

# **QoS Provisioning and Mobility Management for IP-Based Wireless LAN**

**Christos Politis**

Submitted for the Degree of  
Doctor of Philosophy  
from the  
University of Surrey



Centre for Communication Systems Research (CCSR)  
School of Electronics and Physical Sciences  
University of Surrey  
Guildford, Surrey GU2 7XH, UK

March 2004

© Christos Politis 2004

# Summary

Today two major technological forces drive the telecommunication era: the wireless cellular systems and the Internet. As these forces converge, the demand for new services, increasing bandwidth and ubiquitous connectivity continuously grows. The next-generation mobile systems will be based solely or in a large extent, on the Internet Protocol (IP).

This thesis begins by addressing the problems and challenges faced in a multimedia, IP-based Wireless LAN environment. The ETSI HiperLAN/2 system has been mainly selected as the test wireless network for our theoretical and simulation experiments. Apart from the simulations, measurements have been taken from real life test scenarios, where the IEEE 802.11 system was used (UniS Test-bed). Furthermore, a brief overview of the All-IP network infrastructure is presented. An extension to the conventional wireless (cellular) architecture, which takes advantage of the IP network characteristics, is considered. Some of the trends driving the 3G and WLANs developments are explored, while the provision of quality of service on the latter for real-time and non-real-time multimedia services is investigated, simulated and evaluated. Finally, an efficient and catholic QoS framework is proposed.

At the same time, the multimedia services should be offered in a seamless and uninterrupted manner to users who access the all-IP infrastructure via a WLAN, meeting the demands of both enterprise and public environments anywhere and anytime. Thus providing support for mobile communications not only in terms of terminal mobility, as is currently the case, but also for session, service and personal mobility. Furthermore, this mobility should be available over heterogeneous networks, such as WLANs, UMTS, as well as fixed networks. Therefore, this work investigates issues such as, multilayer and multi-protocol (SIP-Mobile IP-Cellular IP) mobility management in wireless LAN and 3G domains. Several local and global mobility protocols and architectures have been tested and evaluated and a complete mobility management framework is proposed.

Moreover, integration of simple yet efficient authentication, accounting and authorisation mechanisms with the multimedia service architecture is an important issue of IP-based WLANs. Without such integration providers will not have the necessary means to control their provided services and make revenue from the users. The proposed AAA architecture should support a robust AAA infrastructure providing secure, fast and seamless access granting to multimedia services. On the other hand, a user wishing a service from the All-IP WLAN infrastructure needs to be authenticated twice, once to get access to the network and the other one should be granted for the required service. Hence, we provide insights into these issues by simulating and evaluating

pre-authentication techniques and other network authentication scenarios based on the well-known IEEE 802.1x protocol for multimedia IP-based WLANs.

**Key words:** WLAN, HiperLAN/2, IEEE 802.11, All-IP Networks, Quality of Service, Mobility Management, AAA, Multimedia Services, Simulations, Test-bed.

**Email:** [c.politis@surrey.ac.uk](mailto:c.politis@surrey.ac.uk)

**WWW:** <http://www.ee.surrey.ac.uk/Personal/C.Politis/>

*Dedicated to Joanna and my family*

# Acknowledgments

I would like to thank my supervisor, Prof Rahim Tafazolli for his guidance, help and support, during this project. It was an endurance and difficult to win race and his advice, technical and managerial were more than valuable.

I would have not finished this work without the emotional support from my parents and I thank them very much. Seamlessly, I would like to thank my partner, Joanna, for being so kind and considerate throughout my studies. Her support has been invaluable.

It is only fair to mention my friends and colleagues, Dr Tasos Dagiuklas, Dr Dorgham Sisalem, Mr Yacine Rebani, Mr Dionisios Gatzounas, Dr Dimitrios Mavrakis, Mr Aris Diamantopoulos and all my colleagues here and in Europe that I have worked with the last five years. It has been fun working with you lot!

I acknowledge the well-established collaboration and mutual support I had with my colleagues in CCSR and in particular Mr Nadeem Akhtar, Mr Kar Ann Chew and Mr Michael Georgiades. I will never forget the times (usually late at night) Mr Jonathan Rodriguez and I spent trying to understand complicated technical topics, meeting deadlines and yet philosophise on the well-hidden secrets of life (in CCSR!); his company and help are deeply appreciated.

Finally, many thanks to the secretaries of CCSR, Stephanie, Emmanuel and Anne for their help with the administrative work, which I must admit, was overwhelming.

**“ΔΙΑ ΜΟΧΘΟΥ ΤΑ ΑΡΙΣΤΑ”**

124 Π.Β.Ε., Τρίπολη

# Table of Contents

SUMMARY.....	II
ACKNOWLEDGMENTS.....	V
TABLE OF CONTENTS .....	VI
LIST OF FIGURES .....	X
LIST OF TABLES .....	XIII
GLOSSARY OF ACRONYMS .....	XIV
INTRODUCTION.....	1
1.1    BACKGROUND .....	1
1.2    MOTIVATION .....	3
1.3    RESEARCH OBJECTIVES .....	4
1.4    STRUCTURE OF THE THESIS .....	5
<b>2    BACKGROUND – ADDRESSED ISSUES .....</b>	<b>7</b>
2.1    INTRODUCTION.....	7
2.2    STATE OF THE ART.....	8
2.2.1 <i>Review of the Cellular Networks</i> .....	8
2.2.1.1    Standards Overview.....	8
2.2.1.2    GSM Network.....	9
2.2.1.3    Extend GSM with GPRS.....	10
2.2.1.4    Phased evolution to UMTS .....	11
2.2.1.5    The all-IP UMTS .....	14
2.2.2 <i>Review of the Wireless Local Area Network Standards</i> .....	15
2.2.2.1    IEEE 802.11b.....	15
2.2.2.2    ETSI HiperLAN/2 .....	18
2.2.2.3    Regulation Status Overview.....	20
2.3    WHY HIPERLAN/2? .....	21
2.4    ALL-IP WLAN.....	21
2.4.1    Overview .....	21
2.4.2    Mobility Management.....	23
2.4.3    End-to-end (ete) QoS.....	26

2.4.4	Security .....	28
2.4.5	Related projects .....	29
2.5	SUMMARY .....	30
<b>3</b>	<b>ALL-IP WLAN SYSTEM DESIGN .....</b>	<b>31</b>
3.1	INTRODUCTION.....	31
3.2	SYSTEM ARCHITECTURE.....	31
3.2.1	Protocol Reference Model.....	31
3.2.1.1	HiperLAN/2 PHY Layer .....	32
3.2.1.2	HiperLAN/2 MAC Protocol.....	35
3.2.1.3	Convergence Layer for IP support.....	37
3.2.2	H2 Resource Management Analysis.....	38
3.3	OFFERED SERVICES AND TRAFFIC REQUIREMENTS.....	40
3.3.1	Services .....	41
3.3.1.1	Internet Access.....	41
3.3.1.2	Video-telephony or videoconferencing.....	42
3.3.1.3	VoIP telephony .....	42
3.3.1.4	Streaming video and broadcast .....	42
3.3.1.5	Other applications.....	43
3.4	QUALITY OF SERVICE PARAMETERS.....	43
3.5	BROADBAND WLANS QOS CLASSES.....	43
3.5.1	Conversational class.....	44
3.5.2	Streaming class.....	44
3.5.3	Interactive class.....	45
3.5.4	Background class.....	45
3.5.5	QoS requirements .....	46
3.5.6	Summary of service requirements.....	47
3.6	TRAFFIC MODELS.....	47
3.6.1	Internet (UMTS 3G model) .....	47
3.6.2	Internet (Choi's model) .....	49
3.6.3	Voice [STR-2.1].....	51
3.6.4	VoIP Model.....	52
3.6.4.1	Audio Codecs.....	52
3.6.4.2	Codec (G.723).....	52
3.6.4.3	Performance Issues .....	53
3.6.4.4	Voice Traffic Characteristics .....	54
3.6.5	A Study on Video Streaming (and VoD).....	56

