*)*. Moreover, due to the frame retransmissions and medium access contentions at the data link layer of wireless links, transmission costs of a wireless hop is higher than that of a wired hop; this effect is denoted by a proportionality constant (ρ = 10).

The processing cost (e.g. processing an RSVP message at an intermediate router or a binding update at the HA) is 30 ($\gamma = 30$). Since the transmission cost of a packet depends on its size, the per-hop transmission cost has been chosen to represent message's packet size: 80 for a binding update message ($\delta_B = 80$), and 140 for an RSVP message ($\delta_R = 140$). Similarly, the per-hop transmission cost of a data packet (δ_D) depends on its traffic type (δ_{VoIP} = 220, $\delta_{VIP} = 630$). Note that

Table 3.3: Signaling Cost Parameters.

$parameter$	$value$	$parameter$	$value$
l_{mc}	10	l_{mm}	3
l_{on}	2	ρ	10
tr	90	γ	30
δ_B	80	δ_R	140
δ_{VoIP}	220	δ_{VIP}	630
β_{VoIP}	220	β_{VIP}	630
P_{VoIP}	7	P_{VIP}	13

even though the maximum VIP data packet size is 1060, the actual video stream is variable bit rate (as depicted in Figure 3.2b). For the particular video stream used in Section 3.2, the average VIP data packet size was measured to be 630 bytes. Similarly, since the buffering space that a packet occupies is relative to its size, the buffering cost (β) has also been assigned according to packet size ($\beta_{VoIP} = 220$, $\beta_{VIP} = 630$).

To obtain the number of packets *(P)* that get tunneled during the execution of a FMIPv6 handoff, the simulation model in Section 3.2.1 was used. For VoIP traffic, the number of tunneled packets was measured to be 7 on average ($P_{VoIP} = 7$) while for VIP traffic 13 packets were tunneled ($P_{VIP} = 13$). The main reason for this difference is due to the different transmission rates of each traffic source (*VoIP = 88 kbps, VIP = 200 kbps*). Finally, the average MN residence time in a subnet is 90 seconds (*tr* = 90) to ensure that each MN undergoes one handoff during the simulation time of 180 seconds

Impact of number of mobile nodes

In this subsection we study the case when the number of mobile nodes (*N*) residing in the subnet is increased from 1 to 20 and the effect on network signaling cost is presented. As can be observed in Figure 3.13, the signaling cost increases linearly,

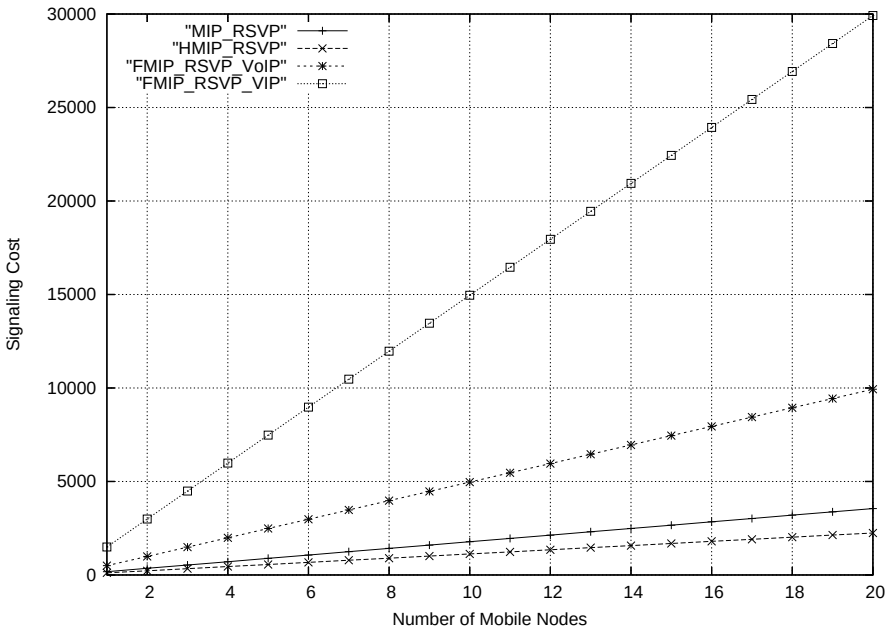

Figure 3.13: Impact of number of mobile nodes on signaling cost.

although at varying rates for the different mobility protocols. HMIPv6 incurs the lowest signaling cost on the network (112 to 2250), while MIPv6 is slightly higher (180 to 3550). This is because in a MIPv6 handoff, the MN exchanges mobility and QoS signaling along the complete end-to-end path with its CN. In HMIPv6 however, a MN confines signaling to its MAP and hence a fewer number of hops are utilized in the process thereby reducing the total signaling cost.

FMIPv6, on the other hand, has the highest signaling cost. As outlined in Section 2.5, FMIPv6 introduces additional mobility signaling on top of MIPv6 such as RtSolPr, PrRtAdv, HI and HAck (Figure 2.6). Although this signaling overhead contributes to FMIPv6's network signaling cost, it does not impose a significant impact as it is only exchanged locally across single hops. As a result, the key contributing factor to FMIPv6's signaling cost is the packet tunneling cost (Ψ^{FPT}) which depends primarily on the specific characteristics of the data traffic used.

In the case of VoIP traffic, FMIPv6 signaling cost ranges from 500 to 10000, while for VIP traffic it ranges from 1500 to 30000. An interesting observation is

that even though VoIP is transmitted at a rate of 88 kbps, and VIP at an average rate of 200 kbps, the signaling cost is tripled (rather than doubled). The reason being that the VoIP traffic source simulates human speech by cycling through active "on" and silence "off" periods. VIP on the other hand is constantly on. Consequently, a larger number of VIP packets are tunneled for the same handoff time than VoIP packets.

Impact of residence time

In the previous subsection, the impact of number of mobile nodes (N) was evaluated by maintaining the residence time (tr) at a constant value of 90 seconds (half the total simulation time of 180 s) while increasing N from 1 to 20. In this subsection, the reverse process is done in order to evaluate the impact of residence time: The number of mobile nodes is fixed at half capacity ($N = 10$) while tr is increased from 10 seconds to 170 seconds (since beyond this point no handoff occurs as the MN would reside in the subnet for the duration of the simulated time).

From Figure 3.14 it is apparent that residence time has a logarithmic impact on signaling cost as illustrated by the exponential decay curves. FMIPv6 decays at the highest rate when using VIP traffic (from 135000 to 8000), followed by VoIP traffic (from 45000 to 2500). In comparison, MIPv6 and HMIPv6 are less effected by tr and decay at approximately (15000 to 1000) and (10000 to 600) respectively.

By comparing Figures 3.13 and 3.14, it is noticeable that tr has a more significant impact on signaling cost than N. Therefore a scenario comprising a large number of MNs with low mobility rates, would be more favourable than a fewer number of MNs at high mobility rates.

3.3.4 Analysis and Discussion

Several analytical models have been developed to assess the signaling costs incurred on the network. The study focused on handoff scenarios, taking into account the

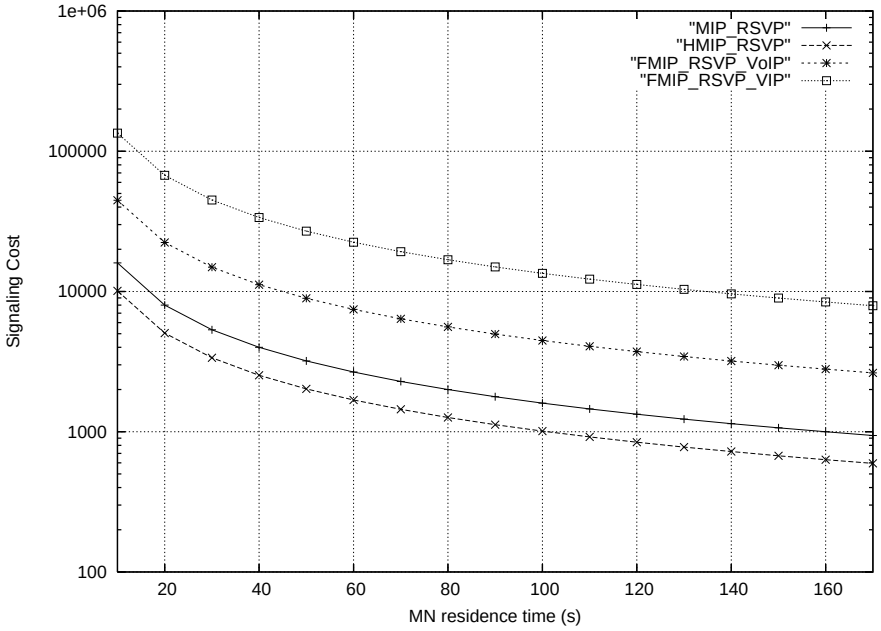

Figure 3.14: Impact of residence time on signaling cost.

mobility and QoS signaling invoked by the MNs residing in the network. HMIPv6 was found to be the most favourable choice from a network provider's perspective as it generates the least amount of signaling cost on the network. MIPv6 comes in second place, although in realistic terms a trade-off has to be made between complexity and signaling cost: While HMIPv6 offers less signaling cost, it requires the network operator to setup and manage a MAP entity at the edge of the network. MIPv6, on the other hand, does not require any additional nodes at the expense of higher network signaling cost.

In contrast, FMIPv6 has the highest signaling cost and is the only traffic-sensitive protocol of the three: For the particular VoIP and VIP traffic sources used in this study, signaling cost triples when using VIP over VoIP. Finally, FMIPv6 was found to be very prone to high mobility scenarios and is hence the least favourable choice from a network provider's point of view.

3.4 Conclusion

The framework presented in this chapter investigated the interaction of RSVP with Mobile IPv6 and some of its extensions. Two approaches were used, one from the end-user's perspective and another from the service provider's point of view. The first was a simulation-based study while the later was comprised of signaling cost models.

The results of the simulation study revealed a number of novel findings, particularly in the case of Hierarchical Mobile IPv6's unexpected decline in performance for Voice-over-IP traffic during periods of high congestion. Other findings include the exceptional application-level performance of Fast Handovers for MIPv6 (FMIPv6); in addition to the notable degradation in quality when using Mobile IPv6 for Video-over-IP traffic, which was due to MPEG4's inter-frame dependencies that trigger errors propagating to subsequent frames.

The analytical model, on the other hand, presented another view point to the framework. HMIPv6 was found to impose the least signaling costs on the network, followed by MIPv6. In contrast, FMIPv6 creates the highest amount of signaling cost, which was found to be proportional to the actual content being delivered. This was due to the packet tunneling cost of the temporary tunnel mechanism utilised by FMIPv6 during the execution handoffs. The results also revealed that the average residence time of a mobile node in a subnet has a greater impact than the actual number of mobile nodes present.

By investigating the interaction of RSVP and the presented mobility protocols from two different perspectives, a more balanced perception was achieved. While MIPv6 and HMIPv6 exhibit a homogeneous performance across the application and networks levels, they offer competitive advantages according to traffic type, handoff rates, and the desired level of complexity. FMIPv6, on the other hand, displays a strong bias in favour of application-level performance with little regard to network signaling cost.

Finally, all studies presented in this chapter use default RSVP behaviour with minor modifications (RSVP signaling is triggered by the mobility protocol when a handoff is completed). This highlights the necessity for a more efficient QoS solution to address the inefficiencies observed. In order to achieve an optimum QoS solution, both mobility and RSVP signaling should be integrated to work as one functional block during handoffs. This would reduce the total disruption in time by means of a single process that re-establishes both IP connectivity and resource reservation. Moreover, since a handoff would most likely change only a small segment of the end-to-end path, the new QoS solution should also be able to distinguish and re-establish only this changed portion of the link.

Chapter 4

Minimising Interruption in QoS: Using Embedded Mobility-Specific Information in RSVP Objects

4.1 Overview

The performance analysis study conducted and presented in Chapter 3 indicates that simple superimposition of RSVP and Mobile IP does not yield an efficient wireless QoS solution. This is mainly due to the independent operation of the two protocols where QoS signaling is not performed until mobility signaling is completed. With this notion in mind, this chapter presents a new protocol, called Mobility Aware Resource Reservation Protocol (*MARSVP*), in which the two protocols mentioned above perform as a single functional block. The key concept of MARSVP is to convey mobility-specific information (binding updates and their associated acknowledgments) using newly defined RSVP objects embedded in existing RSVP messages. This allows a single message exchange to establish both IP-level connectivity as well as QoS guarantees on the new link. An appealing feature of MARSVP is that it adheres to the current RSVP standard (RFC2205) and thus requires minimal

changes to end systems without affecting the operation of unmodified RSVP nodes in between.

The rest of this chapter is organised as follows: The next section outlines the design criteria used in developing the proposed mechanism, followed by a description of features and functionalities. An evaluation of application-level performance is presented in Section 4.4, followed by a network signaling cost analysis in Section 4.5. The chapter is then concluded with an in depth discussion of the results obtained from the various experiments and analytical models.

4.2 Design Criteria

In order to design a more efficient wireless QoS solution than RSVP, problem identification and design criteria have to be clearly defined. When RSVP and Mobile IP are deployed in the same network, two key issues arise: The first issue relates to the failure of intermediate routers to establish reservations for a roaming MN, while the second concerns the interruption in QoS during a handoff. In addition to addressing these two issues, the QoS solution has to follow an important design criterion which is to be compatible with the current RSVP standard.

4.2.1 IP-in-IP encapsulation

As outlined in Section 2.3, when a MN leaves the boundaries of its home subnet and enters a foreign subnet, it acquires a Care-of-Address (CoA) and sends its Home Agent (HA) a Binding Update (BU) message containing this information. The HA then creates an entry for the MN in its binding cache, and all incoming communication for this MN from Correspondent Nodes (CNs), would be encapsulated with the associated CoA and routed to the MN's new location at the foreign network.

This mechanism performs sufficiently well, until reservations are involved: when an RSVP Path message is encapsulated at the HA, the protocol number at the outer

IP header is set to 4 (for IP-in-IP encapsulation) while the original protocol number of 46 (RSVP) is concealed in the inner IP header. Moreover, the outer IP header does not carry the Router-Alert option used to notify RSVP routers to process the message in a hop-by-hop manner. As a result, the Path message is virtually "invisible" to RSVP routers and is forwarded as a normal data packet from the HA to the MN. When the MN's RSVP module inspects the Path message, the Previous Hop (PHOP) entry would still be pointing to the HA's address rather than actual previous hop router (since none of the routers beyond that point have processed the Path message). Consequently, the MN would fail to establish any reservations initiating from CNs while it resides at the foreign network.

An argument can be made that Route Optimisation could be utilised to enable direct communication between the CN and MN, thereby avoiding encapsulation at the HA. This in reality is not the case, since the first few packets from the CN (including the Path message) would always be tunneled to the MN by the HA temporarily until the CN is updated with a BU and responds to the HA with a Binding Acknowledgment (BAck). This results in the failure of establishing reservations in a similar way.

Another argument would be to explicitly program the HA to perform preferential encapsulation for RSVP messages by retaining the protocol number 46, in addition to the router alert option. This would effectively enable an RSVP session to be established between the CN and MN (via the HA). Nonetheless, data packets from the CN would still be encapsulated at the HA. Since the IP-in-IP encapsulation mechanism adds only an IP header as the external wrapper, no distinguishing information such as a UDP port is available in the outer header. As a result, it would be impossible for a packet classifier (Figure 2.3) at any of the RSVP routers between the HA and MN to distinguish between packets that use reservations from those that do not. Hence, data packets belonging to the CN would receive conventional best-effort treatment while the reservations would remain unused for the duration of the session.

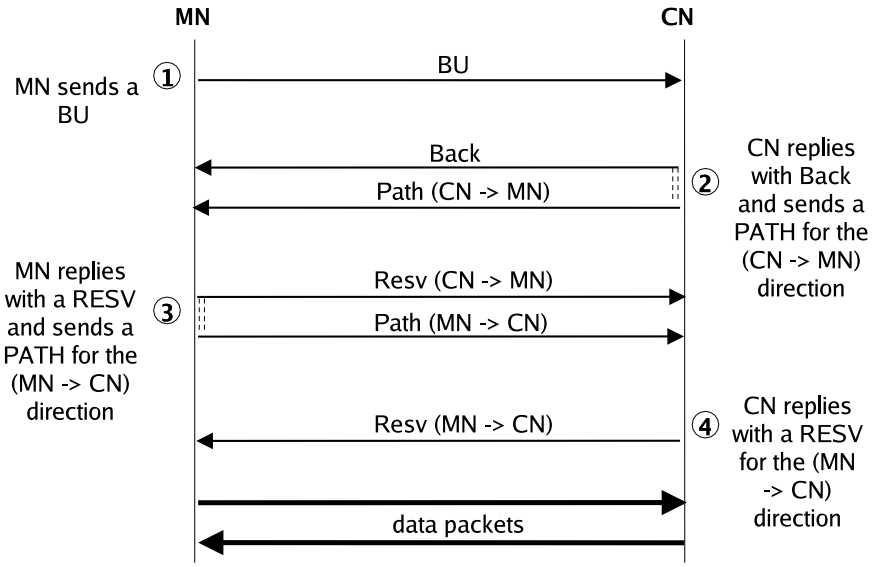

Figure 4.1: Mobility and QoS signaling during a handoff.

4.2.2 Interruption in QoS

Another issue arises in the event of a handoff where the total interruption in QoS (TI_{QoS}) inflicts a notable degradation in application level performance. TI_{QoS} consists of the time required to reconnect the mobile node (Mobile IP signaling) in addition to reserving resources on the new link (RSVP signaling). When a handoff occurs, a MN has to first acquire a new CoA and register it with its HA and CN before resuming data flow at its new location. This process involves the exchange of a BU/BAck message pair with the CN. As a result, mobility signaling takes approximately one round trip time (1RTT). Once connectivity has been established, the MN has to re-establish the full-duplex reservations (to and from) the CN. This requires a Path/Resv pair for the upstream direction, as well as another pair for the down stream connection. As a result the RSVP signaling delay is approximately 2RTT, and hence the total interruption in QoS is three round trip times (TI_{QoS} = 3RTT).

In real life, however, signaling is not performed in such a sequential manner:

The CN can respond with a BU and simultaneously issue a Path message for the CN → MN direction (indicated by the dotted line in step 2, Figure 4.1). Similarly, the MN would respond to the Path message with a Resv message, while at the same time initiating a Path message for the upstream (MN → CN) direction. As a result, the total interruption in QoS is essentially two round trip times. (TI_{QoS} = 2RTT).

The main objective of the proposed wireless QoS solution is to minimise TI_{QoS} as much as possible. This would reduce the number of packets dropped, and hence result in an improved user experience at the application level.

4.2.3 Compatibility with current standards

The proposed QoS solution should require minimal changes to the existing architectures (i.e. backwards compatible with the current RSVP and Mobile IP standards). This entails close examination of RFCs 2205 (RSVP) and 3775 (MIPv6) in order to streamline the deployment and integration of the proposed mechanism into existing RSVP-enabled networks. From a commercial perspective, this permits manufacturers to limit modification to the end systems, avoiding additional costs that would be incurred by modifying or replacing existing routers.

4.3 Features and Functionalities

In order to address the current issues of RSVP performance over wireless networks (Section 4.2), a Mobility Aware RSVP mechanism is proposed (*MARSVP*). This mechanism enables nodes to convey mobility-specific information in RSVP messages through two newly defined RSVP objects. MARSVP exploits the future compatibility built into RSVP (Section 3.10 of RFC 2205), which permits defining new object types. As a result, no changes are required to be made to the legacy RSVP-enabled routers. This method of integrating mobility signaling into QoS signaling should significantly improve RSVP renewing time after a handoff while at the same

time preserving the fundamentals of the RSVP implementation.

4.3.1 Protocol Overview

An RSVP message consists essentially of a common header, followed by a set of objects defining the various parameters of the data flow and its QoS requirements (Figure 2.1). Each object has its own header consisting of the following information:

- ***Object Length***: Total length of the object in bytes.
- ***Class-Num***: A number used to identify the RSVP object.
- ***C-Type***: Identifies the version of the Internet Protocol used (1 for IPv4, 2 for IPv6)

In the proposed mechanism, two new object types are defined: The BU object (Figure 4.2) and the BAck object; used to store the content of BU and the BAck messages respectively. Instead of assigning these new objects the next available class numbers (15 and 16), they are assigned 192 and 193 respectively. According to the RSVP implementation, a class number of 11bbbbbb (in binary format, where *b* represents a bit), is considered an unknown object class and as a result all intermediate nodes (which do not understand these objects) ignore but forward them unexamined and unmodified. This process allows the MN to send mobility information through RSVP objects which are transparent to all intermediate routers along the way. As a result, RSVP signaling continues to operate in its conventional manner, unaffected by the discretely embedded mobility information. Only two types of nodes are configured to be aware of the new mobility objects: The end nodes (CN and MN) and the mobility agents (HA or MAP).

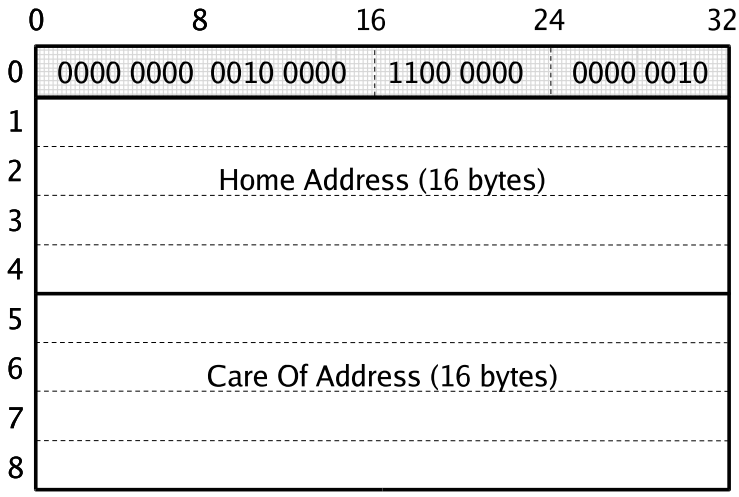

Figure 4.2: BU object format.

4.3.2 Establishing an RSVP Session

When the HA receives a Path message from a CN, it first checks it binding cache for an entry for the concerned MN. If an entry exists, it does not encapsulate the Path message but rather responds back with an RSVP PathTear message and appends a BU object to it. This message serves two simultaneous tasks: It tears down all Path States established in routers from the CN to the HA, in addition to updating the CN with the MN's CoA. Upon receiving the PathTear message, an intermediate node would search its Path State Block (PSB) list for a PSB entry whose (session, sender_template) pair matches the corresponding objects in the message. Once the PSB entry is found, the associated Previous Hop (PHOP) is used as the destination address for the PathTear message. The message is then forwarded upstream (along with the unmodified BU object) while the PSB entry is deleted from the list. When the PathTear message finally reaches the CN, the CoA in the BU object is used to construct a new Path message sent to the MN's exact location at the foreign network.

This new Path message is forwarded downstream to the MN and is processed by all intermediate routers as defined in RFC 2205. If all resources are available, the MN responds with a Resv message, thereby successfully establishing an RSVP

session in the downstream (CN → MN) direction. For duplex applications, another upstream RSVP session is also established for the reverse (MN → CN) direction.

In the case of HMIPv6, however, the MAP separates the RSVP session into a regional one (CN ↔ MAP) and a local one (MAP ↔ MN). This is accomplished by utilizing the inherited behaviour of HMIPv6 which confines signaling to the edge of the domain: When a MAP receives the Path message from the CN with the MN's Regional CoA (which is essentially the MAP's address), it replies with a Resv message; thereby establishing the regional (CN ↔ MAP) RSVP session. The MAP also setups a local (MAP → MN) RSVP session by exchanging local Path and Resv messages (using the local CoA) with the MN. The same approach is used to establish the upstream direction is in the reverse (MN → MAP) RSVP session for duplex applications.

4.3.3 Renewing an RSVP Session after a Handoff

To maintain an acceptable level of application performance, the interruption in QoS during handoffs should be kept to a minimum. A latency of 2RTT (Figure 4.1) may be acceptable at the time of session initiation, but it is not acceptable in the middle of an active session. This is especially true in the case of VoIP or VIP calls. The total interruption in QoS is essentially comprised of three main components: Reconnecting the MN, renewing the downstream (CN → MN) RSVP session, and renewing the upstream (MN → CN) RSVP session. MARSVP reduces this latency from 2RTT to 1.5RTT as follows: As soon as the MN acquires a new CoA on the new link, it sends the CN a Path message with an added BU object. As this Path message propagates upstream towards the CN, it installs a Path state in all routers along the way. Moreover, since these routers are not MARSVP enabled, they do not process the added BU object but rather forward it unexamined and unmodified. Once the CN receives the Path message, it reads the BU object and updates its binding cache with the MN's new CoA. It then uses the remaining RSVP ob-

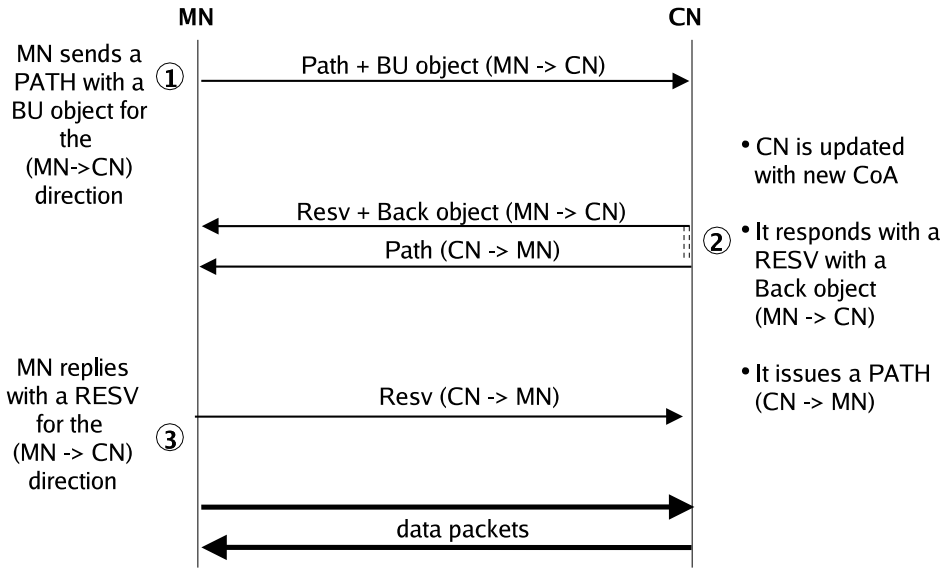

Figure 4.3: MARSVP signaling during a handoff.

jects (Sender_Template and Sender_TSpec) to reply with a Resv message and adds a BAck object to it. Furthermore, since the CN is now aware of the MN's new location it also renews the downstream RSVP session by issuing a Path message (CN → MN) using the new CoA. Note that these two messages are sent simultaneously as depicted by the dotted line in step 2 of Figure 4.3. As the two RSVP messages are routed towards the MN, the Resv message (MN → CN) reserves resources for the RSVP session in the upstream direction while the Path message installs the Path state for the downstream (CN → MN) direction. Finally, once the MN receives the Path message, it replies with a Resv message to reserve resources for the downstream (CN → MN) RSVP session. In the case of HMIPv6, however, only the local (MN ↔ MAP) RSVP session is renewed while the regional (MAP ↔ CN) remains still valid since the RCoA is unchanged.

4.4 Application-Level Performance

4.4.1 Methodology and Assessment

The aim of the simulation model is to evaluate the improvement in performance achieved by implementing the proposed MARSVP mechanism in wireless networks. The main focus is on the total interruption in QoS experienced during a handoff (TI_{QoS}) and the effect it has on application-level performance (using MOS as a performance metric), as observed by a mobile node in a congested wireless network. Offered traffic is fixed at 120% of link capacity in order to ensure intensive contention for bandwidth. The same simulation setup used in Section 3.2 was chosen, with the addition of the MARSVP implementation.

The first study aims to measure the value of TI_{QoS}, both with and without implementing the proposed MARSVP mechanism. The second study examines the effect of TI_{QoS} on VoIP performance by measuring the Mean Opinion Score as the number of handoffs is increased from 1 to 10. The same process is repeated for VIP traffic in the third study.

4.4.2 Results and Observations

Interruption in QoS

As outlined in Section 4.2.2, TI_{QoS} is the amount of time it takes the datastream to return to its pre-handoff level of quality. This time includes the Mobile IP signaling delay and the QoS signaling delay (whether RSVP or MARSVP). Consequently, TI_{QoS} is measured from the last packet received before the handoff occurs, to the first packet received *after* resources have been reserved on the new link.

By measuring TI_{QoS} for handoffs occurring in high congestion levels, the performance of MARSVP was tested against default RSVP. Three mobility scenarios were considered: MIPv6, HMIPv6 and FMIPv6. Table 4.1 summarizes the key findings: In the case of MIPv6, an MARSVP session experiences an average in-

terruption in QoS of 528 ms per handoff while a standard RSVP session experiences 732 ms. This translates into an improvement of 27.9%. Interestingly, while MARSVP still shows an improvement in a HMIPv6 scenario, the improvement is not as dramatic: 168 ms per handoff as opposed to 192 ms, resulting in an overall improvement of 12.5%. The reason for this behaviour can be explained by closely examining the factors contributing to TI_{QoS}: TI_{QoS} consists primarily of the handoff delay ($\lambda_{Handoff}$) and the RSVP signaling delay (λ_{RSVP}). $\lambda_{Handoff}$ can be further broken down into three components: The Layer-2 handoff delay (λ_{L2}) which is the actual time is takes the MN to reconnect at the new access point, the address resolution delay (λ_{Adr}), defined as the amount of time it takes the MN to acquire the new CoA, and finally the Layer-3 handoff delay (λ_{L3}) which represents the amount of time it takes to register the new CoA. TI_{QoS} can therefore be expanded to:

$$TI_{QoS} \;=\; \lambda_{L2} \;+\; \lambda_{Adr} \;+\; \lambda_{L3} \;+\; \lambda_{RSVP}. \tag{4.1}$$

In the presented simulations, λ_{L2} is around 20ms while λ_{Adr} is within the order of 100ms. Therefore, λ_{L3} and λ_{RSVP} remain the factors with the biggest impact. Since HMIPv6 confines signaling to the edge of the wireless domain, the Layer 3 handoff delay is greatly reduced as the BU/BAck messages traverse only to the MAP (as opposed to the CN). Similarly for λ_{RSVP}, only the local RSVP sessions between the MN and MAP are renewed and hence the Path/Resv messages are also sent only to the MAP. In MARSVP, the improvement in TI_{QoS} is 0.5RTT. In the case of MIPv6, this value is quite high while in HMIPv6, RTT is considerably lower since it is the Round Trip Time to the MAP (as opposed to the HA). Consequently, this difference in the value of RTT between MIPv6 and HMIPv6 is reflected on TI_{QoS}.

FMIPv,6 on the other hand, does not show indications of improvement due to two reasons: Firstly, in FMIPv6 the MN's CoA is negotiated and setup in advance (i.e. before the actual handoff occurs) and hence λ_{Adr} is effectively zero. Secondly, the temporary tunnel established during the execution of the handoff enables

Table 4.1: Total Interruption in QoS.

	RSVP	MARSVP	Improvement
MIPv6	732ms	528ms	27.9%
HMIPv6	192ms	168ms	12.5%
FMIPv6	24ms	24ms	0%

data packet to be delivered to the MN. Moreover, an RSVP session was also established over this temporary tunnel to maintain the desired QoS level. As a result, no matter how long it takes to perform mobility and RSVP signaling (whether using the proposed mechanism or not), the MN will maintain its connectivity and QoS level through the FMIPv6 tunnel. From the MN's perspective, there is virtually no interruption during handoffs (except for the initial packet redirection). Therefore, in terms of TI_{QoS}, there is no tangible improvement when using MARSVP over FMIPv6.

Impact of Number of Handoffs

In this study, the effect of the number of handoffs on application-level performance was examined through a series of simulation based experiments. The offered traffic was set at 120% of link capacity, while the number of handoffs was progressively increased from 1 to 10. Figure 4.4 shows the results using VoIP traffic. It can be observed that MARSVP maintains a higher MOS value in both MIPv6 (labeled MIP_MARSVP) and HMIPv6 (HMIP_MARSVP) scenarios than conventional RSVP. Moreover, the improvement in MOS increases as the number of handoffs increase. This is because MARSVP reduces TI_{QoS} by 0.5 RTT for every handoff. Therefore, at five handoffs, the total improvement in TI_{QoS} is effectively 2.5RTT for the whole session. This reduction in TI_{QoS} translates to less packets being dropped, thereby resulting in a higher MOS value.

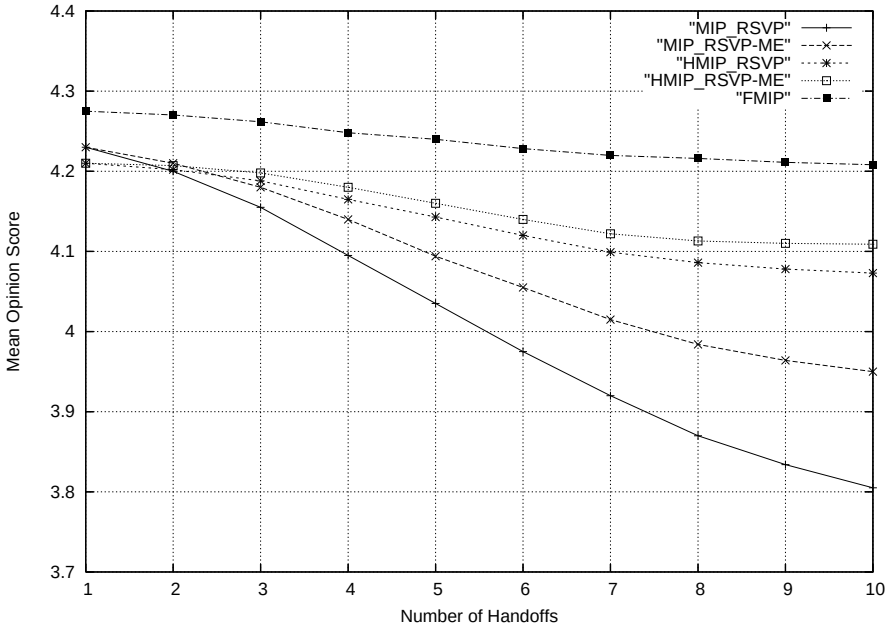

Figure 4.4: Effect of number of handoffs on VoIP performance.

Another observation is that the difference in TI_{QoS} between MIPv6 and HMIPv6 (Table 4.1) can also be noted on MOS performance: At five handoffs, MARSVP shows an improvement in MOS of 0.06 (4.09-4.03) for MIPv6 but only 0.02 (4.16-4.14) for HMIPv6. This difference is further amplified at ten handoffs with 0.14 (3.95-3.81) for MIPv6 and 0.035 (4.11-4.075) for HMIPv6. A final observation is that at one handoff MIPv6 outperforms HMIPv6 (both using RSVP and MARSVP), while beyond two handoffs HMIPv6 provides better performance. This observation was noted and explained in more detail in Section 3.2.3.

Similar observations are noted when using VIP traffic, with a much greater overall degradation in performance (Figure 4.5). Note that the scale used in this plot is different from VoIP (Figure 4.4), and hence to gain a better perspective on the magnitude of the degradation the two are combined into a unified plot in Figure 4.6.

At five handoffs, MARSVP shows an improvement in MOS of 0.17 (3.27-3.1) for MIPv6 but only 0.08 (3.53-3.45) for HMIPv6. This difference is further amplified at ten handoffs with 0.45 (2.65-2.2) for MIPv6 and 0.15 (3.1-2.95) for HMIPv6.

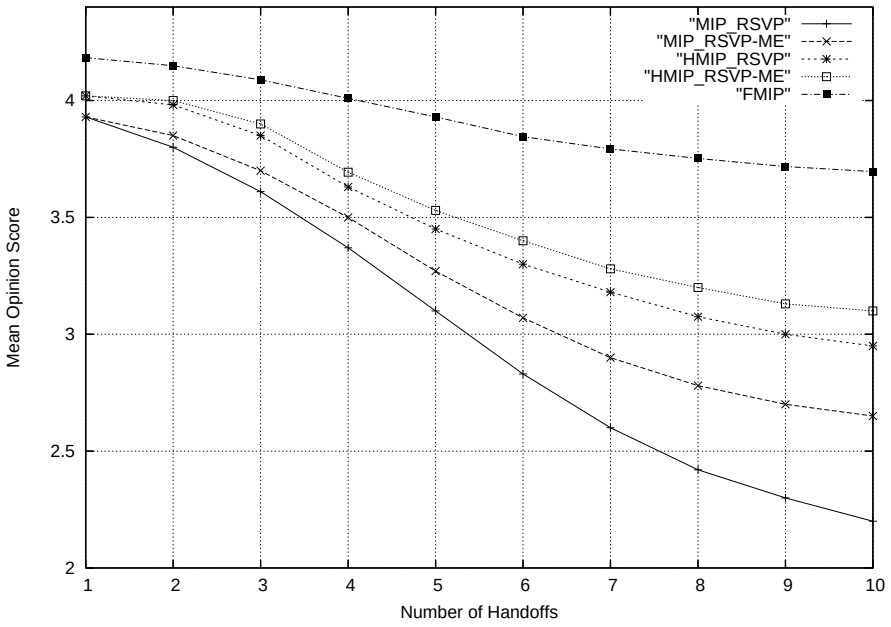

Figure 4.5: Effect of number of handoffs on VIP performance.

Moreover, HMIPv6 starts to outperform MIPv6 from one handoff (as opposed to three handoffs when using VoIP). For a full explanation of this observation, please refer to Section 3.2.3.

4.4.3 Analysis and Discussion

The proposed MARSVP mechanism was assessed for application-level performance in wireless IP networks, using a set of simulation based experiments. The total interruption in QoS (TI_{QoS}) during a single handoff was measured for the three mobility protocols and the relative improvement calculated. MARSVP was found to provide the greatest improvement in TI_{QoS} when deployed over MIPv6, followed by HMIPv6. This is a consequence of the signaling delay savings of 0.5RTT achieved by incorporating BU/BAck objects in RSVP messages.

In the case of MIPv6, RTT is measured between the MN and the CN, while for HMIPv6 it is measured between the MN and the MAP. This is a much shorter

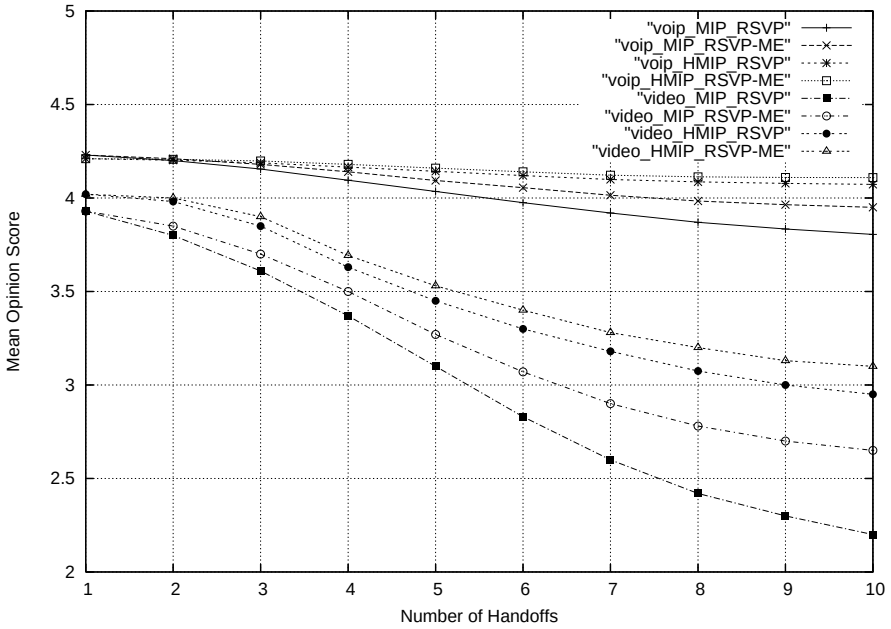

Figure 4.6: Effect of number of handoffs on Mean Opinion Score.

distance since the MAP lies within the boundaries of the same access network as the MN, while the CN could be located anywhere across the Internet. As a result, a saving of 0.5 RTT in MIPv6 translates to a larger improvement than 0.5 RTT in HMIPv6. FMIPv6, on the other hand does not provide any potential for improvement when using MARSVP. This is due to the temporary tunnel established by the fast handoff mechanism during the execution of a handoff. As a result, the MN maintains its connectivity and does not experience an interruption in QoS, except for the initial packet redirection.

When examining the impact of number of handoffs, the results observed for TI_{QoS} were mirrored on Mean Opinion Score: MIPv6 deteriorates the fastest as the number of handoffs is increased, while at the same time provides the biggest improvement when using MARSVP. HMIPv6's performance declines less, but at the same time offers less potential for improvement than MIPv6. FMIPv6 on the other hand, performs the best and does not indicate any improvement when using MARSVP (hence a single plot used for FMIPv6 in Figures 4.4 and 4.5). Finally, the

same behaviour was observed when using VIP traffic, although at a much greater magnitude (Figure 4.6). This is the result of the VIP traffic's MPEG frame dependencies outlined in Section 3.2.3.

4.5 Network-Level Signaling Cost

In this section, the proposed Mobility Aware RSVP mechanism (MARSVP) is evaluated at the network-level using signaling cost analysis. The analytical models derived in Section 3.3.2 are used as a benchmark for comparison.

Although the signaling cost analysis presented in this section follows a logic which is consistent with previous work [FAML05, XA02], it differs in the following manner:

1. The study focuses on signaling invoked due to a handoff and is therefore not concerned with data packet delivery cost.

2. It is assumed that an active session exists between the MN and its CN when the handoff occurs.

3. The CN is assumed to have a direct path of communication with its MN (i.e. Route Optimisation is active).

4. Processing costs at the HA, MAP and other nodes have been normalised by using a single universal processing cost (γ).

4.5.1 Total Signaling Cost using MARSVP

Based on the equations derived in Section 3.3.1, we formulate the total signaling cost of deploying MARSVP over Mobile IP, Hierarchical Mobile IP and Fast Handovers for Mobile IP protocols.