3.6.6	<i>Streaming audio [MENA-00]</i> .....	59
3.7	SUMMARY .....	59
<b>4</b>	<b>QOS FRAMEWORK FOR ALL-IP WLAN</b> .....	<b>60</b>
4.1	INTRODUCTION.....	60
4.2	ENHANCEMENTS TO HIPERLAN/2 PHY FOR QOS SUPPORT .....	60
4.2.1	<i>Link Adaptation Technique</i> .....	61
4.2.2	<i>H2 LA Performance Evaluation</i> .....	62
4.3	ENHANCEMENTS TO HIPERLAN/2 MAC PROTOCOL FOR QOS SUPPORT – CROSS LAYER COOPERATION .....	65
4.3.1	<i>IP Interface Definition</i> .....	65
4.3.2	<i>QoS Support and Scheduling</i> .....	67
4.3.3	<i>Simulation models and Performance analysis</i> .....	71
4.3.3.1	Network Related Aspects .....	71
4.3.3.2	Traffic Generation Related Aspects .....	72
4.3.3.3	Performance Related Aspects [POL-02].....	73
4.3.3.4	Performance Evaluation Analysis .....	74
4.4	CONNECTION ADMISSION CONTROL FOR IP-BASED HIPERLAN/2.....	78
4.4.1	<i>Admission Control Algorithm</i> .....	78
4.4.2	<i>Numerical Results and Analysis</i> .....	81
4.5	DISCUSSION AND SUMMARY .....	83
<b>5</b>	<b>MOBILITY MANAGEMENT FRAMEWORK FOR IP-BASED WLAN</b> .....	<b>84</b>
5.1	INTRODUCTION.....	84
5.2	MOBILITY MANAGEMENT IN ALL-IP NETWORKS .....	85
5.2.1	<i>Basics on mobility management</i> .....	85
5.2.2	<i>Review of the existing IP micro-mobility approaches</i> .....	86
5.2.3	<i>State-of-the-Art of micro-mobility schemes</i> .....	88
5.2.3.1	Micro-mobility Categories .....	89
5.2.3.2	Hierarchical tunnelling .....	89
5.2.3.3	Host-specific forwarding .....	90
5.2.3.4	The comparison .....	91
5.3	MULTILAYER MOBILITY MANAGEMENT .....	94
5.3.1	<i>Differentiation between global and local mobility</i> .....	94
5.3.2	<i>Mobility management in All-IP networks: Protocols and architectures</i> .....	94
5.3.3	<i>SIP-based approach for mobility management</i> .....	96
5.4	SIMULATION RESULTS AND DISCUSSION FOR MICRO-MOBILITY [CH-PO'02] .....	98

5.4.1	<i>Handoff Performance</i> .....	100
5.4.2	<i>Overhead</i> .....	102
5.4.3	<i>Proposed Micro-mobility Protocol for IP-based WLAN Systems</i> .....	103
5.5	<b>SIP/CELLULAR IP-BASED MOBILITY SOLUTION</b> .....	103
5.5.1	<i>SIP/DHCP Test-bed Environment</i> .....	105
5.5.2	<i>SIP/Cellular IP Test-bed Environment</i> .....	105
5.5.3	<i>Handover Procedure using SIP/DHCP Scheme</i> .....	106
5.5.4	<i>Handover Procedure using SIP/Cellular IP Scheme</i> .....	107
5.5.5	<i>Performance Evaluation and Discussion</i> .....	108
5.5.5.1	Performance evaluation results .....	109
5.5.5.2	Discussion .....	110
5.6	<b>MOBILITY MANAGEMENT FRAMEWORK FOR IP-BASED WLAN SYSTEMS</b> .....	110
5.6.1	<i>Comparison of SIP-based and MIP-based mobility approaches [PO-CH'03]....</i>	110
5.6.2	<i>RTP/UDP Real Time Applications</i> .....	115
5.6.3	<i>Non-Real Time Applications (TCP / FTP)</i> .....	119
5.7	<b>DISCUSSION AND SUMMARY</b> .....	124
<b>6</b>	<b>AAA SCHEMES FOR IP-BASED WLAN</b> .....	<b>125</b>
6.1	<b>INTRODUCTION</b> .....	125
6.2	<b>AAA ARCHITECTURE IN ALL-IP NETWORKS</b> .....	125
6.2.1	<i>AAA in WLAN</i> .....	126
6.2.1.1	WEP .....	127
6.2.1.2	802.1x protocol .....	127
6.2.2	<i>AAA in the Internet</i> .....	128
6.2.2.1	General AAA architecture.....	128
6.2.2.2	AAA protocols.....	133
6.3	<b>AAA SIMULATIONS FOR WLAN SYSTEMS</b> .....	135
6.3.1	<i>Network Access Simulation</i> .....	135
6.3.1.1	IEEE 802.1x Model for HiperLAN/2 .....	136
6.3.2	<i>Service Access Simulation</i> .....	141
6.4	<b>DISCUSSION AND SUMMARY</b> .....	142
<b>7</b>	<b>CONCLUSIONS AND FUTURE WORK</b> .....	<b>143</b>
7.1	<b>MAJOR ACHIEVEMENTS</b> .....	143
7.2	<b>FUTURE WORK</b> .....	145
7.3	<b>PUBLICATIONS</b> .....	147
	<b>REFERENCES</b> .....	<b>150</b>

# List of Figures

Figure 2-1: The IMT2000 Evolution of Standards [EVO-D2.1].	9
Figure 2-2: GSM Architecture Overview & Interfaces [EVO-D2.1].	10
Figure 2-3: GPRS Functional Architecture and Interfaces [EVO-D2.1].	11
Figure 2-4: Phases of Standards Development.	11
Figure 2-5: Phases of 2.5/3G Network Evolution.	12
Figure 2-6: Overall Architecture of UMTS Release 2000, All-IP [EVO-D2.1].	14
Figure 2-7: Protocol Reference Model for Wireless LANs and the Internet.	16
Figure 2-8: The past, the present and future of ETSI HIPER-family.	19
Figure 2-9: HiperLAN/2 MAC Frame.	20
Figure 2-10: WLAN network stack.	22
Figure 2-11: a) Hierarchical tunnelling and b) Explicit Terminal Routing Approaches.	25
Figure 3-1: HiperLAN/2 protocol stack (AP and MT).	32
Figure 3-2: Block diagram of a HiperLAN/2 transceiver.	33
Figure 3-3: Broadcast burst preamble.	34
Figure 3-4: MAC frame for HiperLAN/2.	36
Figure 3-5: Mapping of packets to the H2 layers.	37
Figure 3-6: Traffic model for packet data services.	48
Figure 3-7: Structure of a WWW page [STR-2.1].	49
Figure 3-8: Sequence and parameters for a WWW session.	50
Figure 3-9: Choi's Behavioural Model.	51
Figure 3-10: Traffic model for speech service.	51
Figure 3-11: Output of Speech Coder G.723 - High Rate (6.3 kb/s).	53
Figure 3-12: Single Speaker Model.	55
Figure 3-13: Encoded (15, 3) Group of Pictures, hatched frames belong to other Groups of Pictures.	57
Figure 4-1: PER vs. C/I for all PHY modes (channel E).	63
Figure 4-2: PER vs. C/I for $\frac{1}{2}$ and $\frac{3}{4}$ LA.	63
Figure 4-3: Throughput vs. C/I for all PHY modes (channel E).	64
Figure 4-4: Throughput vs. C/I for $\frac{1}{2}$ and $\frac{3}{4}$ LA (channel E).	64
Figure 4-5: DL IP Layer structure.	65
Figure 4-6: DL PHY structure.	66
Figure 4-7: UL IP Layer Structure.	66
Figure 4-8: UL Radio Layer.	67

Figure 4-9: QoS Provisioning for all-IP HiperLAN/2. ....	69
Figure 4-10: IP-based H2 control plane (layers interworking/cooperation). ....	70
Figure 4-11: IP-based H2 data plane (layers interworking/cooperation). ....	70
Figure 4-12: Simulation scenario(s). ....	72
Figure 4-13: Throughput vs. load for http traffic (DiffServ). ....	75
Figure 4-14: Throughput vs. load for http traffic (IntServ). ....	76
Figure 4-15: Ete-delay vs. load for VoIP (DiffServ). ....	76
Figure 4-16: Ete-delay vs. load for VoIP (IntServ). ....	77
Figure 4-17: Ete-delay vs. load for video streaming (DiffServ). ....	77
Figure 4-18: Ete-delay vs. load for video streaming (IntServ). ....	78
Figure 4-19: HiperLAN/2 Protocol Functional Entities. ....	79
Figure 4-20: H2 MT requests resources [POL-TAF]. ....	81
Figure 4-21: Comparison of blocking probability as load varies (CC: WFQ). ....	82
Figure 4-22: Comparison of block probability as load varies (CC: CBQ). ....	82
Figure 5-1: Hierarchical decomposition of mobility management. ....	87
Figure 5-2: Packet delivery under Hierarchical Tunnelling schemes. ....	90
Figure 5-3: Mobile terminal executes registration with FAs. ....	90
Figure 5-4: Classification of IP mobility protocols. ....	93
Figure 5-5: Session establishment with the extended-SIP approach. ....	96
Figure 5-6: SIP-based approach for mobility management. ....	97
Figure 5-7: The simulated network topology. ....	99
Figure 5-8: Packet lost ratio due to handoff (CBR traffic source). ....	101
Figure 5-9: Packet lost ratio due to handoff (on/off traffic source). ....	101
Figure 5-10: Size of host-specific entry created in gateway. ....	102
Figure 5-11: Size of host-specific entry created in access routers. ....	103
Figure 5-12: Subnet handover using SIP/DHCP. ....	105
Figure 5-13: Subnet handover using SIP/CIP. ....	106
Figure 5-14: SIP/DHCP handover signalling. ....	107
Figure 5-15: SIP/Cellular IP handover signalling. ....	108
Figure 5-16: Network architecture for SIP and MIP approaches. ....	111
Figure 5-17: SIP-based approach for mobility management. ....	112
Figure 5-18: MIP-based approach for mobility management. ....	112
Figure 5-19: Message Sequence for the SIP-based approach. ....	113
Figure 5-20: Message Sequence for the MIP-based approach. ....	115
Figure 5-21: PLR for RTP session running over SIP and MIP (MT is static). ....	116
Figure 5-22: RTP stop time for SIP and MIP during handoff. ....	116
Figure 5-23: PLR for SIP (fho) and MIP. ....	117

Figure 5-24: PLR for SIP (hho) and MIP.....	118
Figure 5-25: PLR for SIP and MIP, for $twl = 20$ ms (without old MIP FA forwarding packets to new FA).....	118
Figure 5-26: PLR for SIP and MIP, for $twl = 50$ ms (without old MIP FA forwarding to new FA). .....	119
Figure 5-27: The impact of MIP and SIP mobility management schemes on FTP download response time.....	120
Figure 5-28: The impact of MIP and SIP mobility management schemes on the received TCP segment number.....	120
Figure 5-29: The impact of MIP and SIP mobility management schemes on TCP Congestion window.....	121
Figure 5-30: FTP download time for different TCP settings.....	123
Figure 5-31: File download time for MIP and SIP (tracking agent –enabled).....	123
Figure 6-1: Authentication Process in 802.1x.....	128
Figure 6-2: General AAA architecture [REB-02].....	129
Figure 6-3: Accounting process [REB-02].....	133
Figure 6-4: The basic IEEE 802.1x architecture.....	137
Figure 6-5: Pre-authentication scheme for fast handoff in H2.....	138
Figure 6-6: Message exchanges before handoff.....	139
Figure 6-7: Message exchanges after handoff.....	139
Figure 6-8: Handoff delay for pre-authenticated H2 APs (Local AAA server).....	140
Figure 6-9: Handoff delay for pre-authenticated H2 APs (Remote AAA server).....	141
Figure 7-1: Today and Tomorrow in the Wireless World [DAG-VEL].....	146

# List of Tables

Table 3-1: OFDM modulation main parameters. ....	33
Table 3-2: (a) HiperLAN/2 transmission rates (after coding) [ETSI-475], (b) Characteristics of H/2 transport channels [ETSI-761.1]. ....	38
Table 3-3: Interactive Games.....	46
Table 3-4: Videoconference.....	46
Table 3-5: Broadband Internet Access. ....	46
Table 3-6: High quality streaming .....	46
Table 3-7: HiperLAN/2 services QoS requirements. ....	47
Table 3-8: End-to-end services QoS requirements. ....	47
Table 3-9: Parameters for traffic models.....	49
Table 3-10: Choi WWW traffic model parameters [CHO-99]. ....	50
Table 3-11: Parameters for Voice modelling with Exponential Distribution.....	55
Table 3-12: Usual GOP patterns.....	58
Table 3-13: Statistics (MPEG 640 x 480 pixel). ....	58
Table 4-1: Spectral efficiencies and corresponding H2 PHY modes. ....	61
Table 4-2: Delay requirements for voice and video traffic.....	73
Table 5-1: Comparison of micro-mobility schemes [EVO-D2.2]. ....	92
Table 5-2: Comparison of IP-based mobility protocols.....	93
Table 5-3: Handover delay for SIP/DHCP and SIP/CIP [AKH-POL]. ....	109
Table 5-4: SIP delay in SIP/CIP [AKH-POL]. ....	110
Table 5-5: FTP download time for different TCP settings. ....	122
Table 6-1: Comparison between IEEE 802.11a and HiperLAN/2 [GRE-02]. ....	135

# Glossary of Acronyms

(3)DES	(3)Data Encryption Standard
2G	Second Generation Mobile/Cellular Systems
3G	Third Generation Mobile/Cellular Systems
3GPP	Third Generation Partnership Project
AAA	Authentication Authorisation Accounting
ACF	Association Control Function
ACH	Access Feedback Channel
ADSL	Asymmetric Digital Subscriber Line
AES	Advanced Encryption Standard
AKA	Authentication and Key Agreement
AKI	Authentication Key Identifier
AODV	Ad hoc On-Demand Distance Vector Routing
AP	Access Point
API	Application Programming Interface
AR	Access Router
ARP	Address Resolution Protocol
ARQ	Automatic Resource Request
ASM	Application Specific Module(s)
AVP	Attribute Value Pair(s)
B3G	Beyond 3G
BCCH	Broadcast Control CHannel
BCH	Broadcast CHannel
BER	Bit Error Ratio
BRAN	Broadband Radio Access Networks
BS	Base Station

BU	Binding Update
BW	Bandwidth
CAC	Connection or Call Admission Control
CBQ	Class-based Queuing
CC	Congestion Control
CH	Correspondent Host
CIP	Cellular IP
CL	Convergence Layer
CoA	Care of Address
COPS	Common Open Policy Service
CoS	Class of Service
CP	Cyclic Prefix
CPCS	Common Part Convergence Sublayer(s)
CRC	Cyclic Redundancy Check
CS	Circuit Switched
DAA	Dynamic Address Allocation
DAD	Duplicated Address Detection
DCC	DLC Connection Control
DFA	Dynamic Frequency Allocation
DFS	Dynamic Frequency Selection
DHCP	Dynamic Host Configuration Protocol
DiffServ/DS	Differentiated Services
DiL	Direct Link
DL	Downlink
DLC	Data Link Control layer
DRCP	Dynamic Registration and Configuration Protocol
DREMIP	Distributed Registration Extensions to MIP

DSCP	Differentiated Services Code Points
DSR	Dynamic Source Routing
DSSS	Direct Sequence Spread Spectrum
EAP	Extensible Authentication Protocol
EAPOL	EAP over LAN
EC	Error Control
EMA	Edge Mobility Architecture
ete	End-to-End
ETSI	European Telecommunications Standards Institute
EU	European Union
FA	Foreign Agent
FCCH	Frame Control CHannel
FCH	Frame CHannel
FER	Frame Error Ratio
FFT	Fast Fourier Transform
fho	Full Handoff
FHR	Frequent Handoff Region
FHSS	Frequency Hopping Spread Spectrum
FTP	File Transfer Protocol
FWA	Fixed Wireless Access
GGSN	Gateway GPRS Support Node
GoS	Grade of Service
GPRS	General Packet Radio System
GSM	Global System for Mobile Communications
GSN	GPRS Support Node
GTP	GPRS Tunnelling Protocol
GW	Gateway