MARSVP over Mobile IP

In MARSVP no individual binding update messages are sent to the CN during a handoff, therefore transmission cost for a BU message from the MN to the CN is zero ($T_{mc} = 0$). This effectively equates the total CN registration cost to zero ($R_{mc} = 0$) since the processing cost at the CN is also zero ($P_{mc} = 0$), as outlined in Section 3.3.1. Consequently, the only cost incurred is (R_{mh}) since the BU/BAck message pair is still exchanged with the HA, regardless of the MARSVP mechanism. The new MIPv6 signaling cost (Ψ^{MIP^1}) is therefore formulated as follows:

$$\Psi^{MIP^1} = N[\frac{2(l_{mh} - 1 + \rho)\delta_B + \gamma}{tr}]. \tag{4.2}$$

Moreover, due to the addition of the two mobility objects (BU and BAck objects with a size of 36 bytes) to the first Path/Resv message pair (Figure 4.3), the new per-hop RSVP message transmission cost (δ_{R2}) has to be adjusted accordingly: The first Path/Resv message pair increases in size to 176 bytes while the second pair consists of traditional Path and Resv messages (no added BU/BAck objects) and thus remains at 140 bytes ($\delta_R = 140$). The new average per-hop transmission cost of all four RSVP messages is 158 bytes ($\delta_{R2} = (176 + 140)/2 = 158$) and hence the total signaling cost of MARSVP over MIPv6 (Ψ^{MIP}_{MARSVP}) equates to

$$\begin{aligned}
\Psi^{MIP}_{MARSVP} &= \Psi^{MIP^1} + \Psi^{RSVP^1} \\
&= N[\frac{2(l_{mh} - 1 + \rho)\delta_B + \gamma}{tr}] \\
&\quad + N[\frac{4[(l_{mc} - 1 + \rho)\delta_{R2} + (l_{mc} - 1)\gamma]}{tr}].
\end{aligned} \tag{4.3}$$

MARSVP over Hierarchical Mobile IP

Since only local handoffs are considered, no registrations occur with the HA in HMIPv6 (Equation 3.21). Furthermore, in MARSVP, no individual BU/BAck mes-

sages are exchanged with the MAP since this information is embedded in RSVP messages (used to re-establish local reservations to the MAP). As a result, the HMIPv6 transmission cost to the MAP is zero ($T_{mm} = 0$) and therefore the only registration cost incurred is the single processing cost at the MAP ($R_{mm} = \gamma$) and the new HMIPv6 signaling cost (Ψ^{HMIP^1}) is minimised to

$$\Psi^{HMIP^1} = N\frac{\gamma}{tr}. \tag{4.4}$$

The total signaling cost of MARSVP over HMIPv6 (Ψ^{HMIP}_{MARSVP}) is then formulated by adding Equations 4.4 and 3.18 (using δ_{R2} to denote the new RSVP message transmission cost):

$$\begin{aligned}\Psi^{HMIP}_{MARSVP} &= \Psi^{HMIP^1} + \Psi^{RSVP^1} \\ &= N[\frac{\gamma + 4[(l_{mm} - 1 + \rho)\delta_{R2} + (l_{mm} - 1)\gamma]}{tr}].\end{aligned} \tag{4.5}$$

MARSVP over Fast Handovers for Mobile IP

For the signaling cost of MARSVP over FMIPv6 (Ψ^{FMIP}_{MARSVP}), Equation 3.11 is substituted into Equation 3.23 as follows

$$\begin{aligned}\Psi^{FMIP}_{MARSVP} &= \Psi^{FR} + \Psi^{FPT} + \Psi^{MIP^1} + \Psi^{RSVP^1} + \Psi^{RSVP}_{tunnel} \\ &= N[\frac{(5\rho + 3l_{on})\delta_B + 5\gamma}{tr}] \\ &\quad + N[\frac{P((l_{on} + \rho)\delta_D + 2\gamma + \beta)}{tr}] \\ &\quad + N[\frac{2(l_{mh} - 1 + \rho)\delta_B + \gamma}{tr}] \\ &\quad + N[\frac{4[(l_{mc} + l_{on} - 1 + \rho)\delta_{R3} + (l_{mc} + l_{on} - 1)\gamma]}{tr}].\end{aligned}$$

Note that the last term of the above equation (which consists of Ψ^{RSVP^1} andΨ^{RSVP}_{tunnel}) is being multiplied by δ_{R3}. This value is the average per hop transmission cost of the total of eight RSVP messages exchanged (four for the duplex RSVP session, and another four for the temporary RSVP tunnel between the two access routers). The average per-hop transmission cost of the first four RSVP messages has been calculated earlier as $\delta_{R2} = 158$. The remaining four RSVP messages are traditional Path and Resv messages with a per-hop transmission cost of 140. Therefore the average per-hop transmission cost of the eight RSVP messages, δ_{R3}, equates to 149 ($\delta_{R3} = (158 + 140)/2 = 149$).

4.5.2 Results and Observations

The parameter values presented in Table 3.3 were used to obtain numerical results, with the following MARSVP attributes taken into account:

- Since in MARSVP, the first RSVP message pair carries extra BU/Back objects, the average per-hop RSVP message transmission cost is increased accordingly: $\delta_{R2} = 158$. Similarly, for FMIPv6, δ_{R3} becomes 149.

- Because MARSVP saves a signaling delay equivalent of 0.5 RTT, the temporary tunnel established by FMIPv6 is retained for a shorter period of time; thereby resulting in a fewer number of data packets being tunneled during handoffs. To obtain the exact number of packets (P^1) that get tunneled when using MARSVP over FMIPv6, the simulation model described in Section 3.2.1 is used. The number of tunneled data packets per handoff was measured to be 5 on average for VoIP traffic ($P_{VoIP} = 5$), and 9 for VIP traffic ($P_{VIP} = 9$).

The results presented in Tables 4.2 indicate that MARSVP produces reasonable signaling cost savings for all of the three tested mobility protocols. In the case of MIPv6 and HMIPv6, savings of 9.4% and 11.9% were achieved. Note that even

though MARSVP does not transmit individual BU/BAck messages to the CN, it still embeds their mobility content into RSVP objects. Therefore the difference in total signaling cost between conventional RSVP and MARSVP is the transmission cost of the IP headers of the BU/BAck message pair. This translates to signaling cost savings of 9.4% for MIPv6 and 11.9% for HMIPv6, as illustrated in Figure 4.7 by plotting the signal cost as the number of nodes is increased.

The reason for the slight difference in these two values for MIPv6 and HMIPv6 is as follows: In MIPv6, a MN would send two individual BU/BAck message pairs during a handoff, one to the HA and another to the CN. Moreover, the MARSVP mechanism is applicable only to the (MN $\leftrightarrow$ CN) direction over which the RSVP session exists. Consequently, MARSVP offers cost savings to only of the two BU/BAck message pairs. For HMIPv6, however, only a single local BU/BAck message pair is exchanged with the MAP (since the CN is not notified of the new local CoA), and MARSVP offers signaling cost savings over it. Therefore the IP header cost savings of the MARSVP mechanism results in a relatively better overall improvement for HMIPv6.

For FMIPv6, however, a number of observations are noted: Firstly, the signaling cost savings is higher than that for MIPv6 and HMIPv6 (between 17.9% and 26.7% as outlined in Table 4.2). The main reason for this is MARSVP's reduction of 0.5 RTT in TI_{QoS}. When deploying MARSVP over FMIPv6, this translates to a fewer number of data packets being tunneled: P_{VoIP} is reduced from 7 to 5 packets, while P_{VIP} is reduced from 13 to 9. Moreover, since FMIPv6's Packet Tunneling cost (Ψ^{FPT}) has a significant impact on the total signaling cost of MARSVP over FMIPv6 (Ψ^{FMIP}_{MARSVP}), any reduction in P is mirrored on to Ψ^{FMIP}_{MARSVP}. This results in Ψ^{FMIP}_{MARSVP}being directly proportional to P.

Another observation is that the signaling cost savings for FMIPv6 is traffic dependent (VoIP = 17.9%, while VIP = 26.7%) even though the number of tunneled data packets (P) is reduced proportionally for VoIP and VIP traffic. To help under-

Table 4.2: Improvement in Signaling Cost.

Mobility Protocol	Improvement
MIPv6	9.4%
HMIPv6	11.9%
FMIPv6 (VoIP)	17.9%
FMIPv6 (VIP)	26.7%

stand this, the ratio of the data packet size to the maRSVP packet size is examined:

$$\frac{P_{VoIP}}{\delta_{R2}} = \frac{220}{158} = 1.39,$$

$$\frac{P_{VIP}}{\delta_{R3}} = \frac{630}{158} = 3.99.$$

Therefore, a VoIP packet is 1.39 times larger than an MARSVP packet whereas a VIP packet is 3.99 times larger. Consequently, even if the number of data packets being tunneled is reduced proportionally by MARSVP for VoIP and VIP, this reduction has a bigger impact on the total signaling cost when using VIP traffic simply due to the VIP packet's larger packet size compared to the MARSVP packets. This difference in improvement can be observed when plotting the signaling cost against the number of nodes (Figure 4.8).

4.5.3 Analysis and Discussion

Several analytical models were derived to evaluate MARSVP's performance at the network level. The signaling cost analysis conducted in Section 3.3.2 was used as a performance benchmark to measure the improvement brought about by the proposed mechanism. MARSVP was found to have the least improvement in signaling cost when deployed over a MIPv6 or a HMIPv6 wireless network (9.4% and

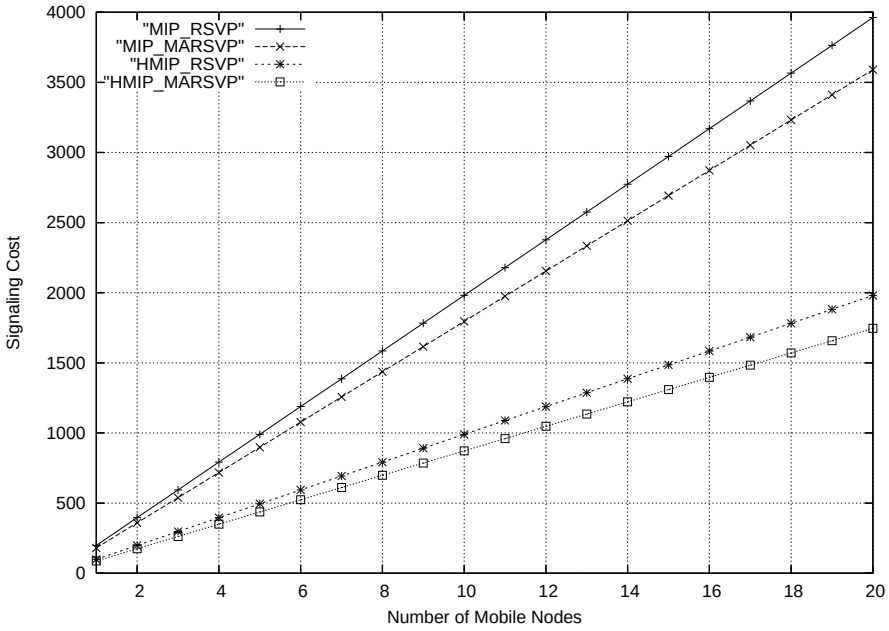

Figure 4.7: MARSVP signaling cost over MIPv6 and HMIPv6, as the number of mobile nodes is increased.

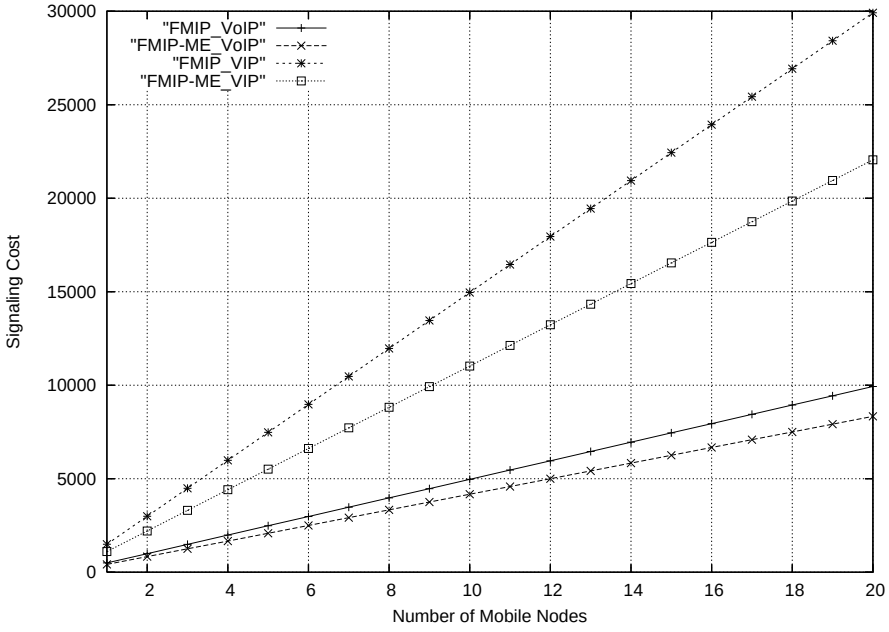

Figure 4.8: MARSVP signaling cost over FMIPv6, as the number of mobile nodes is increased.

11.9%). This is because the only signaling cost saving achieved is the transmission cost of the IP headers of the BU/BAck message pair.

When deployed over a FMIPv6 wireless network, however, MARSVP provides the best improvement in signaling cost, both when using VoIP and VIP traffic. The main reason behind this is the reduced number of data packets being tunneled and buffered between the old and new access points. From a network provider's viewpoint, MARSVP seems to be a favourable option to consider when delivering rich multimedia content over a FMIPv6 wireless network.

4.6 Conclusion

A Mobility Aware RSVP mechanism (MARSVP) for wireless mobile networks was presented and evaluated in this chapter. MARSVP adheres to RFC 2205 and operates concurrently with standard RSVP. It addresses two issues confronted when a MN roams outside its home network: establishing an RSVP session, in addition to the lengthy QoS interruption periods experienced during a handoff.

The proposed mechanism was evaluated from two viewpoints: the end-user's perspective (application-level performance) and the service provider's perspective (network-level performance). The framework presented in Chapter 3 was used as a performance benchmark to quantitatively assess any improvement introduced by the proposed mechanism.

Simulation results show that MARSVP reduces the total interruption in QoS (TI_{QoS}) by 27.9% when using Mobile IPv6, and by 12.5% when using Hierarchical Mobile IPv6. This reduction in TI_{QoS} results in a better end user experience, as indicated by the improvement in the Mean Opinion Score of VoIP and VIP applications used in the simulation study. Improvement gets better as the number of handoffs increases; which indicates the suitability of MARSVP for high mobility, high-handoff scenarios. FMIPv6, however, provides no potential for improvement in TI_{QoS} since a temporary tunnel is always used to maintain the MN's communica-

tion during a handoff regardless of how long it takes to perform the actual mobility and QoS signaling. As a result, MARSVP does not provide any improvement in TI_{QoS} when deployed over a FMIPv6 wireless network.

On the other hand, when examined at the network-level, the proposed mechanism exhibits a bias towards FMIPv6: MARSVP provides the best improvement in total signaling cost when deployed over a FMIPv6 wireless network (17.9% for VoIP applications and 26.7% for VIP). This was due to the significant packet tunneling and buffering cost savings resulting from MARSVP's reduction of 0.5 RTT in TI_{QoS}. In contrast, MARSVP provides a smaller improvement (9.4% and 11.9%) for MIPv6 and HMIPv6. This is because the only signaling cost savings gained is the transmission cost of the IP headers of the BU/BAck message pair.

By examining the proposed mechanism for application and network-level performance, a better overall insight into its features and limitations was achieved. MARSVP offers varying benefits, depending on the performance metrics required: If the end user's experience is a priority, MARSVP provides a notable improvement in Mean Opinion Score when deployed over MIPv6 or HMIPv6; whereas FMIPv6 remains unchanged. If the network signaling cost is the main concern, MARSVP offers considerable cost savings when using FMIPv6; in addition to a lower improvement when using MIPv6 or HMIPv6.

In conclusion, it can be said that MARSVP is a viable alternative to conventional RSVP for the following reasons:

- It is compatible with current standards and can therefore be introduced as a software update to end nodes. This allows for the core and access networks to remain unchanged (operate using conventional RSVP); thereby providing significant cost savings to service providers.

- It facilitates reservations for a roaming MN in a foreign subnet, previously unfeasible using conventional RSVP (see Section 4.2.1).

- It provides a better application-level performance for MIPv6 and HMIPv6

through a 27.9% and a 12.5% improvement in TI_{QoS} (respectively).

- At the network-level performance, it offers significant cost savings for all three mobility protocols (ranging from 9.4% for MIPv6 up to 26.7% for FMIPv6 using VIP traffic).

Finally, even though the mechanism proposed in this chapter provides substantial improvements at the application and network-level, it is limited by a design constraint (to be backwards compatible with current standards). This indicates that there is still potential for improvement if this design constraint is removed. This will allow for a more efficient solution that would further reduce the interruption in TI_{QoS}, albeit imposing a need for software upgrades in all routers along the end-to-end to path.

Chapter 5

Classifying RSVP Flows Using the Home Address Option

5.1 Overview

The proposed mechanism presented in Chapter 4 met its design requirements and produced sufficient results with improvements of up to 27.9% in TI_{QoS} at the application-level (MIPv6) and 26.3% improvement in signaling costs at the network-level (FMIPv6 using VIP traffic). However, this improvement can be further increased by eliminating the compatibility requirement (Section 4.2.3) from the design criteria. By doing so, this chapter presents a new classification mechanism for RSVP in which routers are configured to classify flows based on the home address option in the MIPv6 destination options header.

With this arrangement, intermediate RSVP routers are therefore able to correctly identify an RSVP flow, even after a MN changes its CoA. Moreover, using this mechanism a crossover router (COR) can detect the changed portion of the end-to-end RSVP session and confine RSVP signaling to it. As a result, the RSVP re-establishment time and network signaling costs are substantially reduced.

The rest of this chapter is organised as follows: The next section outlines the

design criteria used in developing the proposed mechanism, followed by a description of features and functionalities. An evaluation of application-level performance is presented in Section 5.4, followed by a network signaling cost analysis in Section 5.5. The chapter is then concluded with an in-depth discussion of the results obtained from the various experiments and analytical models.

5.2 Design Criteria

Since the compatibility requirement of Section 4.2.3 has been removed, greater flexibility in designing a more efficient system is allowed and issues can be addressed: The first issue is dual reservations for the same MN due to the current RSVP packet classification method, while the second issue relates to confining RSVP signaling to the unchanged portion of the end-to-end path. Moreover, the increased design flexibility permits reducing TI_{QoS} even further than the MARSVP method presented in Chapter 4.

5.2.1 Dual Reservations

As outlined in Section 3.2.2, when a MN experiences a handoff, its CoA changes. The MN's reservations, however, are still allocated for its old CoA. Once the MN resumes its transmission, RSVP routers would not be able to recognise the MN's data packets (since the MN is using a different IP address) and would therefore not allocate the reserved resources to the MN's "post-handoff" data flow. As a result, a new RSVP session (using the new CoA) has to be established every time the MN moves into a new subnet.

This new session, however, is established before the old one is relinquished (make before break). Moreover, since a handoff would most likely affect a small portion of the complete end-to-end path, the old and new RSVP sessions would share several links along the path between the MN and the CN. In severe cases, an

existing RSVP session for a MN (using the old CoA) can block the same MN from establishing a new RSVP session after a handoff despite the fact that both sessions are intended to handle the same traffic flow.

Even if a MN manages to establish reservations using its new CoA, the old RSVP session would be retained for a significant period of time. This is because the MN would not be able to explicitly tear down its old reservation using a ResvTear message after a handoff occurs (since the message would traverse along a different route). Consequently, the old reservations would be preserved until the RSVP refresh timer expires (anywhere up to 30 seconds, depending on network settings). In the worst case, a MN might ping-pong between two subnets, thereby creating new reservations as it moves between the two subnets. Although a smaller refresh interval might reduce the impact of this problem, it would also increase signaling overhead as refresh messages would be exchanged more frequently.

Therefore, a new classification method has to be developed in order to reassign existing reservations to a MN after a handoff occurs, thus avoiding dual reservations for the same traffic flow.

5.2.2 Confine RSVP Signaling to affected links

Since only a small portion of the entire end-to-end path is changed, RSVP routers should be modified to be able to identify the changed portion and confine signaling to it accordingly. This is illustrated in Figure 5.1: a MN initially resides in subnet 1 (S1) and moves horizontally towards subnet 3 (S3). When the MN first handoffs to S2, the changed portion between the MN-CN path consists of the last hop (R1-AR2) while the remaining hops remain unchanged. When the MN moves into S3, however, two hops are changed (R3-R2 and R2-AR3). Using conventional RSVP, a new end-to-end session will always be reserved, regardless of the actual number of links that changed during the handoff. By limiting RSVP signaling to the changed links, a fewer number of routers will be involved and therefore both transmission

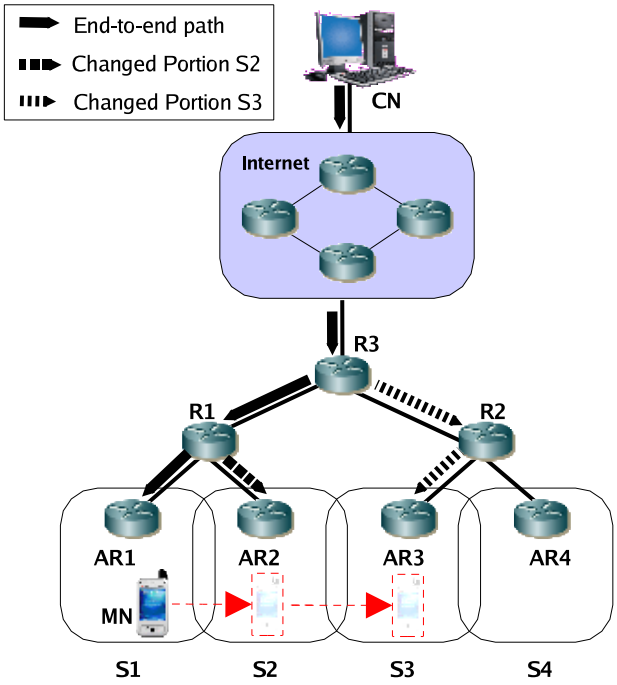

Figure 5.1: Changed links in an RSVP session during handoffs.

and processing costs would be reduced accordingly.