H2	HiperLAN/2 or High Performance LAN (type 2)
HA	Home Agent
HAWAII	Handoff Aware Wireless Access Internet Infrastructure
hho	Half handoff
HL	Higher Layer(s)
HLR	Home Location Register
HMIP	Hierarchical Mobile IP
HO	Handover or Handoff
HR	Home Registrar
HTML	HyperText Markup Language
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
IMT-2000	International Mobile Telecommunications-2000
IN	Intelligent Network
IntServ	Integrated Services
IP	Internet Protocol, version 4 (v4), version 6 (v6)
IPSec	IP Security (protocol)
ISDN	Integrated Services Digital Network
ISP	Internet Service Provider
IST	Information Society Technologies
ITU	International Telecommunications Union
L2 / 3	Layer 2 /3
LA	Link Adaptation
LAN	Local Area Network
LCH	Long Channel
MAC	Medium Access Control layer
MANET	Mobile Ad Hoc Networks

MAP	Mobility Anchor Point
MER-TORA	Temporary Ordered Routing Algorithm with Edge Mobility Support
MGCP/ MEGACO	Media Gateway Control Protocol/Media Gateway Controller
MIP	Mobile IP
MMP	Multicast for Mobility Protocol
MPEG	Motion Picture Experts Group
MSC	Mobile Switching Centre
MSS	Maximum Segment Size/Sequence
MT/ MN/ MH	Mobile Terminal/ Mobile Node/ Mobile Host
MTMR	Multiple Transmit Multiple Receive
MTU	Maximum Transmission/Transfer Unit
NAI	Network Access Identifier
NAS	Network Access Server
NAT	Network Address Translation
nLOS	<i>non-Line Of Sight</i>
OFDM	Orthogonal Frequency Division Multiplexing
OSA	Open Service Access or Architecture
PAN	Personal Area Network
PBCC	Packet Binary Convolutional Code
PCM	Pulse-Code Modulation
PDA	Personal Digital Assistant
PDP	Policy Decision Point
PDU	Packet Data Unit
PEP	Policy Enforcement Point
PER	PDU Error Ratio
PFE	Protocol Functional Entities
PHY	Physical layer

PLR	Packet Loss Ratio
PS	Packet Switched
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RADIUS	Remote Authentication Dial In User Service
RAFA	Regional Aware Foreign Agent
RAN	Radio Access Network
RCH	Random Access Channel
RF	Radio Frequency(ies)
RG	Resource Grant
RG-IE	Resource Grant Information Elements
RLC	Radio Link Control
RMS	Root Mean Square
RNC	Radio Network Controller
RR	Resource Request
RRC	Radio Resource Control
RSA	Rivest-Shamir-Adleman
RSVP	Resource Reservation Protocol
RTP	Real Time Protocol
RTT	Round Trip Delay
RU	Route Update
SAA	Static Address Allocation
SAP	Service Access Point
SAR	Segmentation and Reassembly sublayer
SCH	Short Channel
SCTP	Stream Control Transmission Protocol
SDP	Session Description Protocol

SDU	Service Data Unit(s)
SGSN	Serving GPRS Support Node
SIP	Session Initiation Protocol
SNR	Signal-to-Noise Ratio
SOHO	Small Office Home Office
SSCS	Service Specific Convergence Sub-layer(s)
STUN	Simple Traversal of UDP through NATs
TACACS	Terminal Access Controller Access Control System
TCP	Transmission Control Protocol
TDMA/TDD	Time Division Multiple Access / Time Division Duplex
TEP	Tunnel Establishment Protocol
THEMA	Transparent Hierarchical Mobility Agents
TKIP	Temporal Key Integrity Protocol
TLS	Transport Layer Security protocol
ToS	Type of Service
UA	User Agent
UDP	User Datagram Protocol
UL	Uplink
UMTS	Universal Mobile Telecommunications System
URL	Uniform Resource Locator
UTRAN	Universal Terrestrial Radio Access Network
VBR	Variable Bit Rate
VLR	Visiting Location Register
VoIP	Voice over IP
VPN	Virtual Private Network
WAN	Wide Access Network
WEP	Wired Equivalent Privacy

WFQ	Weighted Fair Queuing
WG	Working Group
WLAN	Wireless Local Area Network
WWW	World Wide Web

# Chapter 1

## Introduction

### 1.1 Background

Since the beginning of the 21st century, the telecommunication services have been undergoing major changes. Information and communication technologies are having an ever-increasing impact on the way we do things, underpinning global business, life style, markets and becoming essential to almost every business sector. The technologies will continue to advance rapidly and the information sector will see a massive increase in the amount of multimedia content accessible remotely through communication networks.

It is envisaged that mobile communication systems beyond the 3rd generation (SB3G) will offer a multitude of services over a variety of access technologies. To fulfil the vision and accomplish the grand objectives for future networks that support the heterogeneity in network accesses, communication services, mobility, security, user devices, etc., it is of great importance to understand significant and high priority requirements for the target future network in terms of services and network capabilities as well as to promote necessary researches on the networking technology by providing a guiding framework of the research areas and technical issues with priority. The new architecture and new technologies will have to be derived from a set of fundamental assumptions and requirements that govern design.

The concept of convergence - using a single Internet-based network to deliver voice and data traffic - is nothing new. For years engineers have tried to put as much traffic, data or voice, as they could down a single pipe. But in recent years, the technology has finally caught up with the concept, and long-term promises can finally be fulfilled. There are many configurations for this idea such as PSTN to IP, GSM to UMTS, ISDN to IN etc. Convergence networks, whether local ("LAN") or Wide-Area ("WAN") have the ability to carry multiple data types through a single channel. Typically voice, video and data are the most common of these. Convergence brings the best of all worlds to the end-user on a single channel, regardless of the transport media (e.g. copper, fibre, wireless, laser, microwave, etc.). One of the major activities undertaken by the 3GPP project is the standardization of 3G mobile systems. Within 3GPP, the first 3G architecture (Release 99) is a logical evolution of 2G, whereas the latest Release (ex-Release 00, now it is

called Release 5)<sup>1</sup> is a complete revolution introducing new concepts and features. The main characteristics of Release 5 include a shift towards the all-IP network architecture, introduction of multimedia capabilities in the core network (through the use of a multimedia server) and use of IP transport in UTRAN. Furthermore, the option 2 of Release 5 reference architecture additionally supports all legacy circuit-switched-only mobile terminals via the MSC server. This All-IP network architecture [3GPP'03] will offer seamless multimedia services to mobile users who access the network via a variety of heterogeneous wireless technologies. For their importance in the near to medium term evolution of global mobile telecommunications, two types of wireless technologies are of prime importance, the UMTS and the WLANs.

In a perfect world all networks would support the same access technology allowing users to utilise the same end device to access all networks. While this surely would simplify the design and complexity of end systems, experience has shown that this is not pragmatic. Already in GSM networks, we can identify at least three frequency bands with different radio frequencies. UMTS networks show already a higher heterogeneity. This heterogeneity is further intensified by considering access through WLAN technologies which use yet another row of radio frequencies. In a highly heterogeneous environment with different network technologies, varying access protocols and different security mechanisms an adhesive common platform should be defined to provide global connectivity, mobility among networks, and service provisioning across the networks: the IP. Furthermore, in order to support seamless communications in such a heterogeneous environment, i.e., to allow users to access a network regardless of the used network technology and to move between these networks, an architecture should be designed and specified; integrating these heterogeneous technologies without requiring a unifying middleware layer other than the Internet Protocol. This is achieved by relying on standardised protocols, as well as, on gateways for translating between heterogeneous technologies.

It is envisaged that WLANs could form a complementary technology to the next generation mobile systems in order to provide high bandwidth connectivity in the range of 10 Mb/s and above [ETSI-957]. WLANs usually operate in the home or business environment (hot spots) with comparably small area coverage, and opposing to UMTS are usually unmanaged or operated on a corporate basis. HiperLAN/2 [H/2-00] and IEEE 802.11 [802.11-94] are the candidate and most popular wireless indoor and short-range outdoor technologies for the wireless local area network architecture.

---

<sup>1</sup> At the time of writing this thesis, another release was under way, namely Release 6.

## 1.2 Motivation

Complicated problems arise when real-time traffic needs to be multiplexed over a wireless packet switched network, as in the IP-based Wireless LAN case. The limited availability of radio resources in a wireless network creates the main difficulty for the implementation and support of the real-time services. The IP layer provides means for transmission/relay that is independent of the lower layer(s) bit-rate, but bandwidth can be provided on demand only when there is in abundance. In WLANs (H2, 802.11a) high bit-rate transmission can be achieved, but the available frequencies are unlicensed (or license-exempt) and resources should be managed in an efficient and flexible manner. However, it is clear, that the use of the IP layer as a medium and base, will permit the existence of the different services in the radio environment, distinguish among them and offer them the expected Quality of Service (QoS). However, the interfacing between IP and MAC/DLC (as well as between PHY and DLC/MAC) is very important and is thoroughly investigated and addressed in this study. Furthermore, the implementation of QoS requires specialised interactions between the wireless and fixed resource allocation entities (IP routers/access points), which are far more complicated than the similar interactions within the fixed network (i.e. connections among the fixed stations), or when no QoS requirements need to be considered. Moreover, the QoS guarantee in the fixed network does not require support and interactions with the layer(s) below the IP; but this is not the case for WLANs. The underlying to the IP layer, the DLC/MAC layer (and through them the PHY), must be able to recognise and support different services, in terms of bandwidth reservation and resource management.

There is another reason that makes the development and support of multimedia based IP services much more challenging than the development and support of non-real time services, over the wireless medium; that is the mobility of the terminal-service-user. Mobility management in heterogeneous wireless environments (i.e. 3G, WLANs) and beyond has to support all these different mobility types, to allow the end user access the network and subscriber services provided in a seamless and transparent manner. To support these types of mobility, one should address this in different *mobility domains*. These domains should work in a complementary manner by utilising a set of mobility protocols (mobility framework) to guarantee transparent mobility to end systems. The motivation for this stems from the requirement of obtaining a new IP address every time the mobile host changes its point of attachment to a network. This may happen quite often in the local mobility scenario (WLAN case) and hence the handoff performance will be affected by the long address acquisition delays. The idea is to have multi-layer and hierarchical mobility management architecture.

Another important issue is the possible enhancements to the network access in order to provide fast and secure access to mobile multimedia services. At the same time, this should give an insight

into a robust Authentication Authorisation and Accounting (AAA) infrastructure granting fast access to the services. On the other hand, a user wishing a service from a WLAN infrastructure needs to be authenticated twice, once to get access to the network and the other one to be granted for the required service. Concurrently, pre-authentication techniques should be investigated in order to provide seamless, fast and secure mobility (support vertical and horizontal handoffs) for the All-IP WLAN systems.

### 1.3 Research Objectives

Considering the issues discussed in the previous paragraphs, three important research areas have been identified within the All-IP WLAN framework:

- **QoS provisioning for IP-based WLAN systems:** To act as a test model for the simulations and evaluations the ETSI HiperLAN/2 was selected (“onerous” justifications are presented later on). The suitability of the HiperLAN/2 MAC protocol to support IP traffic, along with the proposed protocol modifications to PHY layer for ‘optimum’ IP support were investigated. After selecting the architecture for an IP-based WLAN and studying the proposed WLAN protocol stacks and their interactions, the necessity to provide a Quality of Service framework was presented and justified. The interfacing and interworking between PHY, MAC and network (IP) layers were discussed, while the IP role, as an adhesive was presented (convergence layers theory). Summing up, the following technical problems were addressed in this research topic:
  - ✓ A link adaptation algorithms was proposed (Chow technique), in order to optimise the performance of PHY and MAC layers of H2.
  - ✓ Two existing traffic controllers (WFQ and CBQ) were simulated evaluated, under various network and load/traffic conditions for supporting IP-to-MAC QoS in H2.
  - ✓ Resource Request (RR) - Resource Grant (RG) polling-based admission control algorithms was proposed and tested for IP-based H2. Two congestion control algorithms were also simulated.
- **Mobility Management for IP-based WLANs:** The IETF community has recognised the importance of this topic and it has been extremely active, proposing several solutions. From these studies, various solutions have been considered. After a brief comparison among them, our approach is being presented. We propose a hierarchical and multilayer/multiprotocol mobility management scheme to efficiently handle mobility for different types of services, using application-layer (e.g. SIP), along with various network-layer solutions (Mobile IP) and other micro-mobility approaches (local network mobility).

However, in order to make a suitable selection out of the various micro-mobility approaches, a thorough investigation was undertaken and completed with the performance evaluation of the nominated protocols (i.e. Cellular IP and Hierarchical Mobile IP). Summing up, the following technical solutions were investigated:

- ✓ Micro-mobility management, two existing protocols (Cellular IP and Hierarchical Mobile IP) were simulated and evaluated in terms of reliability and scalability for their suitability for WLANs.
  - ✓ Cellular IP was combined with Session Initiation Protocol (SIP), in order to provide mobility support on the application level. Their performance was compared with the performance of a layer-2 protocol (i.e. DHCP) in a real-time (testbed) scenario.
  - ✓ The suitability of SIP for handling mobility was discussed and possible solutions were proposed (*c-field* approach). Then, two global mobility management architectures were studied, simulated and evaluated, namely SIP-based and Mobile IP-based. The outcome a multilayer, multi-protocol and hierarchical solution could provide mobility management in All-IP 3G and WLAN infrastructures.
- **AAA services for secure and fast multimedia access over an All-IP WLAN:** The access to multimedia services through a WLAN infrastructure is achieved in two steps. Firstly, the access to the network is required. Secondly, the user needs to get authenticated, before being authorised to use the services provided by the SIP Provider. The AAA simulation work handled here, concentrated on the performance evaluation of the proposed AAA architecture for network access. We have thoroughly studied several AAA protocols (e.g. Radius, Diameter) for their suitability to fit into the proposed WLAN architecture. In the network authentication front, an extended study and simulation work was performed regarding the use of the IEEE801.x protocol for supporting fast handoffs in IP-based HiperLAN/2 systems.

## 1.4 Structure of the Thesis

The rest of the thesis is organised as follows:

Chapter 2 reviews the recent developments in the IP world and wireless communication areas. It also identifies the issues that need to be addressed for supporting IP services in the radio environment. At the end the chapter presents the recent developments and discussions in the All-IP networks, highlighting the points that are closely related to the MAC layer algorithms and the mobility management schemes.

Chapter 3 presents the expected network architecture for the WLAN and the respective protocol layering (based on H2). The description of the services and traffic characteristics such a network will support along with their modelling that has been used both in simulation and analytical studies, follows.

Chapter 4 discusses the suitability of the HiperLAN/2 wireless MAC protocol in supporting IP traffic. After a short presentation of its functionality, results both from simulation and theoretical analysis are illustrated. Finally, the necessary changes for 'optimum' and interoperable IP support are addressed in all layers (PHY-MAC-IP). In order to harmonise the findings, we propose a complete QoS framework, suitable for IP-based WLAN systems, while the interfacing and interworking between the PHY and MAC as well as the MAC and the (CL) IP are exhaustively discussed.

Chapter 5 furnishes the design and evaluation details of a complete suite for mobility management. After the description of various mobility management protocols, a hierarchical and multilayer scheme is proposed to efficiently handle mobility in WLANs for different types of services, using application-based (e.g. SIP) along with various network-based (CIP, MIP) mobility protocols. However, in order to make a suitable selection out of the various local mobility protocols, a thorough investigation was undertaken and completed with the performance evaluation of the nominated protocols (i.e. Cellular IP and Hierarchical Mobile IP). Finally, an integrated and multilayer mobility management solution for IP-based infrastructures was fully analysed, evaluated and proposed.