5.2.3 Interruption in QoS

As explained in Section 4.2.2, the total interruption in QoS (TI_{QoS}) consists of the time required to reconnect the mobile node (Mobile IP signaling) in addition to reserving resources on the new link (RSVP signaling). The mechanism proposed in Chapter 4 minimises TI_{QoS} by merging mobility with QoS signaling. This allowed a single message exchange to perform mobility signaling and establish reservations for one of the two directions. However, another RSVP messages exchange was required to renew the reverse RSVP session, resulting in a total signaling delay of 2 RTT.

From a MN's perspective this is experienced as 1 RTT delay of disruption in service (until CN is updated with the MN's new CoA), followed by another 1 RTT

of substandard level of QoS (until reservations are established for the reverse path). The objective of the new mechanism is to reduce TI_{QoS} as close as possible to 1 RTT, which would essentially consist solely of the mobility signaling delay. This effectively means that the reservations should be renewed during the time frame of 1 RTT used to conduct mobility signaling between the MN and CN. Therefore, by the time the MN and CN resumed their transmission, the reservations should have already been renewed and ready to handle the traffic flow's QoS requirements.

5.3 Features and Functionalities

Taking into account the aforementioned design criteria, a mechanism incorporating a new packet classification method is proposed as an extension to RSVP. The new mechanism called *RSVP-HoA* classifies traffic flows based on the MN's Home Address rather than the source or destination address stored in a packet's IP header. According to the MIPv6 specification [JP04], IP packets sent by a roaming MN should explicitly include a home address option in the destination options extension header. RSVP-HoA utilises this readily available information and thus no changes are required to the mobility protocols. On the other hand, changes are still required to RSVP to allow routers to inspect the home address option and limit RSVP signaling to the changed portion.

5.3.1 Protocol Overview

Several options are available in the current IPv6 specification (RFC 2402) and are handled as extension headers. These extension headers are inserted into the IPv6 header as needed and can consist of any of the following:

- Hop-by-hop options header,
- Routing header,

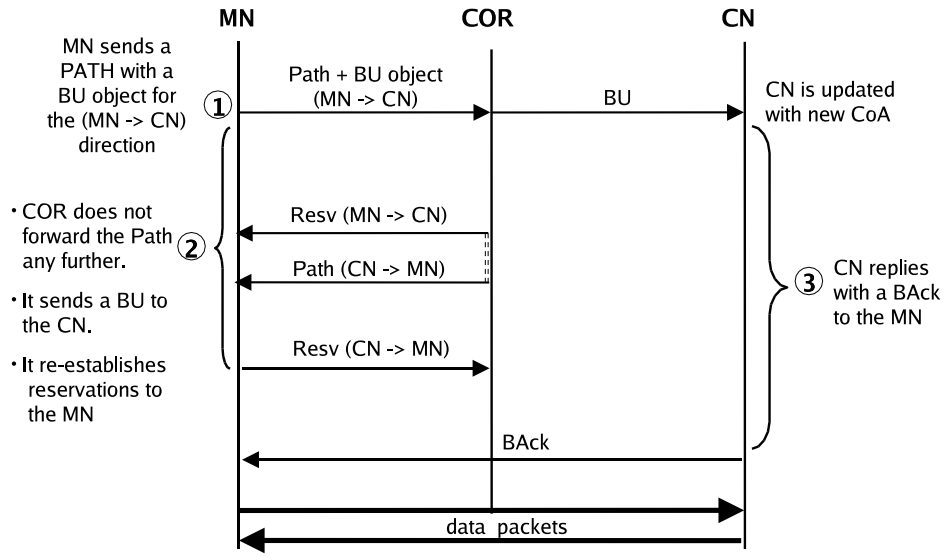

Figure 5.2: RSVP-HoA Signaling during a handoff.

- Fragment header,
- Destination options header,
- Authentication header,
- Encrypted security payload header.

The Destination options header is used to carry information to be processed at the destination node. Mobile IPv6, for example, specifies the Home Address option to be carried by the destination options header. The main objective behind this is to provide transparency at the application layer. Therefore if a MN starts a session with a CN using its Home Address (HoA) and then undergoes a handoff and starts using a CoA, the home address option must be included in all of the MN's data packets transmitted to the CN. At the receiving end, the CN's transport layer retrieves the MN's HoA from the destination options header and places it in the source address field of the data packet (instead of the CoA), which is then passed on to the higher layers. As a result, the MN's changed IP address would be transparent to the CN's

application-layer. The reverse process is done by the CN's network layer when its application layer replies with data packet destined to the MN: the HoA in the destination field is replaced with the MN's CoA (which is the topologically correct IP address).

This process provides application layer transparency, enabling applications to run smoothly, unaffected by the MN's changed IP address. In a similar fashion, the proposed RSVP-HoA mechanism utilises the home address option to provide transport layer transparency for RSVP. Moreover, the BU/BAck objects introduced by the MARSVP mechanism are once again utilised in the proposed RSVP-HoA mechanism to capitalise on the 0.5RTT savings (as depicted in Figure 5.2).

5.3.2 Path Message Processing

Using the proposed RSVP-HoA mechanism, an RSVP router processes a Path message according to the flowchart diagram depicted in Figure 5.3. The router first examines the Path message's headers for a destination options header containing a home address option.

If a home address option is not found, the router would behave in the conventional RSVP manner: The RSVP module's Path State Block (PSB) list would be searched for an existing session using the sender address found in the session object of the received Path message. If an entry exists, the Path message is considered a refresh message and the corresponding path state timer is reset and the message is forwarded upstream. However, if no existing PSB entry is found, the session is treated as new one and a new PSB entry is created. The router's IP address is recorded in the Previous Hop (PHOP) field of the Path message, which is then forwarded upstream towards the destination.

On the other hand, if a home address option is included in the Path message, the HoA is retrieved and swapped with the CoA in the session object. The PSB list is then searched for an existing entry using the HoA. If no entry is found, it

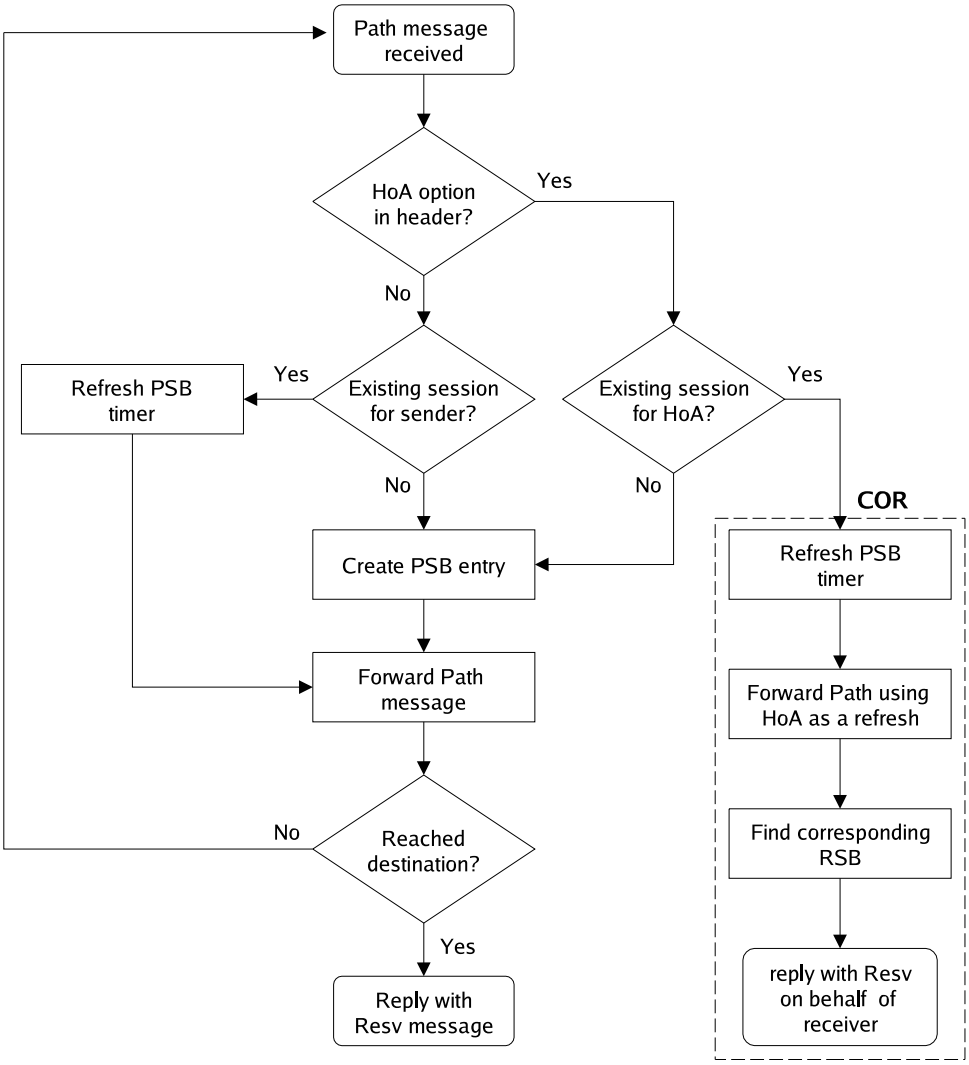

Figure 5.3: Path message processing using the RSVP-HoA mechanism.

indicates that the current RSVP router is a new one being added to the end-to-end path of a MN that has undergone a handoff (home address option included but no PSB found). A new PSB entry is therefore created and the message is forwarded upstream (router's IP address is recorded in the PHOP entry of the Path message). However, if an entry for the HoA exists, it indicates that the current RSVP node is the COR for a MN that has undergone a handoff. The COR will therefore not forward the Path message any further, and using the associated Reservation State

Block (RSB) entry, would issue a Resv message back to the MN on behalf of the CN.

5.3.3 Data Flow Classification

An RSVP router would initially inspect an incoming data packet for a home address option. If a home address option is not found, the router would function in the conventional RSVP manner by classifying the data packet according to the sender address field in the IP header. If an RSVP session exists, resources are allocated accordingly and the packet receives its desired QoS level. However if a home address option is included, the HoA is retrieved and used to classify the data packet. If an RSVP session exists for the HoA, the designated resources are allocated and the packet receives its desired level of QoS.

This method insures that, when a MN undergoes a handoff and resumes communication using its new CoA, the change in IP address would be transparent to the RSVP module's packet classifier. As a result, the MN is no longer required to create new end-to-end reservations using its new CoA since the reservations on the unchanged portion of the end-to-end path would be reallocated to the MN's flow (using the home address option). As a result, the dual reservation issue outlined in Section 5.2.1 is avoided.

Moreover, the Path message sent by the MN after the handoff would only travel to the COR which would reply with a Resv to reserve resources on the changed portion of the link. This confines RSVP signaling to the COR (Section 5.2.2) and would therefore result in a shorter RSVP signaling delay since the Path and Resv messages only traverse a few hops. This, in return, reduces the total interruption in QoS (Section 5.2.3). The signaling cost of RSVP would likewise be reduced due to the smaller number of hops and fewer number of routers involved in processing the Path message.

Table 5.1: Total Interruption in QoS.

	RSVP	MARSVP	Improvement	RSVP-HoA	Improvement
MIPv6	732ms	528ms	27.9%	312ms	57.4%
HMIPv6	192ms	168ms	12.5%	168ms	12.5%
FMIPv6	24ms	24ms	0%	24ms	0%

5.4 Application-Level Performance

5.4.1 Results and Observations

The application-level performance of RSVP-HoA was measured and compared against standard RSVP and MARSVP by monitoring the effect on the total interruption in QoS under high congestion levels. Offered traffic was fixed at 120% of link capacity while the number of handoffs was increased from one to ten. The same simulation environment and topology used in Section 3.2 was utilised, with the addition of the RSVP-HoA implementation. Table 5.1 presents a performance matrix of TI_{QoS} using the different combinations of the three mobility protocols and RSVP mechanisms.

Mobile IPv6

As shown in Table 5.1, TI_{Qos} for MIPv6 was further minimised using RSVP-HoA to 312 ms per handoff, compared to 528ms for MARSVP and 732 ms for standard RSVP. This translates to an improvement of 57.4%, which is significantly higher than 27.9% for MARSVP.

Since RSVP-HoA confines RSVP signaling to the COR (node N4 as depicted in Figure 3.1), the associated RSVP signaling delay is reduced accordingly. Moreover, because the transmission delay from the MN to the CN is much higher than

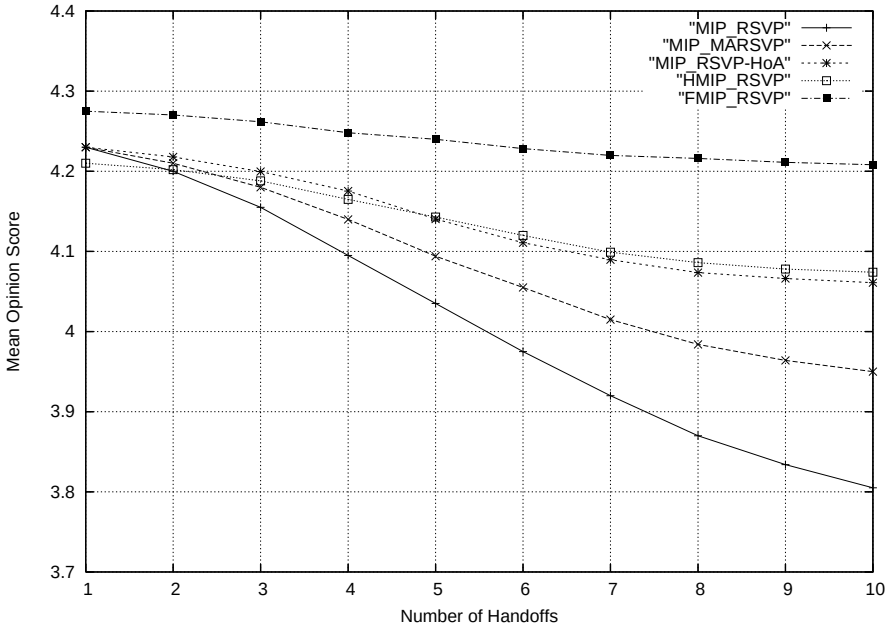

Figure 5.4: Effect of number of handoffs on VoIP performance.

that to the COR, RSVP signaling is effectively performed in parallel with (and is completed before) mobility signaling. As a result, by the time the CN is ready to resume transmission, the resources would have already been reserved for it at the new nodes, N3 and AR1. This essentially means that TI_{QoS} consists of the mobility signaling delay between the CN and MN, in addition to the inherited MAC contention delay at the access routers due to high traffic congestion. The improvement in TI_{QoS} can also be noted on application-level performance, as illustrated in the MOS graphs for VoIP and VIP traffic (Figures 5.4 and 5.5 respectively).

In the case of VoIP traffic, MARSVP (labeled *MIP_MARSVP*) provides an improvement in MOS of 0.06 (4.09 - 4.03) over standard RSVP (labeled *MIP*) at five handoffs. However, RSVP-HoA (labeled *MIP_RSVP-HoA*) provides a better improvement in MOS of 0.11 (4.14 - 4.03) for the same number of handoffs. This difference is further amplified at ten handoffs with 0.14 improvement in MOS for MARSVP and 0.24 for RSVP-HoA.

An interesting observation is that using RSVP-HoA with MIPv6 delivers a MOS

performance comparable to that of HMIPv6 and standard RSVP (labeled *HMIP-RSVP*). A slight variation exists where MIP_RSVP-HoA performs slightly better than HMIP-RSVP below five handoffs, after which HMIP_RSVP starts to outperform it. Note that even though HMIP_RSVP provides a better TI_{QoS} of 192 ms per handoff (compared to 312 ms for MIP_RSVP-HoA), the additional 40 byte IP-in-IP 40 header (18% increase in data packet size) results in higher packet delay and loss in RSVP routers and access routers (MAC contention). This affects HMIPv6's MOS performance to the point where it is slightly worse than MIP_RSVP-HoA. As the number of handoffs increase, the lower TI_{QoS} of HMIPv6 results in less packets being dropped and delayed (for every handoff) and thus by five handoffs it starts to compensate for the increased data packet size.

For VIP traffic (Figure 5.5), HMIP_RSVP outperforms MIP_RSVP-HoA regardless of the number of handoffs. As explained in Section 3.2.3, this is due to the larger packet size of VIP traffic which reduces the relative impact of HMIPv6 IP-in-IP encapsulation (3.8% increase for VIP, compared to 18% for VoIP). Finally, RSVP-HoA results in a more noticeable MOS improvement when using VIP traffic than VoIP: By ten handoffs, MIPv6 MOS is enhanced from 2.2 for standard RSVP to 3.1 for RSVP-HoA, compared to 3.805 and 3.95 for VoIP traffic.

Hierarchical Mobile IPv6

For HMIPv6, RSVP-HoA and MARSVP provide the same TI_{QoS} value of 168 ms (Table 5.1). This is because the simulation topology used (Figure 3.1) is a single level hierarchy in which the MAP (node N4) is the COR between AR1 and AR2. Moreover, since RSVP Operation over IP Tunnels is used, mobility and QoS signaling are confined to the MAP regardless of the RSVP mechanism used. As a result, the only advantage RSVP-HoA provides at single level hierarchy is the 0.5RTT saving in TI_{QoS} achieved by implementing embedded BU/Back objects in RSVP messages. This results in the same improvement in TI_{QoS} of 12.5% achieved by MARSVP.

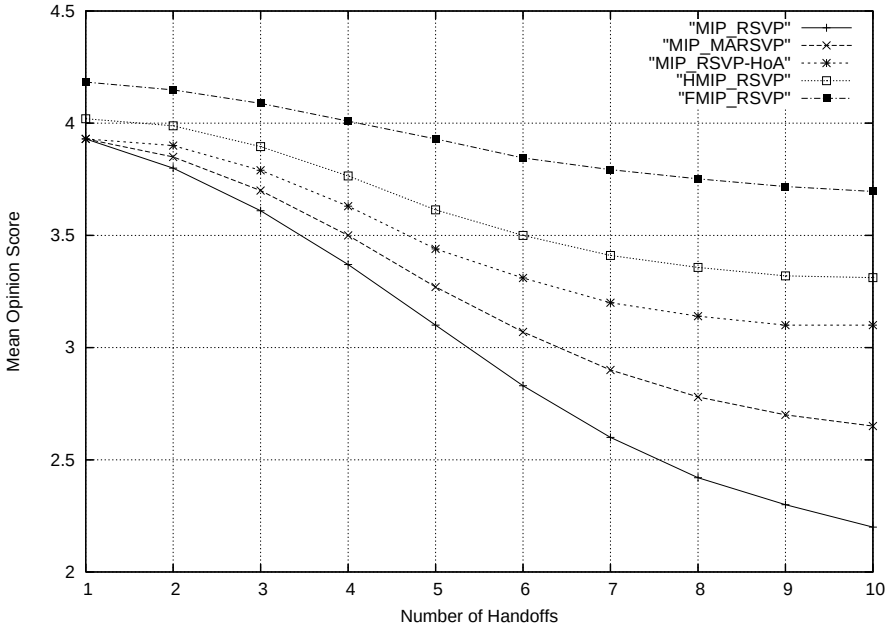

Figure 5.5: Effect of number of handoffs on VIP performance.

Therefore, in order to measure the advantage of RSVP-HoA, a larger HMIPv6 hierarchy architecture should be considered. For example, for a 2-level hierarchical topology, the COR would be one hop away from the MN while the MAP would be two hops away. As the number of hierarchy levels increases, the advantage of RSVP-HoA becomes more apparent: At five handoffs, RSVP-HoA re-establishes reservations to the COR (one hop away), while MARSVP re-establishes them to the MAP (5 hops).

With this notion in mind, a new set of simulations were conducted using various HMIPv6 levels of hierarchy for the three RSVP mechanisms. As can be observed in Table 5.2, RSVP-HoA's advantage over MARSVP increases as the number of hierarchy levels is increased: At a 5-level hierarchy, standard RSVP results in a TI_{QoS} of 433 ms, while MARSVP is 326 ms (24.7%) and RSVP-HoA 248 ms (42.7%).

In order to illustrate the effect of TI_{QoS} on MOS performance, a three-dimensional plot is used for the three parameters involved: the number of handoffs (x-axis), the Mean Opinion Score (y-axis), and the number of hierarchy levels (z-axis). By ex-

Table 5.2: Improvement in TI_{QoS} according to HMIPv6 hierarchy depth.

Levels of Hierarchy	RSVP	MARSVP	%	RSVP-HoA	%
1	192 ms	168 ms	12.5%	168 ms	12.5%
2	250 ms	205 ms	18%	189 ms	24.4%
3	311 ms	247 ms	20.6%	206 ms	33.8%
4	372 ms	289 ms	22.3%	227 ms	39%
5	433 ms	326 ms	24.7%	248 ms	42.7%

amining the VoIP MOS performance graph for MARSVP (Figure 5.6) a few observations are made: MOS performance decreases as the number of handoffs is increased (as illustrated in earlier two-dimensional plots). However, the magnitude of this drop is higher as the number of hierarchy levels is increased. For example, at one-level of hierarchy MARSVP's MOS drops from 4.23 (one handoff) to 3.95 (ten handoffs), resulting in a 0.28 drop in MOS. At five-levels of hierarchy, MOS decreases from 4.28 to 3.4, resulting in a larger drop of 0.88.

By comparing MARSVP's performance against RSVP-HoA (Figures 5.6 and 5.7), it can be observed that RSVP-HoA's advantage is more visible at larger hierarchy levels: At five-level hierarchy, RSVP-HoA's MOS decreases from 4.27 to 3.49 while MARSVP's MOS from 4.28 to 3.4. Therefore, even at one handoff, RSVP-HoA still performs better than MARSVP since RSVP signaling only travels a single hop to the COR, while with MARSVP it travels five hops to the MAP.

When examining VIP traffic performance (Figures 5.8 and 5.9), the effect of hierarchy level seems minimal (the MOS plot does not curve on the z-axis as much as it did for VoIP traffic). This is due to the larger scale used since the number of handoffs has a bigger impact on VIP traffic than it did for VoIP (y-axis ranges from 2.6 to 4, as opposed to 3.4 to 4.3 for VoIP). By closely examining the impact of each

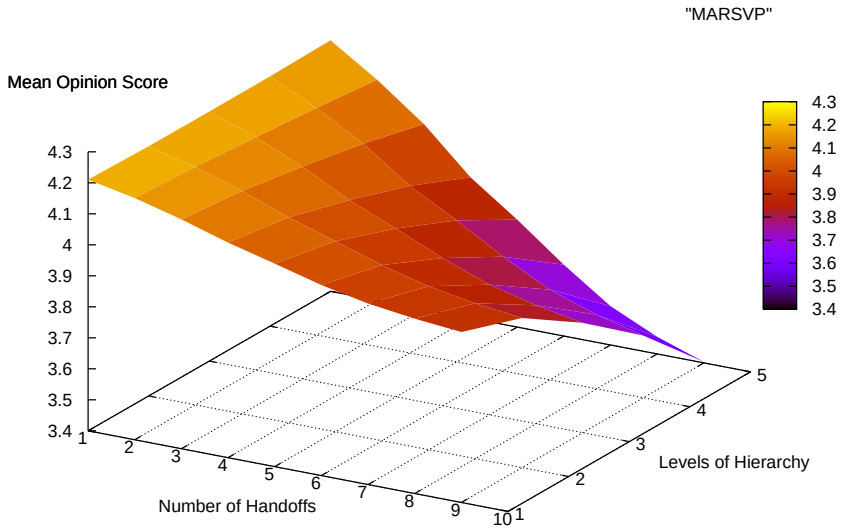

Figure 5.6: Effect of number of handoffs on VoIP performance (MARSVP).

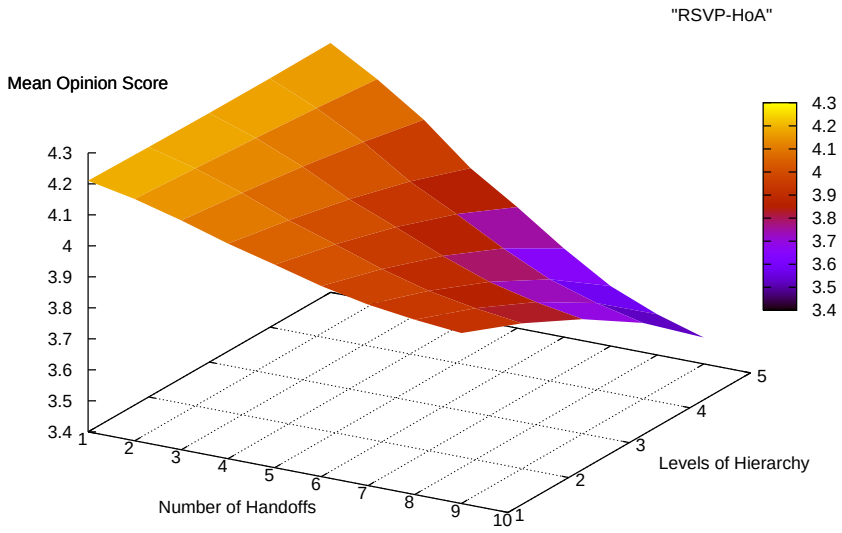

Figure 5.7: Effect of number of handoffs on VoIP performance (RSVP-HoA).

of the two variables, it becomes clear that in the case of VIP traffic, the number of handoffs has a bigger impact on MOS performance than the number of hierarchy levels:

- **Effect of Number of Handoffs:** At single level hierarchy, MOS drops from 4.02 (one handoff) to 3.1 (ten handoffs), resulting in a difference of 0.92. Similarly, at five-level hierarchy it drops from 3.94 to 2.73 (a difference of 1.21).

- **Effect of Number of Hierarchy Levels:** For a single handoff, MOS drops from 4.02 (1-level hierarchy) to 3.94 (5-level hierarchy), resulting in a difference of 0.08. Similarly, for ten handoffs, MOS drops from 3.1 to 2.73 (a difference of 0.37).

As a result, the relatively small change due to hierarchy levels (0.08 to 0.37) is not as clearly visible on the larger y-axis scale used for VIP traffic. In terms of RSVP-HoA's performance against MARSVP, Figures 5.8 and 5.9 illustrate MARSVP's lower MOS value as shown in the darker template colour on the upper-right region (large number of handoffs and hierarchy levels). RSVP-HoA, on the other hand, provides a better MOS performance as illustrated by the lighter colour template for the same upper-right region. Therefore, RSVP-HoA is a better candidate than MARSVP for large HMIPv6 wireless networks, where the number of hierarchy levels is larger than two.

Fast Handovers for Mobile IPv6

For FMIPv6, no improvement in TI_{QoS} is achieved (Table 5.1). This is due to the pre-handoff CoA configuration and the FMIPv6 tunnel utilised during handoffs (as explained in Section 4.4.2). Therefore, in FMIPv6's case, all three RSVP mechanisms provide the same TI_{QoS} value of 24 ms, which consists of the initial delay due to packet redirection.

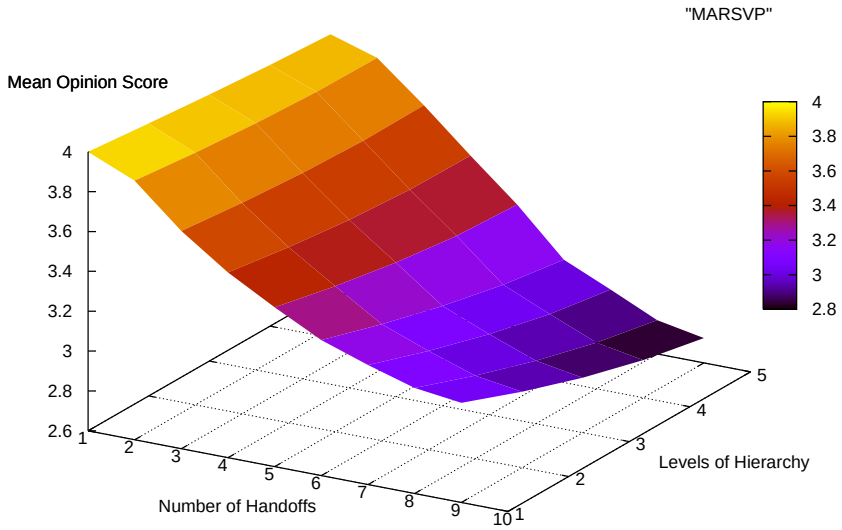

Figure 5.8: Effect of number of handoffs on VIP performance (MARSVP).

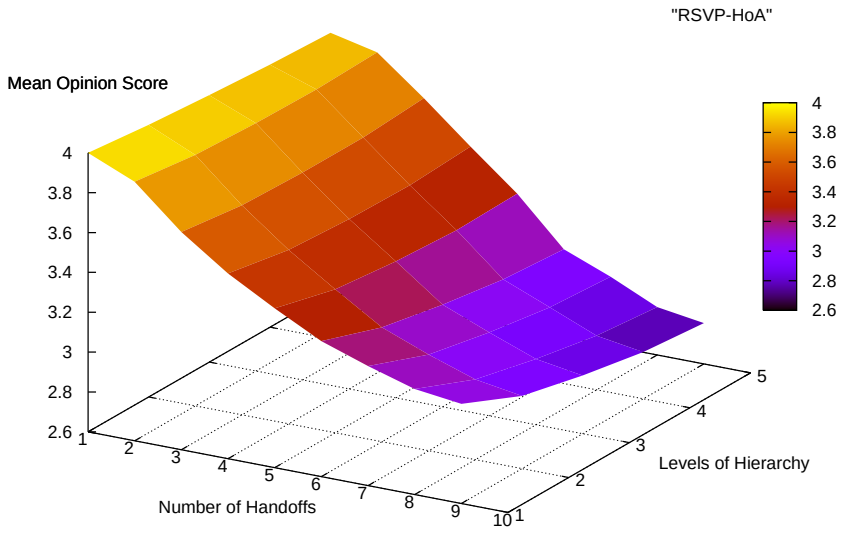

Figure 5.9: Effect of number of handoffs on VIP performance (RSVP-HoA).

5.4.2 Analysis and Discussion

The proposed RSVP-HoA mechanism was assessed for application-level performance in wireless IP networks, using a set of simulation based experiments. The total interruption in QoS (TI_{QoS}) during a single handoff was measured for the three mobility protocols and the relative improvement calculated. RSVP-HoA was found to provide the greatest improvement in TI_{QoS} when deployed over MIPv6 (57.4% over standard RSVP), followed by HMIPv6 (12.5%), whereas no improvement was achieved for FMIPv6.

To further assess RSVP-HoA's improvement in HMIPv6 wireless networks, another set of simulations was conducted using simulation topologies consisting of multiple levels of hierarchy. As the number of hierarchy levels increased, RSVP-HoA provided better performance than MARSVP. This is because reservations were renewed only to the COR (one hop) whereas MARSVP renewed them to the MAP (multiple hops, depending on hierarchical depth).

The corresponding effect of TI_{QoS} was observed on MOS performance. In the case of HMIPv6, three-dimensional plots were used to illustrate the effect of both the number of handoffs and hierarchy level on MOS performance.

5.5 Network-level Signaling Cost

In this section, the proposed reservation model (RSVP-HoA) is evaluated at the network-level using signaling cost analysis. The analytical models derived in Section 3.3.2 are used as a benchmark for comparison.

5.5.1 Total Signaling Cost using RSVP-HoA

Based on the equations derived in Section 3.3.1, we formulate the total signaling cost of deploying RSVP-HoA over the different mobility protocols.

RSVP-HoA over Mobile IPv6

Similar to the MARSVP signaling cost analysis conducted in Section 4.5.1, RSVP-HoA does not transmit a binding update message to the CN during a handoff but rather embeds it in a BU object sent in a Path message to the CN. Nonetheless, this holds true for RSVP-HoA until the Path message reaches the COR, which would not transmit it any further. The COR would read the BU object and send a standard BU message to the CN, while at the same time replying to the MN with a Resv message and another Path/Resv message pair for the reverse path. The CN then replies with a standard BAck message to the MN (consult Figure 5.2 for detailed message signaling).

Therefore the CN registration cost (R_{mc}) consists of the transmission cost of the BU message from the COR to the CN (MN to COR is not included since the BU information is transmitted in an BU object in the Path message), in addition to the BAck message from the CN to the MN. Furthermore, the number of hops from the COR to the CN can be calculated as simply the number of hops between the MN and CN (l_{mc}), minus the number of hops between the MN and COR (l_{mcr}). The CN registration cost is then calculated as the sum of the transmission costs of the BU ($COR \rightarrow CN$) and BAck message ($CN \rightarrow MN$) as follows:

$$\begin{aligned} R_{mc} &= N\frac{(l_{mc} - l_{mcr})\delta_B}{tr} + N\frac{(l_{mc} - 1 + \rho)\delta_B}{tr} \\ &= N\frac{(2l_{mc} - l_{mcr} - 1 + \rho)\delta_B}{tr}. \end{aligned} \tag{5.1}$$

As for the HA registration cost (R_{mh}), the BU/BAck message pair is still exchanged with the HA regardless of the RSVP-HoA mechanism. R_{mh} is simply the transmission costs of the two messages (BU/BAck) in addition to the processing cost at the HA:

$$R_{mh} = N\frac{2(l_{mh} - 1 + \rho)\delta_B + \gamma}{tr}. \tag{5.2}$$

The new MIPv6 signaling cost (Ψ^{MIP_2}) is formulated by combining Equations 5.1 and 5.2 as follows

$$\begin{aligned}\Psi^{MIP_2} &= R_{mc} + R_{mh} \\ &= N\frac{(2l_{mc} - l_{mcr} - 1 + p)\delta_B}{tr} + N\frac{2(l_{mh} - 1 + p)\delta_B + \gamma}{tr} \\ &= N[\frac{(2l_{mc} + 2l_{mh} - l_{mcr} - 3 + 3\rho)\delta_B + \gamma}{tr}] \end{aligned} \quad (5.3)$$

As for the RSVP signaling cost, it consists of the first Path message (with an embedded BU object, $\delta_R = 176$) to the COR, in addition to the remaining three Path/Resv messages exchanged between the MN and COR (do not include any BU/BAck objects, therefore $\delta_R = 140$). In order to simplify the equation, the average value of δ_R for the four RSVP messages is used: $\delta_{R4} = [(140 \times 3) + 176]/4 = 149$. The RSVP signaling cost becomes:

$$\Psi^{RSVP_2} = N[\frac{4[(l_{mcr} - 1 + \rho)\delta_{R4} + (l_{mcr} - 1)\gamma]}{tr}.] \quad (5.4)$$

Finally, the total signaling cost of RSVP-HoA over MIPv6 ($\Psi^{MIP}_{RSVP-HoA}$) is formulated by combining Equations 5.3 and 5.4 as follows:

$$\begin{aligned}\Psi^{MIP}_{RSVP-HoA} &= \Psi^{MIP_2} + \Psi^{RSVP_2} \\ &= N[\frac{(2l_{mc} + 2l_{mh} - l_{mcr} - 3 + 3\rho)\delta_B + \gamma}{tr}] \\ &\quad + N[\frac{4[(l_{mcr} - 1 + \rho)\delta_{R4} + (l_{mcr} - 1)\gamma]}{tr}. \end{aligned} \quad (5.5)$$

RSVP-HoA over Hierarchical Mobile IPv6

Since only local handoffs are considered, no registrations occur with the HA or the CN in HMIPv6 (R_{mh} = R_{mc} = 0). The only registration cost incurred is that to the MAP (R_{mm}). Similar to the RSVP-HoA signaling cost analysis for MIPv6, the total

signaling cost of RSVP-HoA over HMIPv6 ($\Psi^{HMIP}_{RSVP-HoA}$) is formulated as follows, using the number of hops between the MN and the MAP (l_{mm} instead of l_{mc}) and excluding R_{mh}:

$$\begin{aligned}\Psi^{HMIP}_{RSVP-HoA} &= N\frac{(2l_{mm} - l_{mcr} - 1 + p)\delta_B + \gamma}{tr} \\ &\quad + N[\frac{4[(l_{mcr} - 1 + \rho)\delta_{R4} + (l_{mcr} - 1)\gamma]}{tr}. \end{aligned} \tag{5.6}$$

RSVP-HoA over Fast Handovers for Mobile IPv6

For the signaling cost of RSVP-HoA over FMIPv6 ($\Psi^{FMIP}_{RSVP-HoA}$), Equation 3.11 is substituted into Equation 3.23 as follows:

$$\begin{aligned}\Psi^{FMIP}_{RSVP-HoA} &= \Psi^{FR} + \Psi^{FPT} + \Psi^{MIP_2} + \Psi^{RSVP_2} + \Psi^{RSVP}_{tunnel} \\ &= N[\frac{(5\rho + 3l_{on})\delta_B + 5\gamma}{tr}] \\ &\quad + N[\frac{P((l_{on} + \rho)\delta_D + 2\gamma + \beta)}{tr}] \\ &\quad + N[\frac{(2l_{mc} + 2l_{mh} - l_{mcr} - 3 + 3\rho)\delta_B + \gamma}{tr}] \\ &\quad + N[\frac{4[(l_{mc} + l_{on} - 1 + \rho)\delta_{R5} + (l_{mc} + l_{on} - 1)\gamma]}{tr}.] \end{aligned} \tag{5.7}$$

Similar to the signaling cost analysis of MARSVP over FMIPv6 presented in Section 4.5.1, the last term of the above equation, consisting of Ψ^{RSVP_2}andΨ^{RSVP}_{tunnel}, is being multiplied by δ_{R5}. This value is the average per hop transmission cost of the total of eight RSVP messages exchanged (four for the duplex RSVP session, and another four for the temporary RSVP tunnel between the two access routers). The average per-hop transmission cost of the first four RSVP messages has been calculated earlier as $\delta_{R4} = 149$. The remaining four RSVP messages are traditional Path and Resv messages (no added BU/BAck objects) and therefore have a per-

Table 5.3: Improvement in Signaling Cost.

Mobility Protocol	MARSVP	RSVP-HoA
MIPv6	9.4%	44.7%
HMIPv6	11.9%	11.9%
FMIPv6 (VoIP)	17.9%	45.4%
FMIPv6 (VIP)	26.7%	56.7%

hop transmission cost $\delta_R = 140$. The average of the eight RSVP messages (δ_{R5}) becomes:

$$\delta_{R5} = \frac{\delta_{R4} + \delta_R}{2} = \frac{149 + 140}{2} = 144.5.$$

Moreover, since the RSVP-HoA mechanism limits RSVP signaling to the COR, the total interruption in QoS is reduced and hence the temporary tunnel established by FMIPv6 is retained for a shorter period of time. This results in a fewer number of data packets being tunneled during handoffs. To obtain the exact number of packets (P_2) that get tunneled when using RSVP-HoA over FMIPv6, the simulation model described in Section 3.2.1 was used. The number of tunneled data packets per handoff was measured to be 3 on average for VoIP traffic ($P_{VoIP} = 3$), and 5 for VIP traffic ($P_{VIP} = 5$).

5.5.2 Results and Observations

The network-level performance of RSVP-HoA was analysed and compared against standard RSVP and MARSVP using the associated signaling cost models for the three mobility protocols. The parameter values presented in Table 3.3 were used to obtain numerical results, with the following RSVP-HoA attributes taken into account:

- The average per-hop transmission cost (δ_{R2}) of the four RSVP messages used in MIPv6 and HMIPv6 is 149.

- The average per-hop transmission cost (δ_{R3}) of the total of eight RSVP messages used in FMIPv6 (four for the RSVP session, and four for the temporary RSVP tunnel) is 144.5.

- The average number of tunneled data packets (P)in FMIPv6 using RSVP-HoA is 3 for VoIP traffic, and 5 for VIP traffic.

Mobile IPv6

Table 5.3 compares the signaling cost savings achieved using RSVP-HoA and MARSVP for each of the three mobility protocols. In the case of MIPv6, the signaling cost savings achieved using RSVP-HoA is 44.7%; compared to 9.4% when using MARSVP. This substantial improvement is a consequence of the new packet classification system (using the HomeAddress Option) which confines RSVP signaling to the COR. As a result, the RSVP-HoA mechanism saves RSVP message transmission and processing costs on the unchanged portion of the link ($COR \leftrightarrow CN$). Moreover, since the RSVP session is full-duplex, savings are made in both directions while the mobility signaling remains unchanged; thereby resulting in the 44.7% overall signaling cost savings for MIPv6.

Figure 5.10 illustrates the signaling costs incurred by the three RSVP mechanisms when deployed over a MIPv6 wireless network, as the number of nodes is increased from 1 to 20. Conventional RSVP (labeled *MIP_RSVP*) ranges from 200 to 4000, while MARSVP (labeled *MIP_MARSVP*) offers a slight improvement, ranging from approximately 180 to 3600. RSVP-HoA, however, offers the highest improvement, with signaling costs ranging from 110 to 2200 (labeled *MIP_RSVP-HoA*).