Chapter 6 outlines the interworking of mobility management and security (AAA) infrastructures by handling the secure access to multimedia services through a WLAN system. The access to the network is discussed and its importance is illustrated. Then, the user needs to exchange AAA messages, in order to use the services provided by the SIP Provider. The proposed AAA architecture as well as various protocols and scenarios are simulated, while their evaluations are based on performance metrics in terms of security, efficiency and reliability. Finally, an extended simulation study was performed regarding the use of the pre-authentication concept based on IEEE801.x protocol for supporting fast and secure handoffs for WLAN (HiperLAN/2) systems.

Finally, chapter 7 concludes the thesis, highlighting the major contributions of this work and discussing proposals for future work, in the IP-based WLAN research area.

It is suggested to the readers, in order to obtain a complete understanding of the different issues addressed in this work, to read chapters 2 and 3 first. Then, they can proceed to any chapter they find more interesting.

# Chapter 2

## 2 Background – Addressed Issues

This chapter addresses the problems and challenges faced in a multimedia, IP based, WLAN environment. Initially, a brief overview of the All IP network architecture is presented. The chapter then continues with the description of the current wireless communication systems, focusing on their MAC protocols, and explains the need for the WLAN network evolution. It briefly presents the major issues related to the WLAN concept, discusses its requirements and highlights the state of the art developments that are being discussed in the IETF, ETSI and other standardization and research bodies. The rest of the chapter gives an overview of related EU (IST) projects along with the design issues that have been considered and their major contributions in this research area.

### 2.1 Introduction

Not too long ago, communications meant voice and mobility meant cellular. But today we see that subscribers are increasingly relying upon diverse communications solutions for a complex array of voice, data, and multi-media needs, *many of which are being addressed by Internet/Intranet connected networks e.g. at offices, homes, shopping areas, transport facilities, and the like.* However, due to the progress in high speed networking and the unexpected success in mobile communications (i.e. the GSM growth, 3G potential), people are researching ways on how to achieve the integration of the wireless and cellular systems with the Internet. As these forces converge, the demand for new services, increasing bandwidth and ubiquitous connectivity continuously grows. The next-generation mobile systems will be based solely, (or in a large extent), on IP, offering end-user services, including voice telephony over IP networks. The objective is to offer seamless multimedia services to users who access an all IP-based infrastructure via a variety of different access technologies, meeting the demands of both enterprise and public environments anywhere and anytime. A key role of IP in next-generation mobile systems will be the efficient and cost-effective interworking between overlay networks for the seamless provisioning of current and future applications and services. IP is assumed to act as an adhesive to provide global connectivity, quality of service, mobility among networks, and a common platform for service provisioning across different types of access networks.

## **2.2 State of the Art**

### **2.2.1 Review of the Cellular Networks**

The Internet created a new paradigm in wireline networks with new service deployments heavily oriented toward IP applications. The reach of IP technology influence is now extending to encompass cellular networks where both operators and RAN vendors have embraced IP as the fountainhead for a new class of service application offerings. As a result, the third generation of cellular architecture definitions are focused on delivering end-to-end IP-based services to an increasing number of voice and data subscribers.

The goal of UMTS is to enable networks that offer higher bandwidth and can support a wide range of voice, data and multimedia services, all of them offered at competitive prices in a dynamic marketplace. These new UMTS networks will build on the success of GSM and on operator's existing investment in the development of customer base and infrastructure.

The first stage of service and network evolution is from today's GSM systems, through GPRS to commercial UMTS systems supporting interworking with GSM systems and reusing investment in GPRS. UMTS will enter the market at a time when fixed-mobile integration is becoming a reality, the telecoms, computer and media industries have converged on Internet Protocol (IP) as a shared standard, as well as, that data traffic accounts for a significant proportion of the traffic carried on the mobile networks.

Within this section, the new world of wireless architecture and the benefits of adopting a phased and emerging all-IP UMTS architecture, specifically engineered for the rapid deployment, as well as producing new revenue from IP-based applications and services will be briefly but concisely described.

#### **2.2.1.1 Standards Overview**

The International Telecommunications Union began its Third Generation mobile communications initiative in 1985 with the vision that there would be a single global standard. However, market conditions and drivers have proved to vary so widely in different regions that the ITU have now moved from this concept of one single system to a vision, which accommodates a family of systems. The goal is now to establish global roaming among the various Third Generation technologies, with this task currently being led by 3GPP/3GPP2 forums.

This family of systems is known as IMT-2000. The evolution paths from the first two generations of cellular standards ("1G" and "2G") through intermediate standards ("2.5G") to the Third Generation family of IMT-2000 standards ("3G") is illustrated in Figure 2-1.

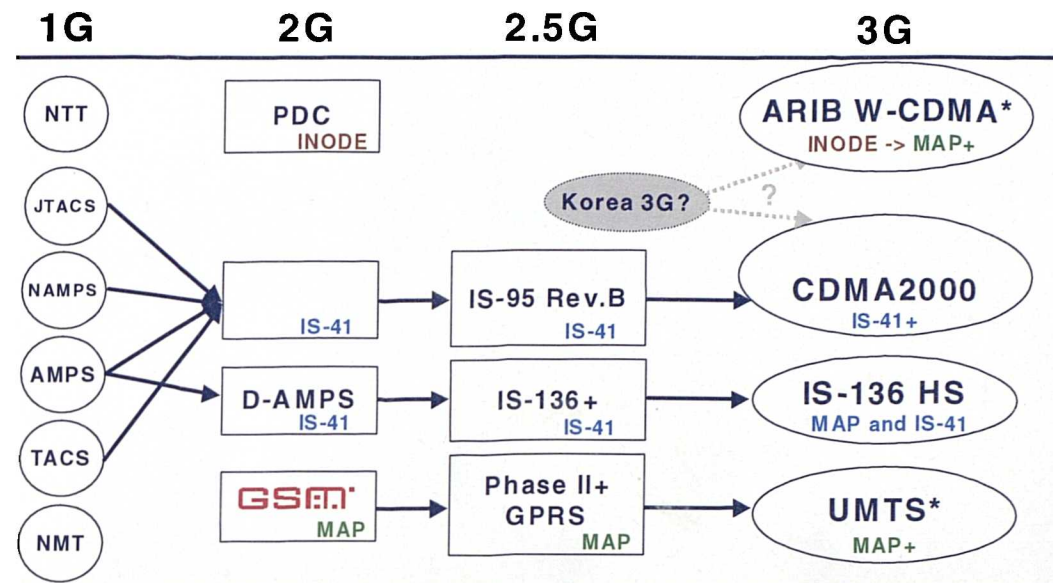


Figure 2-1: The IMT2000 Evolution of Standards [EVO-D2.1].

Regional choice of a Third Generation standard is based on a number of factors including customer requirements, existing investment in Second Generation systems and desire to support the local industry. Japan has been very active in Third Generation development, with capacity being a key driver in that market and NTT DoCoMo leading the run.

In America, there were different proposals for the Third Generation system e.g. cdma2000 for migration from cdmaOne, UMTS for migration from GSM1900 and UWC-136 for migration from IS-136 TDMA. In Korea, ETRI has established a consortium to develop Korean IMT-2000.

In Europe, the technology choice is UMTS with UTRA (UMTS Terrestrial Radio Access) defined as the air interface. Key Drivers are the implementation of advanced services (to develop the European Union's concept of the "Information Society") and to introduce increased competition through a new business model.

### 2.2.1.2 GSM Network

Global System for Mobile (GSM) is a second generation cellular standard that was developed to solve the fragmentation problems of the first cellular systems in Europe. It is the world's first cellular system to specify digital modulation and network level architectures and services. GSM was originally developed to serve as the pan-European cellular service and promised a wide range of network services through the use of ISDN. Its success has exceeded the expectations of

virtually everyone, and it is now the world's most popular standard for personal communications. Figure 2-2 presents an overview of the GSM architecture including the main components of the network and interfaces.