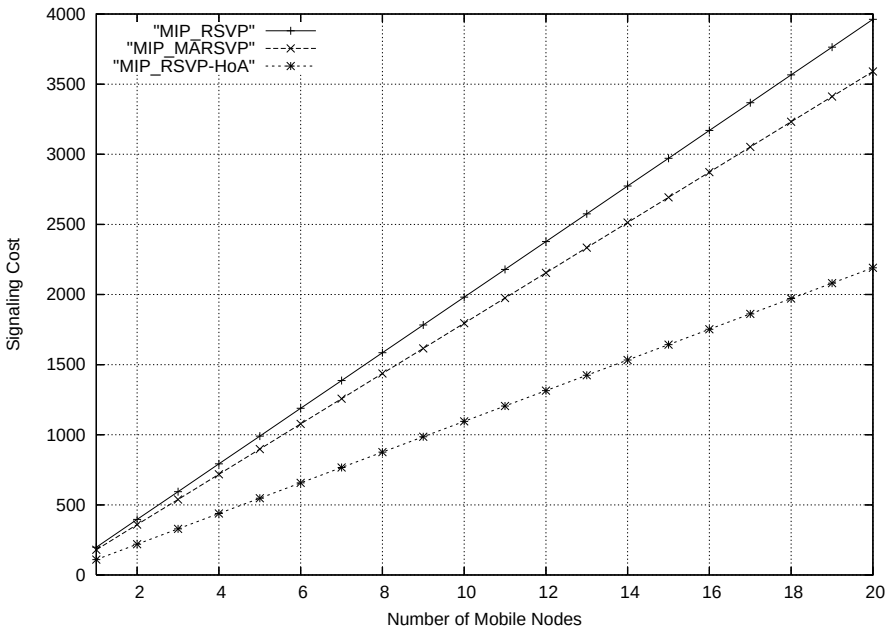

Figure 5.10: MIPv6 signaling cost using the different RSVP mechanisms.

Hierarchical Mobile IPv6

Similar to the results of the application-level performance (Section 5.4.1), RSVP-HoA offers the same signaling cost savings achieved by MARSVP (11.3%, as shown in Table 5.3). This is due to the simulation topology used (Figure 3.1) in which the COR is the MAP (consult Section 5.4.1 for a detailed explanation). Therefore, for a single-level hierarchy, RSVP-HoA and MARSVP operate in essentially the same manner and hence incur the same signaling costs.

In order to investigate the advantage of RSVP-HoA, larger HMIPv6 hierarchical topologies are considered. Table 5.4 presents the incurred signaling costs of the three RSVP mechanisms as the number of hierarchy levels is increased from one to ten. For MARSVP, signaling cost savings are slightly reduced (from 11.9%) for every added level of hierarchy, until it reaches 11.2% for a ten-level hierarchy. This is due to the additional transmission costs (for RSVP and mobility messages) and processing costs (RSVP message processing at intermediate nodes) for each

Table 5.4: Improvement in Signaling Cost according to HMIPv6 hierarchy depth.

Levels of Hierarchy	RSVP	MARSVP	%	RSVP-HoA	%
1	990	873	11.9%	873	11.9%
2	1083	956	11.7%	890	17.8%
3	1177	1040	11.6%	908	22.3%
4	1270	1123	11.5%	926	27.1%
5	1363	1207	11.5%	944	30.8%
6	1457	1290	11.4%	962	34%
7	1550	1374	11.4%	979	36.8%
8	1643	1458	11.3%	997	39.3%
9	1737	1541	11.3%	1015	41.6%
10	1830	1625	11.2%	1033	43.6%

additional hop introduced towards the MAP.

The RSVP-HoA mechanism, however, operates in a contrary manner: signaling cost savings increase considerably as the number of hierarchy levels is increased (11.9% at single-level hierarchy, and up to 43.6% for a ten-level hierarchy). This is because the higher the number of hierarchy levels, the further away the MAP is positioned from the MN. According to the RSVP-HoA mechanism, RSVP signaling is limited to the COR which is always one hop away (rather than perform it across multiple hops towards the MAP, as is the case with RSVP and MARSVP).

This can be better understood by examining Figure 5.11 which illustrates the signaling costs incurred by the three RSVP mechanisms when deployed over a HMIPv6 wireless network as the number hierarchy levels is increased from 1 to 10. As can be observed, MARSVP's signaling cost (labeled *HMIP_MARSVP*) increases linearly with conventional RSVP (labeled *HMIP_RSVP*). Note that even though signaling cost savings are higher at a ten-level hierarchy ($1830 - 1625 = 205$) than at

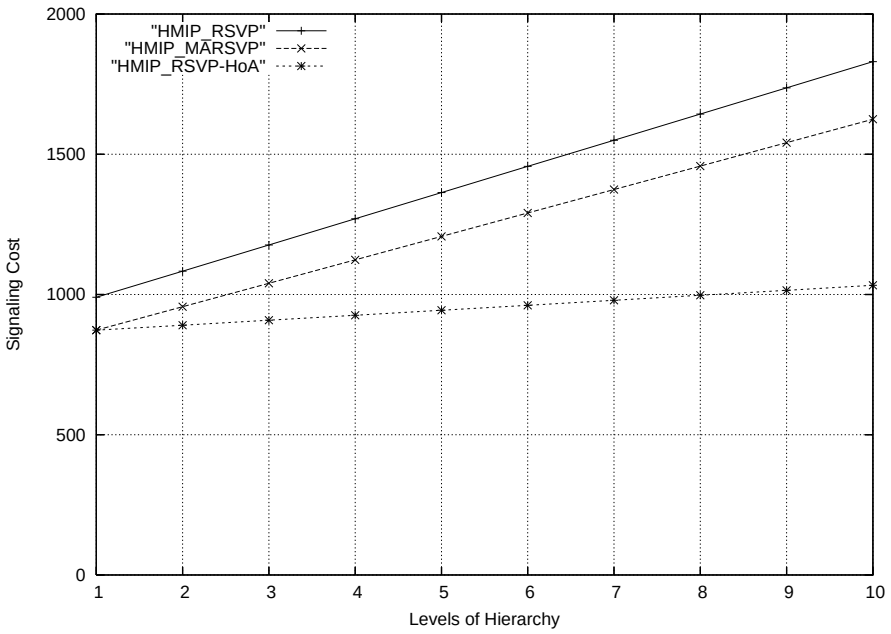

Figure 5.11: HMIPv6 signaling cost as the number of hierarchy levels is increased.

single-level hierarchy ($990 - 873 = 117$), the percentage signaling cost savings are slightly lower (11.2% as opposed to 11.9%).

In the case of RSVP-HoA (labeled *HMIP_RSVP-HoA*), the larger signaling cost savings are apparent in RSVP-HoA's signaling cost plot which has a much lower slope than those of RSVP and MARSVP. The reason for this is that RSVP-HoA always limits RSVP signaling to the COR (typically one hop away), regardless of the total number of hops to the MAP (i.e. number of hierarchy levels). Therefore the gradual increase in signaling cost of RSVP-HoA as the number of hierarchy levels is increased is solely due to the transmission cost of the BU/BAck message pair to the MAP (increases by a single hop for every added level of hierarchy). Note, that unlike RSVP signaling, this does not include any additional processing cost since mobility signaling is only processed at the end nodes.

Table 5.5: Number of Tunneled Packets (P) in a FMIPv6 Handoff.

$QoSMechanism$	TI_{QoS}	P_{VoIP}	P_{VIP}
RSVP	732 ms	7	13
MARSVP	528 ms	5	9
RSVP-HoA	312 ms	3	5

Fast Handovers for Mobile IPv6

For FMIPv6, higher signaling cost savings are achieved (45.4% for VoIP traffic, and 56.7% for VIP traffic) than MIPv6's signaling cost savings of 44.7%. This is because FMIPv6 not only saves the transmission and processing costs of RSVP messages from the COR to the MAP, but the smaller value of TI_{QoS} (312 ms) results in less data packets (P) being tunneled during handoffs (3 for VoIP traffic and 5 for VIP traffic). By observing the values of TI_{QoS} and P for the different RSVP mechanisms and traffic types (Table 5.5), an association can be made with the relative signaling cost plots for VoIP and VIP traffic presented respectively in Figures 5.12 and 5.13: As can be observed by comparing the two figures, slightly better signaling cost savings are achieved when using VIP traffic than VoIP traffic. This is due to the VIP traffic's larger packet size which has a higher impact on signaling cost savings of FMIPv6 when considering tunneling and buffering costs. For example, 13 VIP packets are tunneled when using RSVP compared to 5 for RSVP-HoA. This difference in 8 VIP packets reduces the signaling costs accordingly since the larger VIP packets naturally incur higher transmission and buffering costs than VoIP packets. This is mirrored on to the respective signaling cost plots of VoIP and VIP traffic (Figures 5.12 and 5.13), where the latter inflicts a slightly larger impact on signaling cost savings.

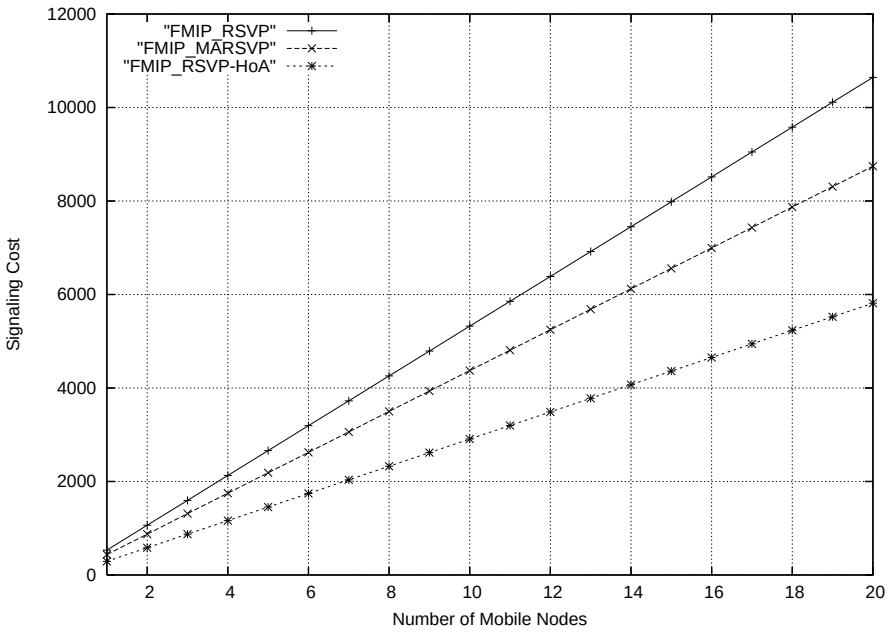

Figure 5.12: FMIPv6 signaling cost using VoIP traffic.

5.5.3 Analysis and Discussion

Several analytical models were derived to evaluate RSVP-HoA's performance at the network level. The signaling cost analysis conducted in Section 3.3.2 was used as a performance benchmark to measure the improvement brought about by the proposed mechanism. RSVP-HoA was found to provide the largest signaling cost savings when deployed over a FMIPv6 wireless network (45.4% for VoIP traffic, and 56.7% for VIP traffic). The difference in performance of the two traffic sources was due to the different number of packets being tunneled during handoffs, in addition to the packet size of each. In MIPv6 wireless networks, on the other hand, RSVP-HoA provided signaling cost savings of 44.7%, while it provided the same signaling cost savings as MARSVP (11.9%) when deployed over a HMIPv6 wireless network. This was due to simulation topology used were the MAP was the COR.

To further assess RSVP-HoA's improvement in HMIPv6 wireless networks,

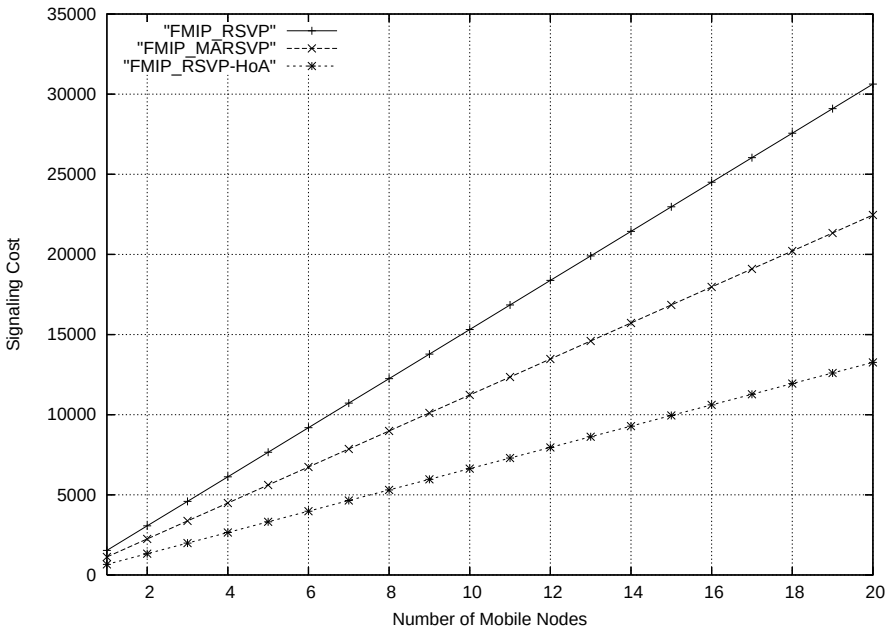

Figure 5.13: FMIPv6 signaling cost using VIP traffic.

larger HMIPv6 hierarchical topologies were considered. As the number of hierarchy levels increased, RSVP-HoA provided greater signaling cost savings while MARSVP's signaling cost savings were slightly reduced. This is because reservations were renewed only to the COR (one hop) whereas MARSVP renewed them to the MAP (multiple hops, depending on hierarchical depth) and hence savings were made in transmission and processing costs.

5.6 Conclusion

A new RSVP mechanism (RSVP-HoA) for wireless mobile networks was presented and evaluated in this chapter. RSVP-HoA presents a new classification mechanism for RSVP in which routers are configured to classify flows based on the Home Address option in the MIPv6 destination options header. Intermediate RSVP routers are therefore able to correctly identify an RSVP flow, even after a MN changes its CoA. Moreover, using this mechanism a crossover router (COR) can detect the

changed portion of the end-to-end RSVP session and confine RSVP signaling to it. As a result, the RSVP re-establishment time and network signaling costs are substantially reduced.

According to the MIPv6 specification [JP04], IP packets sent by a roaming MN should explicitly include a Home Address option in the destination options extension header. RSVP-HoA utilises this readily available information and thus no changes are required to the mobility protocols. On the other hand, changes are still required to RSVP to allow routers to inspect the home address option and limit RSVP signaling to the changed portion.

The proposed mechanism was evaluated for application and network-level performance, and was compared against standard RSVP and the MARSVP mechanism presented in Chapter 4. Simulation results reveal that, in a MIPv6 network, the total interruption in QoS can be reduced by 57.4% when using RSVP-HoA, compared to 27.9% for MARSVP. This significant improvement is due to RSVP signaling being confined to the COR by the RSVP-HoA mechanism, which reduces the associated RSVP signaling delay accordingly.

For HMIPv6 however, RSVP-HoA provides the same TI_{QoS} value of 168 ms as MARSVP (12.5% improvement). This is due to the single-level simulation topology used in which the MAP (node N4) is the COR. Further simulations were conducted using higher levels of hierarchy to investigate RSVP-HoA's advantage over MARSVP. RSVP-HoA was found to provide better improvements in TI_{QoS} than MARSVP as the number of hierarchy levels was increased: At a 5-level hierarchy, standard RSVP results in a TI_{QoS} of 433 ms, while MARSVP is 326 ms (24.7%) and RSVP-HoA 248 ms (42.7%). In the case of FMIPv6, however, no further improvement in TI_{QoS} was achievable. This is due to FMIPv6's anticipation of handoffs and the associated tunneling mechanism used which ensures a MN's connectivity during the execution of the actual handoff.

When examined at the network-level, RSVP-HoA produces significant signaling cost savings for MIPv6 with 44.7% improvement compared to 9.4% when using

MARSVP. This is because the RSVP-HoA mechanism confines RSVP signaling to the COR and hence saves RSVP message transmission and processing costs on the unchanged portion of the link ($COR \leftrightarrow CN$). For HMIPv6, however, RSVP-HoA once again produced similar results to MARSVP due to the single-level hierarchal topology used. Further signaling cost analysis were conducted using larger levels of hierarchy to investigate RSVP-HoA's advantage over MARSVP. RSVP-HoA was found to produce better signaling cost savings as the number of hierarchy levels was increased, while MARSVP slightly deteriorated.

The highest signaling cost savings were achieved when using RSVP-HoA over FMIPv6 wireless networks with 45.4% improvement when using VoIP traffic, and 56.7% for VIP traffic. This is because FMIPv6 not only saves the transmission and processing costs of RSVP messages from the COR to the MAP, but the smaller value of TI_{QoS} (312 ms) results in less data packets (P) being tunneled during handoffs (3 for VoIP traffic and 5 for VIP traffic).

After examining the proposed RSVP-HoA mechanism for application and network-level performance, it was found to be the best alternative compared to MARSVP and standard RSVP: RSVP-HoA delivers the best end user experience (as shown in the lower value of TI_{QoS} and the associated MOS plots). RSVP-HoA also proves to be the lightest in terms of signaling costs incurred on the network. The only drawback of the RSVP-HoA is in its implementation which requires modifications to be made to all RSVP-enabled nodes. This, however, is justifiable by the significant signaling cost savings and application-level performance achieved. Moreover, the new packet classification method is considered a minor modification which could be introduced as a firmware upgrade to existing RSVP routers.

Chapter 6

Conclusion and Future Work

6.1 Summary

The focus of this research was to develop reservation models for improved resource allocation in wireless All-IP networks in order to meet the QoS requirements of real-time applications whilst maintaining resource utilisation at high levels. The following requirements were taken into account in the design process of the proposed models:

- Interoperable with Mobile IP (specifically with IP-in-IP encapsulation).
- Minimise TI_{QoS} in the event of a handoff.
- Localise RSVP signaling to the affected sections of the end-to-end path.

Two models were proposed as extensions to the existing RSVP standard: namely, MARSVP and RSVP-HoA. The first adheres to the current RSVP standard (RFC 2205) [BZB+97] and is hence backwards compatible with it; while the latter significantly reduces TI_{QoS} at the expense of strict adherence to the existing RSVP standard.

Depending on a service provider's desired level of complexity, any of the two proposed reservation models could be implemented to improve the performance

Table 6.1: Comparison of QoS Mechanisms.

Proposal	*Mobility Prediction*	*New Messages*	*Modifications*	*New Nodes*
MRSVP	Y	9	MN and CN	Proxy Agents
SMRP	N	3	All network entities	-
WLRP	Y	3	MN and MA	-
ARSVP	N	1	MN and all Internal Routers	-
RSVP-MP	N	0	DAT on all RSVP packets	Mobility Proxy
QoS-Conditionalised Handoff	N	0	MN and all Internal Routers	-
QoS-Aware Handoff	Y	2	MN and all Internal Routers	EVL Processors
MARSVP	N	0	MN and CN	-
RSVP-HoA	N	0	MN and all Internal Routers	-

in wireless networks: While MARSVP provides a simple and efficient alternative, RSVP-HoA delivers superior application-level performance to the end user, and at the same time imposes fewer signaling costs on the network. This, however, is achieved at the expense of requiring changes to be made to all RSVP-capable nodes in the network.

Finally, Table 6.1 compares MARSVP and RSVP-HoA with the surveyed Qos mechanisms from Chapter 2. In order to identify the most suitable protocol for a

head-to-head comparison with MARSVP and RSVP-HoA, an elimination approach is used: Since both proposed protocols do not introduce any additional nodes, protocols that do require the installation of new nodes are eliminated (MRSVP, RSVP-MP, and QoS-Aware Handoff). Furthermore, both proposal do not introduce any new messages and hence SMRP, WLRP and ARSVP are also eliminated. This leaves QoS-Conditionalised Handoff (QoS-CH) as the nominated proposal for comparison.

As outlined in Section 2.6.3, QoS-CH performs mobility and QoS signaling as a single functional block, thus resulting in a similar theoretical reduction in TI_{QoS} to that of MARSVP. Having said that, MARSVP still represents a more appealing choice due to the following features:

1. MARSVP provides a similar improvement in TI_{QoS} to that of QoS-CH, while limiting modifications to the end nodes.

2. QoS-CH explicitly assumes that the coverage areas of wireless subnets overlap, while MARSVP does not .

3. QoS-CH is designed over a HMIPv6 network; while MARSVP can operate over a MIPv6, HMIPv6 or a FMIPv6 network.

On the other hand, when comparing RSVP-HoA to QoS-CH, a similar number of modifications is required (MN and all routers). Nonetheless, RSVP-HoA has the advantage of confining QoS signaling to the changed portion while QoS-CH confines it to the MAP. When considering multiple-hierarchy levels in a HMIPv6 network (Table 5.2), RSVP-HoA delivers superior application level performance with a reduced network signaling cost. Furthermore, RSVP-HoA (like MARSVP) neither assumes overlapping wireless subnets, nor does it assume a HMIPv6 network.

6.2 Thesis Contributions

This section recaps the major contributions of the work presented in this thesis.

Chapter 2 presented a critical review of the research community's related work in the area of QoS provisioning mechanisms in wireless IP networks. Chapter 3 introduced a performance analysis study to investigate the interaction of RSVP and Mobile IPv6 (including its extensions). The analysis framework was comprised of a simulation-based section to measure application-level performance, and a signaling cost analysis section to measure network-level performance.

Chapter 4 presented a mechanism for enhancing RSVP performance over Mobile IPv6 and its extensions, called Mobility Aware Resource Reservation Protocol (MARSVP). The key concept of MARSVP is to convey mobility-specific information using newly defined RSVP objects embedded in existing RSVP messages. This allows a single message exchange to establish both IP-level connectivity as well as QoS guarantees on the new link. The appealing attribute of MARSVP is that it requires minimal changes to end nodes and is hence compatible with the current RSVP standard.