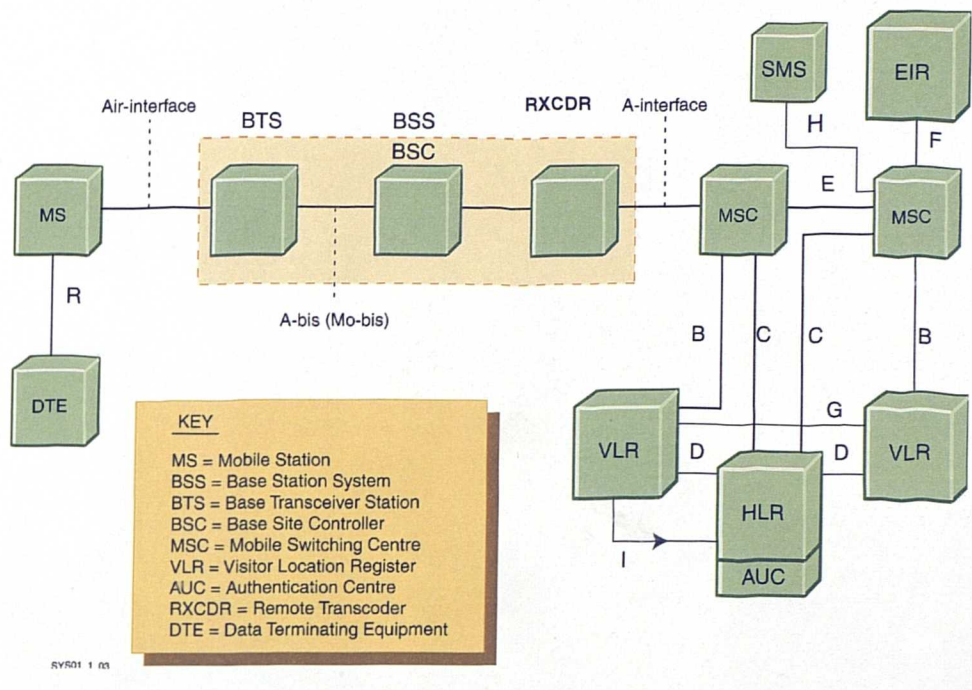


Figure 2-2: GSM Architecture Overview & Interfaces [EVO-D2.1].

### 2.2.1.3 Extend GSM with GPRS

GPRS was first proposed in December 1992, using a packet-mode technique to transfer high-speed and low-speed data and signalling in an efficient manner as well as optimising the use of network and radio resources. It is a non-voice value added service that allows information to be sent and received across a mobile telephone network, with theoretical maximum speeds of up to 171.2 kilobits per second (kb/s) e.g. using all eight timeslots at the same time. Its main purpose is to facilitate new data applications that have not previously been available over GSM networks due to the limitations in speed of Circuit Switched Data (9.6 kb/s), as well as enabling Internet applications such as the ones used by personal computers e.g. web browsing, downloading, FTP, chat etc.

GPRS makes use of the existing GSM network but it also requires that the backbone network to be consisted of relevant GPRS support nodes (GSNs). This network is a private (with access control) IP-network. Within this network, data is sent as packets in GPRS using the tunnelling protocol (GTP). The SGSNs roughly fulfil the same functionality as a MSC and a VLR of a GSM. GGSNs, provide access to the external packet data networks e.g. the public Internet. The GPRS functional architecture and interfaces are shown in Figure 2-3.

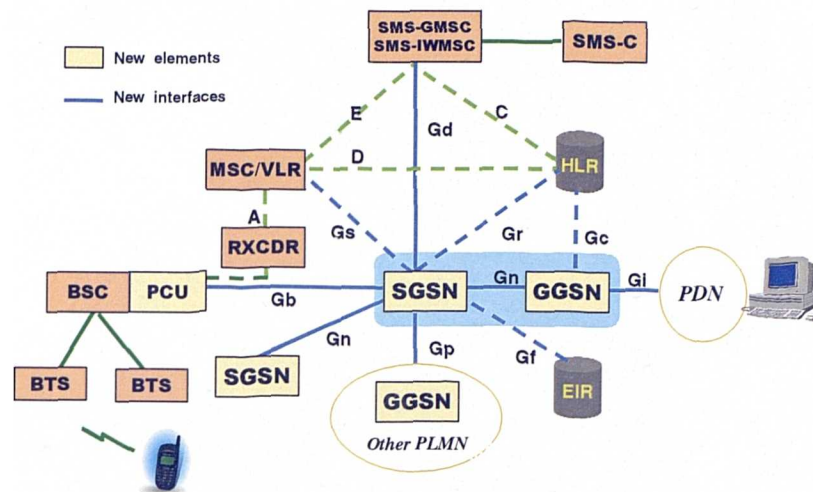


Figure 2-3: GPRS Functional Architecture and Interfaces [EVO-D2.1].

### 2.2.1.4 Phased evolution to UMTS

GSM network is mature and relatively stable, continuing to provide voice-centric services in 2.5G systems e.g. GPRS. In UMTS networks, a circuit-switched core network domain is being deployed to provide circuit-switched voice and other related services. In a later stage, an “all-IP” based core network architecture will support voice and multimedia service without relying on the circuit-switched core network domain. The evolution process, from standard development perspective, can be shown in Figure 2-4.

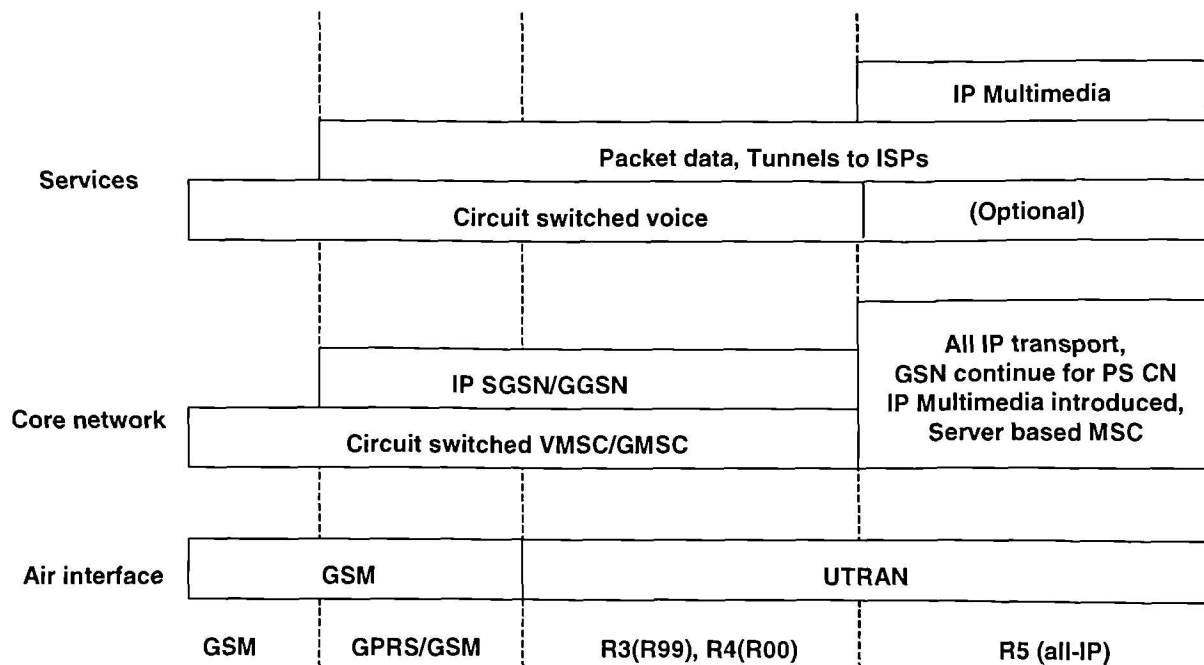


Figure 2-4: Phases of Standards Development.