Finally, a new packet classification mechanism for RSVP (called RSVP-HoA) was proposed in Chapter 5. According to RSVP-HoA, routers are configured to classify flows based on the home address option in the MIPv6 destination options header. Through this approach, intermediate RSVP routers are able to correctly identify an RSVP flow, even after a MN changes its CoA. Moreover, a crossover router (COR) using this mechanism can detect the changed portion of the end-to-end RSVP session and confine RSVP signaling to the changed nodes.

These mechanisms will improve the QoS provisioning in next-generation wireless networks. Efficient QoS provision remains a crucial factor in enabling future 4G mobile technology which will be based on an all-IP core network. This thesis presented a detailed outline of the major issues encountered in such an environment and conducted studies provided feasible solutions towards a fully integrated, efficient multi-service network.

6.3 Recommendations for Future Work

Throughout the course of this thesis various limitations and opportunities for improvement were encountered, some of which had to be overlooked due to time restrictions. This section presents these observations as recommendations for future research work.

6.3.1 Simulation Accuracy

Links

The link between nodes N1 and N4 (Figure 3.1) was configured to model the Internet by introducing an 80ms delay. Although this simple approach provides viable results, an Internet topology generator such as Inet [WJCJ] should be used for more realistic results. The Inet generator creates random network topologies with characteristics similar to those of the Internet from November 1997 and beyond. A drawback however is that it generates a minimum of 3037 nodes. This would consume significant processing power and may make each individual simulation run last for a considerably long time.

Mobility

Each simulated RSVP session consists of a mobile end node (MN) communicating with its respective fixed node (CN). It would be interesting if both end nodes were mobile, undergoing random handoffs. This would not only provide more realistic results (since we expect to place mobile to mobile calls in real life) but also test the robustness of the proposed models.

Data Traffic

A VoIP traffic generator was used to simulate voice calls over the Internet. However, a more realistic approach would be to sample real voice conversations and use

them as data traffic in the simulation environment. For VIP traffic on the other hand, it was noted that the MPEG-4 inter-frame dependencies had a significant impact on application-level performance (Section 3.2.3). A feasible solution would be to implement a selective packet dropping mechanism at RSVP routers that distinguishes the different frame types of an MPEG-4 data stream and assigns different levels of priority accordingly: an I-frame, for example, would be assigned the highest level of priority and would therefore have a lower probability of being dropped.

6.3.2 Further Research

During the progression of this thesis, the Next Steps In Signaling (NSIS) [HKLdB05] framework was proposed. NSIS is argued by many researchers to have the potential of replacing RSVP due to the following reasons [Blo]:

- **Transport:** RSVP is transmitted over UDP while NSIS can be transmitted over TCP or UDP.
- **Reservation Model:** RSVP is initiated by the receiver, while NSIS could be initiated by either the sender or the receiver.
- **Multicasting:** Unlike RSVP, NSIS does not support multicasting. This reduces the complexity of applications, which mostly consist of unicast communications.
- **Two-way Reservations:** NSIS enables two-way reservations by performing bindings to sessions in both directions.
- **QoS Models:** NSIS can be used within any QoS model signaling model such as DiffServ [BBC+98] or IntServ [BCS94]. In contrast, RSVP is closely tied to the IntServ architecture.
- **Mobility:** RSVP identifies a session using the 5-tuple flow identifier (source and destination IP addresses, source and destination port numbers, and pro-

tocol ID). NSIS however, uses a random session identifier and therefore does not rely on IP addresses that may change due to mobility. This makes mobility support in NSIS much easier.

- **Security:** In RSVP, various security issues have been pointed out in the initial specification, but were later on addressed as extensions. In NSIS however, security is built-in.

Further research should be conducted for the investigation of the interaction of NSIS with the different mobility protocols presented in this thesis. Such a work would provide a quantifiable comparison between RSVP and NSIS; and prove (if it is the case) that NSIS is a suitable candidate for replacing RSVP in future.

Appendix A

NS-2 and RSVP Simulation Models and Modifications

A.1 NS-2 Simulation Package for Performance Analysis

Performance analysis is considered an integral part of any computer or communications research undertaking. Researchers develop models to evaluate various aspects of an actual system. This may range from the investigation of the accuracy of the developed model itself, to analysing the performance of a proposed protocol deployed using the model. This section first presents a comparison of analytical and simulation models, followed by a brief introduction to the Network Simulator 2 (ns-2).

A.1.1 Analytical and Simulation Models

Performance analysis can be conducted using either analytical or simulation models. Analytical models consist of a set of mathematical models (with their associated approximations and assumptions), while simulation models consist of a computer

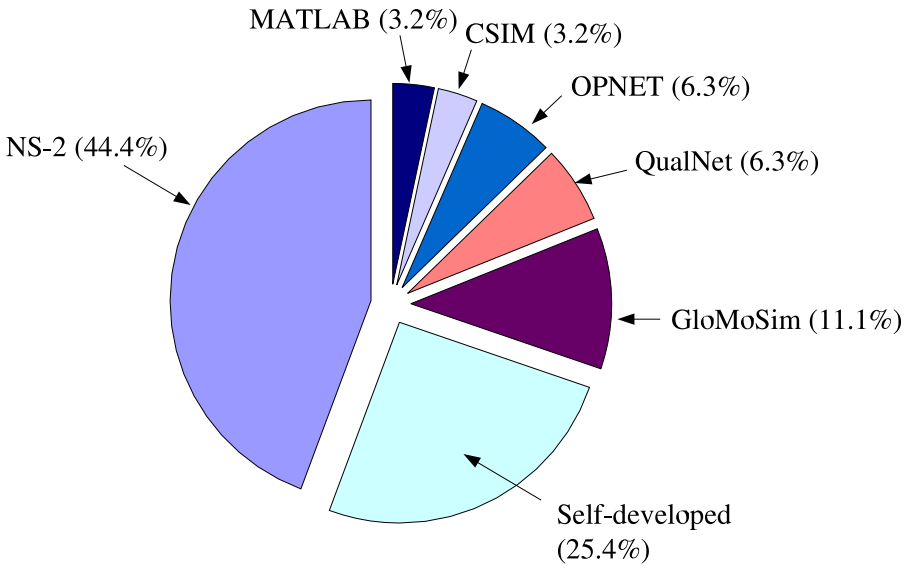

Figure A.1: Simulation package usage in MobiHoc survey.

program developed to mimic various aspects of the actual system.

Traditionally, analytical models were used more commonly since simulation models took a relatively longer time to build and execute. In 1978, the prevailing perception of simulations was lead by Kobayashi: *It is quite often found, however, that a simulation model takes much longer to construct, requires much more time to execute, and yet provides much less information than the model writer expected* [Kob78]. However, as the studied systems became larger and more complex, analytical models require making unrealistic assumptions and approximations [BGdMK06].

With the phenomenal evolution of computer processing power and data storage, performance analysis shifted in favour of today's accurate and large scale simulation models. Nowadays researchers have a broad choice of simulation packages, both commercial and freeware, for modeling complex computer networks. A survey conducted on the usage of simulation models in the ACM International Symposium on Mobile and Ad Hoc Networks (MobiHoc) from 2000 to 2004 [KCC05], revealed that 75.7% of the published papers used simulations to present research results. As

can be observed from Figure A.1 the most popular simulation package of choice was NS-2 (44.4%) followed by GloMoSim (11.1%); whereas 24.5% of authors opted to develop their own simulation code.

Upon surveying the various simulation packages available, NS-2 was selected as the simulator of choice for the experiments presented in this thesis. This decision was based on the following features:

- NS-2 is open-source [1].
- At the time of this research, NS-2 was the only open-source simulation package with both RSVP and Mobile IP modules readily available.
- NS-2 is used extensively in academia, and is supported by a large group of researchers.

A.1.2 Network Simulator 2

NS-2 is discrete-event [2] for applications in networking. It performs two main tasks, and is consequently based on two languages: an object oriented simulator, written in C++, and an interpreter, written in OTcl (Object-oriented Tool Command Language) to execute user's command scripts [AJ02].

The compiled C++ code achieves high efficiency, and is therefore best suited for detailed definition and operation of protocols where run-time speed is critical. On the other hand, network topologies and parameters are variables that are defined at the beginning of a simulation run. For these dynamic configurations, a more friendly command language is more appropriate (OTcl).

The main features of NS-2 can be summarised as follows [BGdMK06]:

[1] Publicly available source code permitting users to use, change, and improve the software in a collaborative manner.

[2] In a discrete-event simulator, the operation of a system is represented as a chronological sequence of events.

- It provides canned sub-models for several network protocols (such as TCP and UDP), router queue management mechanism (tail-drop, RED, CBQ) and routing algorithms (Dijkstra).
- It provides several traffic sources (Telnet, FTP, and CBR).
- It implements some MAC protocols for LAN multicast protocols.
- It contains a simulation event scheduler and a large number of network objects (such as routers and links) which are interconnected to form a network.
- Events can be visualized through a graphical interface and/or data logged in a file for post-simulation analyses.

A.2 Hierarchical Addressing Problem

When running Marc Greis' simulation model of "RSVP/ns"[Mur] over Mobile IP (incorporated in NS-2 simulation package by default) a conflict occurs due to the type of addressing used. RSVP/ns was developed based on ns-2's default flat addressing system in which nodes are assigned integer values for their addresses. For consistency, these node addresses are assigned according to the nodes' respective node numbers (i.e. node 1 is assigned an address of 1, node 2's address is 2 and so on). Mobile IP models however, use a hierarchical addressing system similar to traditional IP. The number of hierarchical levels can be set to four to resemble IPv4 addresses (e.g. 192.168.0.1) or six for IPv6.

Since most of ns-2's code assumes the default flat addressing system (i.e. node id = node address), RSVP/ns was developed to pass node id's (integer values) to all its methods. However, when Mobile IP's hierarchical addressing is used, routing problems occur. For example, in a Mobile IP scenario an RSVP sender might have a node id of 1 and an IP address of 192.168.0.1. According to the RSVP/ns code, the sender would record its node id as the sender address in the generated Path mes-

sage. At the receiving end, the RSVP receiver would reply with a Resv message to the sender address stored in the Path message. Although this would work for flat addressing (node id = node address = 1), in hierarchical addressing the destination would be unreachable since the actual sender address would be a hierarchical address (e.g. 192.168.0.1) and not the integer value of 1. As a result, no reservations would be established due to the different addressing systems used by RSVP/ns and Mobile IP simulation models.

A.2.1 Converting Hierarchical and Flat Addresses

To resolve this issue, two new methods have been added: `hier_addr` (returns a node's hierarchical address), and `flat_addr` (returns a node's flat address).

```
int RSVPAgent::flat_addr(nsaddr_t flat) {
 Tcl& tcl = Tcl::instance();
 tcl.evalf("[Simulator instance] id_by_addr %d %d",flat, node_number_);
 int n1=atoi(tcl.result());
 if (n1==0 && flat != 0)
  return flat;
 else
  return n1;
}
int RSVPAgent::hier_addr(int id) {
 Tcl& tcl = Tcl::instance();
 tcl.evalf("[Simulator instance] addr_by_id %d %d",id, node_number_);
 int n1=atoi(tcl.result());
 if (n1==0 && id != 0)
  return id;
 else
  return n1;
}
```

The two methods are used as interfaces between RSVP/ns and Mobile IP (Figure A.2)

- **`flat_addr()`**: Used to convert a hierarchical address to its flat equivalent (which is in essence the node's id) for RSVP/ns's internal processing, such as creating or deleting PSB and RSB lists. This is done by adding the method

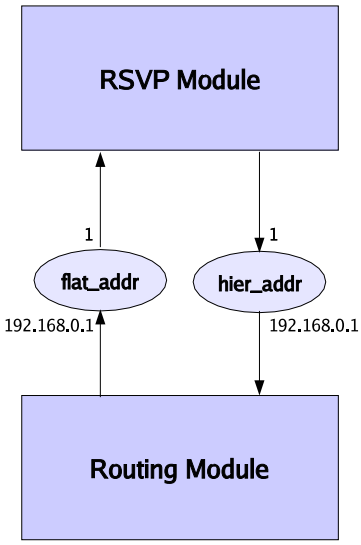

Figure A.2: Address conversion using `hier_addr` and `flat_addr` .

node_id() whenever `here_.addr_`[3] is called. For example:

```
if ((p = find_psb(s, flat_addr(here_.addr_)))
```

- **`hier_addr()`**: Used to convert a flat address to its "actual" hierarchical address for routing purposes. The flat address is retrieved from the RSVP/ns code (stored in either a PSB,RSB or session list) and converted using `hier_addr` method to the equivalent hierarchical address. This address can then used to send packets using `dst_.addr_`[4]:

```
dst_.addr_ = hier_addr(s->s->get_dest())
```

A.2.2 Retrieving Node IDs using oTcl

Since node numbers are generated at the beginning of a simulation run (i.e. when running `common/simulator.cc`), the `hier_addr` and `flat_addr` methods of

[3] `here_.addr_`: A default ns-2 method that returns a node's actual address (in Mobile IP's case, the hierarchical address).

[4] `dst_.addr_`: A default ns-2 method used to set a packet's destination address for routing purposes.

Appendix A.2.1 can only gain access to node numbers by invoking OTcl scripts. Two methods, named `id-by-addr` and `addr-by-id` , have been added accordingly to the simulation code in order to invoke the OTcl commands `node_list_[]` and `nodeid()`.

```
if (strcmp(argv[1], "id_by_addr") == 0) {
 int address=atoi(argv[2]);
 int num_n= atoi(argv[3]);
 for (int i=0; i<((num_n)+0); i++) {
  if(nodelist_[i]->address() == address) {
   tcl.resultf("%d",nodelist_[i]->nodeid());
   return TCL_OK;
   }
  }
 }

if (strcmp(argv[1], "addr_by_id") == 0) {
 int id=atoi(argv[2]);
 int num_n= atoi(argv[3]);
 for (int i=0; i<((num_n)+0); i++) {
  if(nodelist_[i]->nodeid() == id) {
   tcl.resultf("%d",nodelist_[i]->address());
   return TCL_OK;
   }
  }
 }
```

A.3 Packet Tracing Problem

To allow the user to analyse simulations efficiently, NS simulator stores all network events (trace data) in a file to be post-processed and analysed. In order to distinguish RSVP messages and log them accordingly (as opposed to logging them as generic or UDP packets), modifications have to be made to the tracing code `(trace/cmu-trace.cc)`.

A.3.1 Tracing RSVP messages

To distinguish the different RSVP messages, switch cases have to be added accordingly in the method `CMUTrace::format()`.

```
switch(ch->ptype()) {
 case PT_RSVP_PATH:
  format_ip(p, offset);
  break;
 case PT_RSVP:
  format_ip(p, offset);
  break;
 case PT_RSVP_RESV:
  format_ip(p, offset);
  break;
 case PT_PATH_TEAR:
  format_ip(p, offset);
  break;
 case PT_RESV_TEAR:
  format_ip(p, offset);
  break;
 case PT_RESV_ERR:
  format_ip(p, offset);
  break;
 case PT_RESV_CONF:
  format_ip(p, offset);
  break;
```

A.4 Validation

To validate the operation of the modified RSVP/ns code, the network topology depicted in Figure A.3 is used. The main objective is to create contention in a wireless environment and use RSVP to reserve sufficient bandwidth for one of the contending traffic flows.

The network topology is comprised of three sender nodes ($N1$, $N2$ and $N3$) connected via nodes $N4$ and $N5$ to two wireless subnets (serviced by $N6$ and $N7$). Three wireless receiver nodes (MN_1, MN_2 and MN_3) reside in their shared home subnet. The three sender nodes transmit 500 kbps UDP data streams to their respective receiving wireless nodes.

All wired links have a link capacity of 1 Mbps, creating a bottleneck at the $N4$-$N5$ link since three 500 kbps data streams contend for a total link capacity of 1 Mbps. An RSVP session is established between $N1$ and MN_1 at the beginning of the simulation run, and at t = 100 s, MN_1 moves from its Home Network (HoA =

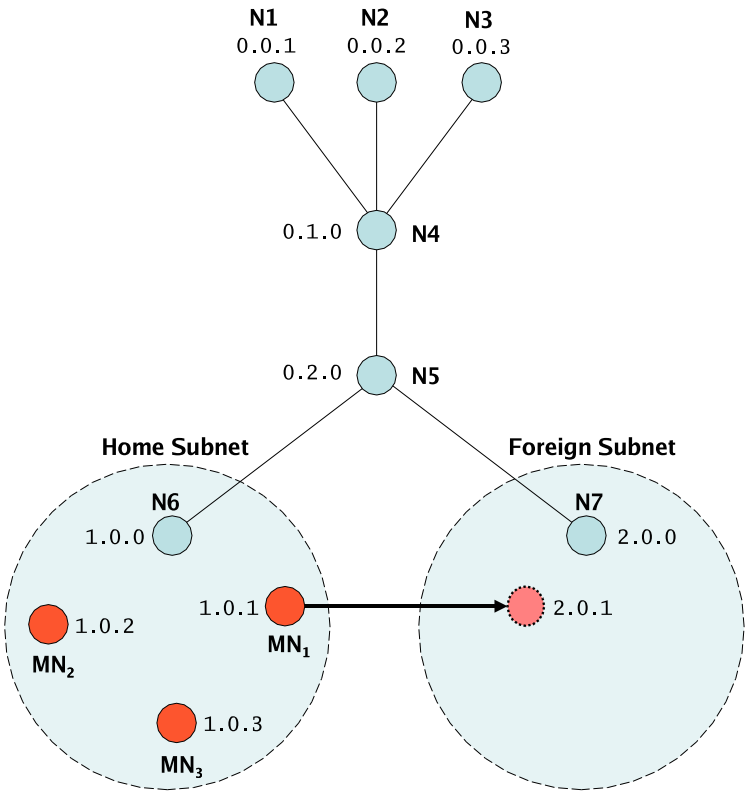

Figure A.3: Test Scenario.

1.0.1) and roams into the Foreign Network (CoA = 2.0.1).

Note that the two subnets are deliberately placed further apart in order for their respective coverage areas not to overlap. This creates an extended period of time in which MN_1 would be temporarily disconnected as it moves from one subnet to another.

A.4.1 Relinquishing Reservations

The simulation was run for a total of 200 seconds and the received throughput at each of the mobile nodes was measured. As can be observed from Figure A.4, MN_1 maintains a throughput of about 500 kbps (reserved bandwidth using RSVP) while MN_2 and MN_3 share the remaining 500 kbps link capacity amongst themselves (best-effort traffic) resulting in an average throughput of 250 kbps each.

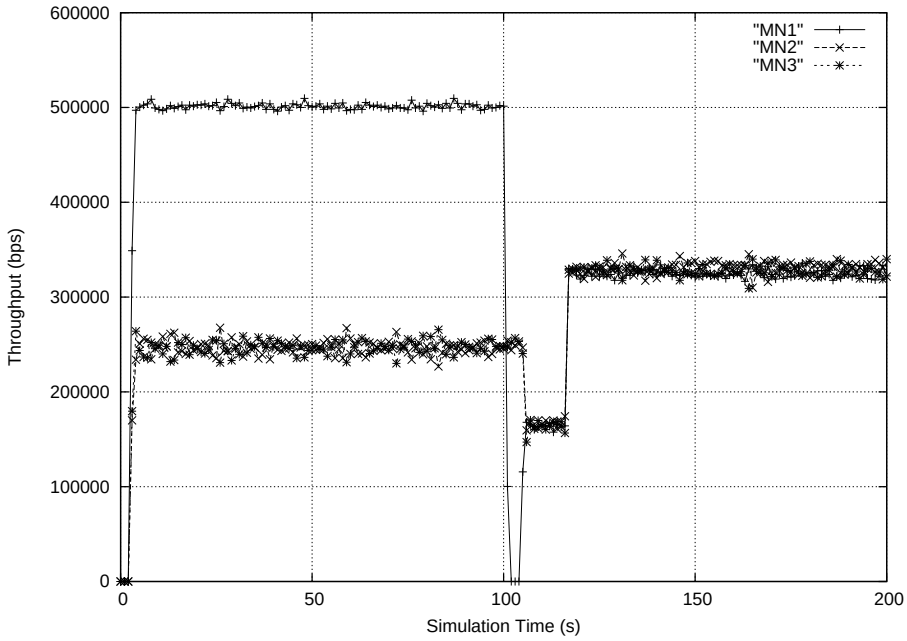

Figure A.4: Throughput of the three mobile receiving nodes without re-establishing the reservations for MN_1.

At t = 100 s, MN_1 starts moving to the foreign subnet (its throughput is effectively zero for about 3 seconds). However, MN_2 and MN_3 remain stationary at the home network and therefore continue to receive their respective 250 kbps throughput. Once MN_1 reconnects at the foreign network, its throughput reduces considerably from its pre-handoff level of 500 kbps to around 166 kbps temporarily. MN_2's and MN_3's respective throughputs are also reduced temporarily to 166 kbps.

The reason for this behaviour has been outlined theoretically in Section 5.2.1 and is now embodied in this simulation scenario as follows: The reservation for MN_1 held at the $N4$-$N5$ link was setup using MN_1's home address ($1.0.1$) whereas after the handoff, MN_1 uses a different IP address (CoA = $2.0.1$). As a result, the 500 kbps would remain allocated for MN_1's old IP address ($1.0.1$) even though in reality it would not be able to gain access to this bandwidth due to the IP address mismatch. Consequently, all the three data streams would share the remaining bandwidth (500 kbps) amongst themselves (166 kbps each).

After around 13 seconds, MN_1's old reservation is automatically relinquished since no Path/Resv message pairs are exchanged to refresh the reservation (as outlined in Section 2.2). Note that the default value for the reservation refresh timer is 20 s. This means that the last Path/Resv refresh message pair were sent at around 7 seconds before the handoff occurred. Since resources have been freed up at the $N4$-$N5$ link, all three data streams share the 1 Mbps link capacity amongst themselves (333 kbps each).

A.4.2 Re-establishing Reservations

While the scenario presented in Appendix A.4.1 examined how the simulation models default RSVP behaviour, the scenario presented in this section tests the simulation's ability to re-establish reservations for a mobile node after it completes its handoff.

In order to perform this, MN_1 is configured to explicitly relinquish its old reservations when a handoff occurs by sending a ResvTear message. Once MN_1 reconnects at the foreign subnet, CN_1 is configured to re-establish reservations at foreign subnet using MN_1's new CoA ($2.0.1$) as the destination address (since it is notified of MN_1's CoA through the BU message).

As can be observed from Figure A.5, at t = 100 s, MN_1's throughput is reduced to zero (handoff occurs) while MN_2's and MN_3's throughput increases to 500 kbps. This is due to MN_1 explicitly tearing down its old reservation, thereby freeing up resources at the $N4$-$N5$ link.

Once MN_1 reconnects at the foreign subnet, all three mobile nodes share the available bandwidth (333 kbps each) for a small period of time while the reservations are re-established for MN_1 using its new CoA. As soon as resources are reserved for MN_1, all three mobile nodes return to their respective pre-handoff throughputs (MN_1 = 500 kbps, MN_2 = MN_2 = 250 kbps).

Note that for the plot presented in Figure A.5, reservation re-establishment has

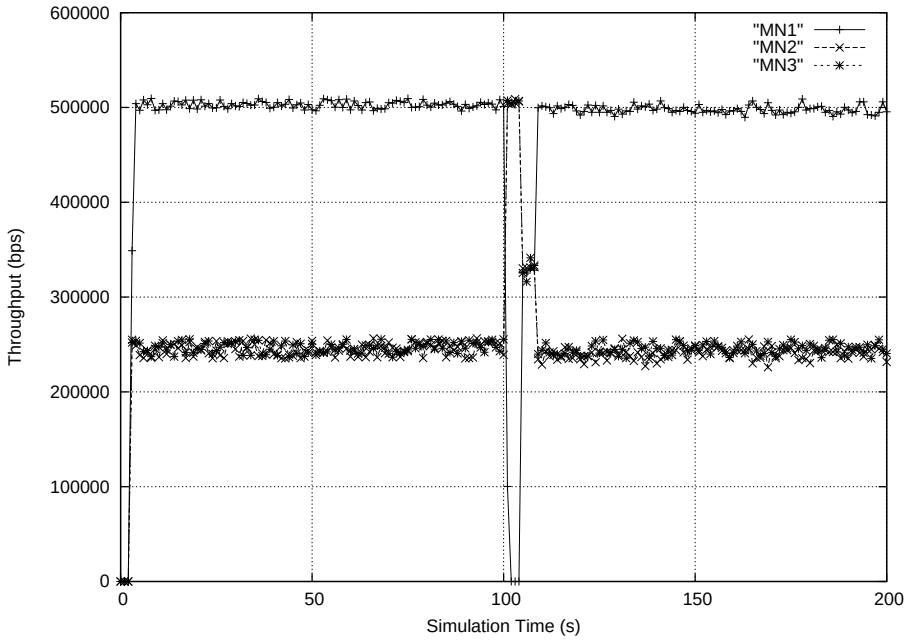

Figure A.5: Throughput of the three mobile receiving nodes with re-establishment of reservations for MN_1.

been deliberately delayed for 3 seconds (CN_1 waits for 3 s before sending the new Path message to MN_1). This is done for testing and visualisation purposes, since the default behaviour (CN_1 immediately send a new Path message once it receives the BU) would be barely notable in the plot (typically less than 700 ms of best-effort).

Bibliography

[3GP] 3GPP2. IS-856, High Rate Packet Data Air Interface Specification. URL reference: http://www.3gpp2.org.

[AJ02] E. Altman and T. Jimenez. Lecture Notes: NS Simulator for Beginners, 2002. URL reference: http://www-sop.inria.fr/mistral/personnel/Eitan.Altman/COURS-NS/n3.pdf.

[BA03] B. Baykal and O. B. Akan. A QoS-Aware Handoff using RSVP in Next Generation Wireless Networks. In *Proceedings of the 8th IEEE Symposium on Computers and Communications (ISCC'03)*, pages 125–130, Antalya, Turkey, July 2003.

[BBC+98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. RFC 2475 An Architecture for Differentiated Services, December 1998. URL reference: http://www.ietf.org/rfc/rfc2475.txt.

[BcM04] A. Belhoul, Y. A. Şekercioğlu, and N. Mani. A Survey of QoS Provisioning Mechanisms in Wireless All-IP Networks. In *Proceedings of the Third Workshop on the Internet, Telecommunications and Signal Processing (WITSP'04)*, Adelaide, Australia, December 2004.

[BcM06a] A. Belhoul, Y. A. Şekercioğlu, and N. Mani. A Mechanism for Enhancing VoIP Performance over Wireless Networks using Embedded Mobility-Specific Information in RSVP Objects. In *Proceedings*

of the IEEE Global Conference on Telecommunications (GLOBECOM'06), volume 3, pages 1–5, San Francisco, USA, November 2006.

[BcM06b] A. Belhoul, Y. A. Şekercioğlu, and N. Mani. Interaction of RSVP and Mobile IPv6 Protocols: An Evaluation of Performance on Wireless Networks. In *Proceedings of the IEEE International Conference on Communications (ICC'06)*, volume 3, pages 973–978, Istanbul, Turkey, June 2006.

[BCS94] R. Braden, D. Clark, and S. Shenker. RFC 1633 Integrated Services in the Internet Architecture: an Overview, June 1994. URL reference: http://www.ietf.org/rfc/rfc1633.txt.

[BGdMK06] G. Bolch, S. Greiner, H. de Meer, and Trivedi K. *Queuing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*. John Wiley & Sons, 2006.

[Blo] My Tech Blogs. NSIS: The New Signalling. URL reference: http://mytechblogs.com/network/2007/05/13/nsis-the-new-signalling/.

[BZB+97] B. Barden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. RFC 2205 Resource Reservation Protocol (RSVP) Version 1 Functional Specification, September 1997. URL reference: http://www.ietf.org/rfc/rfc2205.txt.

[Cas00] Claude Castelluccia. HMIPv6: A Hierarchical Mobile IPv6 Proposal. *ACM SIGMOBILE Mobile Computing and Communications Review*, 4(1):48–59, 2000.

[CDM] CDMA Development Group. URL reference: http://www.cdg.org.

[Cia05] Ciaran Buckley. Turf War Brewing over 3G Spectrum, August 2005. URL reference: http://www.electricnews.net/news.html?code=9629821.

[CR01] R. Cole and J. Rosenbluth. Voice over IP Performance Monitoring. *ACM SIGCOMM Computer Communication Review*, 2(1):9–24, 2001.

[Dat05] Datacomm Research Company. Hard Numbers and Experts' Insights on Migration to Evolved 3G and 4G Wireless Technology. URL reference: http://www.datacommresearch.com/competitiveedge/summary/ofdm.asp, October 2005. Research Report.

[DBV+03] R. Droms, J. Bound, B. Volz, T. Lemon, C. Perkins, and M. Carnet. RFC 3315 Dynamic Host Configuration Protocol for IPv6 (DHCPv6), July 2003. URL reference: http://www.ietf.org/rfc/rfc3315.txt.

[DoC07] NTT DoCoMo. Achievements in 4G System Development, 2007. URL reference: http://www.nttdocomo.com/pr/files/20070209_attachment02.pdf.

[Dur01] Durlacher Research. UMTS Report. URL reference: http://www.durlacher.com/wimax.html, May 2001.

[FAG95] D. Falconer, F. Adachi, and B. Gudmundson. Time Division Multiple Access Methods for Wireless Personal Communications. *IEEE Communications Magazine*, 33(1):50–57, 1995.

[FAML05] S. Fu, M. Atiquzzaman, L. Ma, and Y.-J. Lee. Signaling Cost and Performance of SIGMA: A Seamless Handover Scheme for Data Networks. *Wireless Communications and Mobile Computing*, 5(7):825–845, 2005.

[FCX+03] L. Ferreira, L. Correia, D. Xavier, L. Vasconcelos, and E. Fledderus. Final Report on Traffic Estimation and Services Characterization. Technical Report IST-TUL_WP1_DR_PUB_200_WL_05_D1.4, Information Society Technologies, May 2003. URL reference: http://momentum.zib.de/paper/momentum-d14.pdf.

[Glo] Global Mobile Suppliers Association. EDGE Operators Forum. URL reference: http://www.gsacom.com/eof/index.php4.

[GSMa] GSM Association. GPRS Platform. URL reference: http://www.gsmworld.com/technology/gprs/index.shtml.

[GSMb] GSM Association. GSM World - the Website of the GSM Association. URL reference: http://www.gsmworld.com.

[GSMc] GSM Association. HSPA Mobile Broadband Today. URL reference: http://hspa.gsmworld.com.

[HCS01] L. Hanzo, P. J. Cherriman, and J. Streit. *Wireless Video Communications: Second to Third Generation and Beyond*. IEEE Series on Digital & Mobile Communications. IEEE Press, 1^{st} edition, 2001.

[HKLdB05] R. Hancock, G. Karagiannis, J. Loughney, and S. Van den Bosc. RFC 4080 Next Steps in Signaling (NSIS), June 2005. URL reference: http://www.ietf.org/rfc/rfc4080.txt.

[HS02] R. Hsieh and A. Seneviratne. Performance analysis on Hierarchical Mobile IPv6 with Fast-handoff over End-to-End TCP. In *Proceedings of the IEEE GLOBECOM 2002*, volume 3, pages 2488–2492, Taipei, Taiwan, November 2002.

[Hsi] R. Hsieh. FHMIP ns-extension for Network Simulator. URL reference: http://mobqos.ee.unsw.edu.au/~robert/nstut.php.

[HY03] S. Y. Hui and K. H. Yeung. Challenges in the Migration to 4G Mobile Systems. *IEEE Communications Magazine*, 41(12):54–59, 2003.

[IEEa] IEEE 802 LAN/MAN Standards Committee. The IEEE 802.16 Working Group on Broadband Wireless Access Standards. URL reference: http://www.ieee802.org/16/.

[IEEb] IEEE 802 LAN/MAN Standards Committee. The IEEE 802.20 Working Group on Mobile Wireless Access (MBWA). URL reference: http://www.ieee802.org/20/.

[IEEc] IEEE 802.11 WLAN Standards Committee. The IEEE 802.11 Working Group Wireless LAN. URL reference: http://www.ieee802.org/11/.

[IR02] ITU-R. *Vision, Framework and Overall Objectives of the Future Development of IMT-2000 and Systems Beyond IMT-2000*. PDNR WP8F, 2002.

[IT96] ITU-T. *Recommendation P.800: Methods for Subjective Determination of Transmission Quality*. ITU-T, 1996.

[IT00] ITU-T. *Recommendation G.107: The E-model, a Computational Model for use in Transmission Planning*. ITU-T, 2000.

[IT03] ITU-T. *Recommendation P.59: Artificial Conversational Speech*. ITU-T, 2003.

[IT05] ITU-T. *Recommendation H.263: Video Coding for Low Bit Rate Communication*. ITU-T, 2005.

[ITU02] ITU-D. *Mobile cellular, subscribers per 100 people*. International Telecommunications Union Statistics, 2002. URL reference: http://www.itu.int/ITU-D/ict/statisitcs/at_glance/cellular02.pdf.

[JP00] D. Johnson and C. Perkins. Route Optimization in Mobile IP. Internet Draft, February 2000. URL reference: http://people.nokia.net/charliep/txt/optim.txt.

[JP04] D. Johnson and C. Perkins. RFC 3775 Mobility Support in IPv6, June 2004. URL reference: http://www.ietf.org/rfc/rfc3775.txt.

[JTKK05] R. Jain, M. B. Tariq, J. Kempf, and T. Kawahara. *The All-IP Next-Generation Network Architecture*. Wiley InterScience, 1^{st} edition, 2005.

[KCC05] S. Kurkowski, T. Camp, and M. Colagrosso. MANET Simulation Studies: The Current State and New Simulation Tools. Technical Report MCS-05-02, Information Society Technologies, February 2005. URL reference: http://toilers.mines.edu/Public/PublicationList/CSM-MCS-05-02.pdf.

[KJC+03] Y. Kim, B. J. Jeong, J. Chung, C. Hwang, J.S. Ryu, K. Kim, and Y. K. Kim. Beyong 3G: Vision, Requirements, and Enabling Technologies. *IEEE Communications Magazine*, 41(3):120–124, 2003.

[Kob78] H. Kobayashi. *Modeling and Analysis: An Introduction to System Performance Evaluation Methodology*. Addison Wesley, 1978.

[Koo05] R. Koodli. RFC 4068 Fast Handovers for Mobile IPv6, July 2005. URL reference: http://www.ietf.org/rfc/rfc4068.txt.

[Mar07] Maravedis Research Company. Spectrum Analysis - The Critical Factor in WiMAX versus 3G. URL reference: http://www.maravedis-bwa.com/wimax.html, February 2007. White Paper.

[MH99] A. Mahmoodian and G. Haring. A Resource Allocation Mechanism to Provide Guaranteed Service to Mobile Multimedia Applications. In

Proceedings of the First IEEE/Popov Workshop on Internet Technologies and Services, volume 9, pages 9–17, Moscow, Russia, October 1999.

[MTT+03] T. Murakami, T. Takazawa, M. Takahashi, M. Bandai, and I. Sasase. An Adaptive RSVP Connection Renewing Scheme with th IP Address of the Neighbor Routers in Mobile IP. In *Proceedings of the 2003 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM'03)*, pages 875–878, Victoria, Canada, August 2003.

[Mur] S. Murphy. Updated version of Marc Greis' RSVP/ns models. URL reference: http://www.eeng.dcu.ie/~murphys/ns-work/rsvp/.

[Nat] National Chiao Tung University. NCTU Codec. URL reference: http://megaera.ee.nctu.edu.tw/mpeg.

[Ope] Open Mobile Alliance. Messaging Working Group. URL reference: http://www.oponmobilealliance.org/tech/wg_committees/mwg.html.

[Par03] S. Parameswaran. WLRP: A Resource Reservation Protocol for Quality of Service in Next-Generation Wireless Networks. In *Proceedings of the 28th IEEE International Conference on Local Computer Networks (LCN'03)*, pages 295–296, Bonn, Germany, October 2003.

[PCTMH03] X. Perez-Costa, M. Torrent-Moreno, and H. Hartenstein. A Performance Comparison of Mobile IPv6, Hierarchical Mobile IPv6, Fast Handovers for Mobile IPv6 and their Combination. *ACM SIGMOBILE Mobile Computing and Communications Review*, 7(4):5–19, 2003.

[Per96] C. Perkins. RFC 2003 IP Encapsulation within IP, October 1996. URL reference: http://www.ietf.org/rfc/rfc2003.txt.

[Per02] C. Perkins. RFC 3344 Mobility Support for IPv4, August 2002. URL reference: http://www.ietf.org/rfc/rfc3344.txt.

[PKZM02] S. Paskalis, A. Kaloxylos, E. Zervas, and L. Merakos. Evaluating the RSVP Mobility Proxy Concept. In *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'02)*, volume 1, pages 270–274, Lisbon, Portugal, sep 2002.

[RSC+02] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. RFC 3261 SIP: Session Initiation Protocol, June 2002. URL reference: http://www.ietf.org/rfc/rfc3261.txt.

[S. 06] S. Lawson. Intel Readies Flexible WiMAX Chipset, July 2006. URL reference: http://www.networkworld.com/news/2006/072406-intel-readies-flexible-wimax.html.

[SCMB05] H. Soliman, C. Castelluccia, K. Malki, and L. Bellier. RFC 4140 Hierarchical MIPv6 Mobility Management, August 2005. URL reference: http://www.ietf.org/rfc/rfc4140.txt.

[SK03] S. Sroka and H. Karl. Performance Evaluation of a QoS-Aware Handover Mechanism. In *Proceedings of the 8th IEEE Symposium on Computer and Communications (ISCC'03)*, pages 117–124, Antalya, Turkey, July 2003.

[SMMA06] M. Shin, J. Ma, A. Mishra, and W. A. Arbaugh. Wireless Network Security and Interworking. *Proceedings of the IEEE*, 94(2):455–466, 2006.

[SSK00] X. Shangguan, W. K. G. Seah, and C. C. Ko. Evaluation of a Lightweight Resource Reservation Protocol for Mobile Internet Hosts. In *Proceedings of the Third IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'00)*, pages 119–127, California, USA, December 2000.

[Tac03] K. Tachikawa. A Perspective on the Evolution of Mobile Communications. *IEEE Communications Magazine*, 41(10):66–73, 2003.

[TB99] A. K. Talukdar and B. R. Badrinath. Integrated Services Packet Networks with Mobile Hosts: Architecture and Performance. *Wireless Networks*, 5(2):111–124, 1999.

[TB01] A. K. Talukdar and B. R. Badrinath. MRSVP: A Resource Reservation Protcol for an Integrated Services Networks with Mobile Hosts. *Wireless Networks*, 7(1):5–19, 2001.

[Thea] The 3rd Generation Partnership Project. URL reference: http://www.3gpp.org.

[Theb] The 3rd Generation Partnership Project 2. URL reference: http://www.3gpp2.org.

[Tho02] M. Thomas. Analysis of Mobile IP and RSVP Interactions. Internet Draft, October 2002. URL reference: http://tools.ietf.org/id/draft-thomas-nsis-rsvp-analysis-00.txt.

[TKWZ00] A. Terzis, J. Krawczyk, J. Wroclawski, and L. Zhang. RFC 2746 RSVP Operation over IP Tunnels, January 2000. URL reference: http://www.ietf.org/rfc/rfc2746.txt.

[TN98] S. Thomson and T. Narten. RFC 2462 IPv6 Stateless Address Autoconfiguration, December 1998. URL reference: http://www.ietf.org/rfc/rfc2462.txt.

[UCB] UCB/LBNL/VINT. Network Simulator ns-2. URL reference: http://www.isi.edu/nsnam/ns.

[UMT] UMTS World. WCDMA Specification and Information Page. URL reference: http://www.umtsworld.com/technology/wcdma.htm.

[Wid] J. Widmer. Extensions to the ns Network Simulator. URL reference: http://www.icsi.berkeley.edu/~widmer/mnav/ns-extension/.

[WJCJ] J. Winick, C. Jin, Q. Chen, and S. Jamin. Inet Topology Generator. URL reference: http://topology.eecs.umich.edu/inet.

[WP02] S. Wolf and M. Pinson. Video Quality Measurement Techniques. Technical Report TR-02-392, NTIA, June 2002. URL reference: http:www.its.bldrdoc.gov/pub/ntia-rpt/02-392.

[XA02] J. Xie and I. F. Akyildiz. An Optimal Location Management Scheme for Minimizing Signaling Cost in Mobile IP. In *Proceedings of the IEEE International Conference on Communications (ICC'02)*, volume 5, pages 3313–3317, New York, USA, April 2002.

[ZAB99] M. Zeng, A. Annamalai, and V. K. Bhargava. Recent Advances in Cellular Wireless Communications. *IEEE Communications Magazine*, 37(9):128–138, 1999.

Publications Related to the Study Presented in this Thesis

Journal Articles

[with Y. Ahmet Şekercioğlu and Nallasamy Mani] Interaction of RSVP and Mobile IPv6 Protocols: An Evaluation of Performance on Wireless Networks. *Wiley InterScience Journal of Wireless Communications and Mobile Computing*, (accepted for publication), September 2007.

[with Y. Ahmet Şekercioğlu and Nallasamy Mani] Mobility-Aware RSVP: A Framework for Improving the Performance of Multimedia Services over Wireless IP-based Mobile Networks. *Elsevier Journal of Computer Communications*, (submitted), August 2007.

Conference Papers

[with Y. Ahmet Şekercioğlu and Nallasamy Mani] Survey of QoS Provisioning Mechanisms in Wireless All-IP Networks. In *Proceedings of the Third Workshop on the Internet, Telecommunications and Signal Processing WITSP'04*, Adelaide, Australia, December 2004.

[with Y. Ahmet Şekercioğlu and Nallasamy Mani] Interaction of RSVP and Mobile IPv6 Protocols: An Evaluation of Performance on Wireless Networks. In *Proceedings of the IEEE International Conference on Communications ICC'06*, Istanbul, Turkey, June 2006. IEEE Communications Society.

[with Y. Ahmet Şekercioğlu and Nallasamy Mani] A Mechanism for Enhancing VoIP Performance over Wireless Networks using Embedded Mobility-Specific

Information in RSVP Objects. In *Proceedings of the IEEE Global Conference on Communications Globecom'06*, SanFrancisco, USA, November 2006. IEEE Communications Society.