Migration to an IP based wireless network will be a phased activity, aligning with the technology advances and standardisation progress, as well as to enable flexibility for meeting different operator's needs. It will adapt an evolution path through GPRS technology and the initial phases of UMTS (e.g. Release 99), followed up with the introduction of Voice over IP. The later UMTS releases will involve a single IP core network serving different access systems. The major phases are illustrated in Figure 2-5, as well as outlined below.

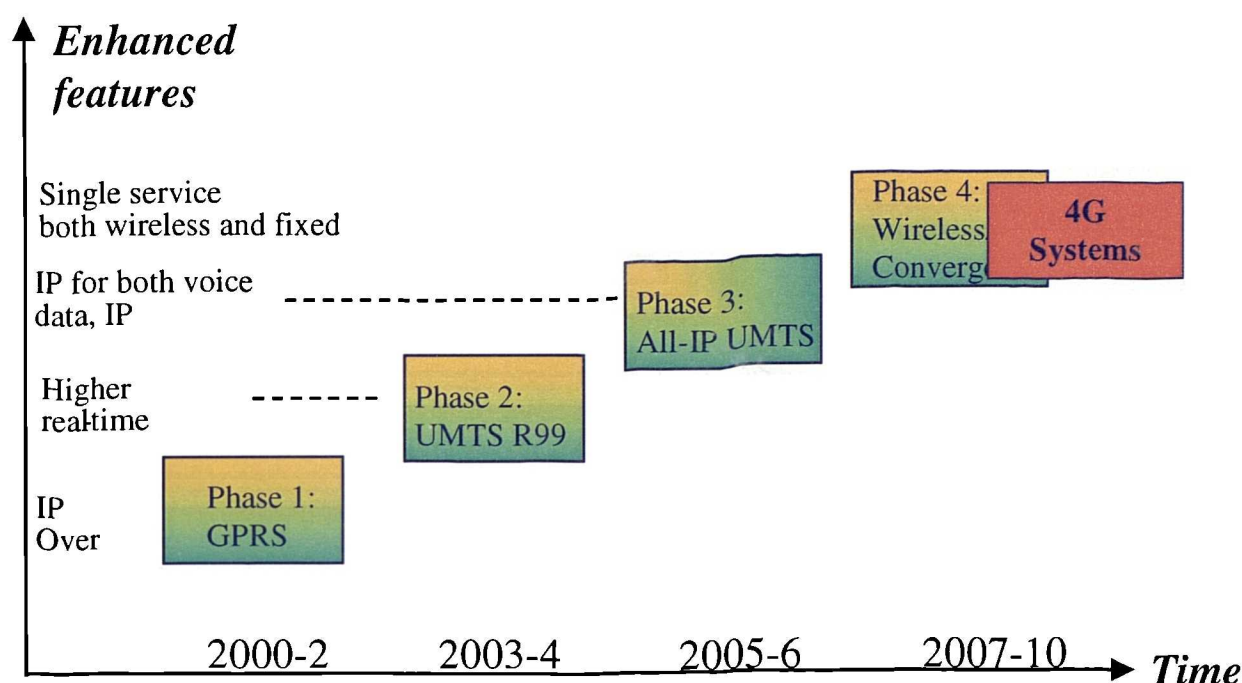


Figure 2-5: Phases of 2.5/3G Network Evolution.

**Phase 1 – GSM + GPRS.** The first step on the road to IP is GPRS, which is more efficient for packet access. A new IP based core network is introduced in parallel to the MSC, delivering IP to the user's terminal and enabling data or "always on" services. Operators can complement their existing circuit voice based services with IP based data applications, e.g. email and corporate network access.

**Phase 2 – UMTS R99.** GPRS is a good starting point for wireless IP applications, as it enables the operators to introduce the IP technology and its applications in the wireless environment. More importantly it will create market demand for new services, which are more bandwidth and QoS demanding. This will drive the deployment of UMTS.

The second phase sees the introduction of the new UMTS air interface with Release 99, providing higher data rates with better QoS support for real time services, allowing multimedia services like video/audio conferencing to be delivered over the IP path e.g. as new applications.

Prior to Phase 3 (e.g. an all-IP architecture based on 3GPP standards) being fully supported, an IP based solution for circuit-switch voice call will be supported with limited features e.g. IP MSC or MSC server. This solution will allow Greenfield operators to deploy UMTS network without investing in a circuit-switched core network. The IP solution for CS call would involve development of 24.008 call model function (i.e. to support standard R99 handsets) via Iu-CS interface.

**Phase 3 – UMTS all-IP Architecture.** The evolution to All-IP UMTS architecture has two aspects. From a transport aspect, an IP based UTRAN needs to be fully standardised, so that it can be used throughout the whole network as a unified transport infrastructure. From a service aspect, an all-IP standard will enable the voice services (e.g. used to be supported by MSC) to be delivered over IP transport infrastructure using VoIP technologies. This will help to achieve the goal of having a single core network for all services, including voice, multimedia and other data applications. The VoIP or IP Multimedia Subsystem (IMS) servers (including call control, media gateway and signalling gateway) can also provide voice and multimedia services over other IP access systems e.g. ADSL for nomadic devices.

As far as the voice service is concerned, two stages of deployment are involved. In the first stage, voice service can be supported by both circuit domain and packet domain. Operators can offload traffic from circuit domain to packet one and, at the same time try new service features with packet voice users. Later while the packet domain voice service is stable and mature, majority of traffic will be over packet domain.

For Greenfield operators, the evolution includes the feature enhancement on IP based MSC and the availability of voice support over Iu-PS interface.

**Phase 4 – Fixed/Wireless Convergence.** In the fourth phase the fixed IP access networks can be integrated providing seamless service between the wireless and wireline domains. Although VoIP services can be accessed via other IP access medium with phase 3, service level integration is to be enhanced within phase 4. The mobility management databases and subscriber profile database will be enhanced for both wireless and fixed domains and provide the mapping between them. This allows a single subscription account to be used for both cell phone and laptop use from both wide-area cellular and fixed access schemes. At the same time, the core network and radio access network will evolve to a true peer-to-peer architecture which gives a more resilient and scalable solution that flexibly and more cost effectively meets the increasingly volatile voice and data traffic demands on our networks.



- IP call control and service control interworking with legacy system, including roaming between all-IP network and legacy circuit-switch networks.
- Spectrum efficiency – normal transport of VoIP would impose a heavy penalty on spectrum efficiency, as the overhead of various layers is quite high. A better compression scheme or removal of IP headers are under discussion.
- QoS is certainly a key for any real time service over IP transport.

The implications to the UMTS architecture due to an all-IP migration concept are being illustrated in Figure 2-6.

## **2.2.2 Review of the Wireless Local Area Network Standards**

### **2.2.2.1 IEEE 802.11b**

The 802.11 protocol [802.11-WG] is a standard for WLANs. It enables a group of wireless computers to share the same frequency and space and provides reliable means of wireless data transfer. The 802.11 standard defines physical layer parameters for wireless transmission and MAC layer for the use of a shared medium.

#### ***2.2.2.1.1 Network topology***

The WLAN IEEE802.11 standard suggests two ways to configure a network, Ad-hoc and Client/Server networks.

##### **2.2.2.1.1.1 Ad-hoc network**

This simple network consists of a set of computers, which are able to communicate with each other, so there are no fixed points and no structure. The stations communicate directly with each other on a peer-to-peer level, sharing a given cell coverage area. A good example of an Ad-hoc network is a meeting where employees are using their laptops to communicate and share data with each other.

##### **2.2.2.1.1.2 Client/Server network**

In this type of network used in WLANs, the mobile nodes can communicate with some fixed network Access Points (AP). The main function of an AP is to form a bridge between wireless and wired LANs. Thus, all communications between stations or between a station and a wired

network client go through the AP. This also means that a wireless user can roam from a WLAN coverage area to another since WLANs areas, in this case, are connected together via a wired network and through APs. This structure is very similar to the cellular networks architecture and provides better throughput performance. The APs could also be IP Routers that control the traffic in and out of the Internet. Thus, routing wireless packets into and out of the Internet as shown in the figure below (Figure 2-7).

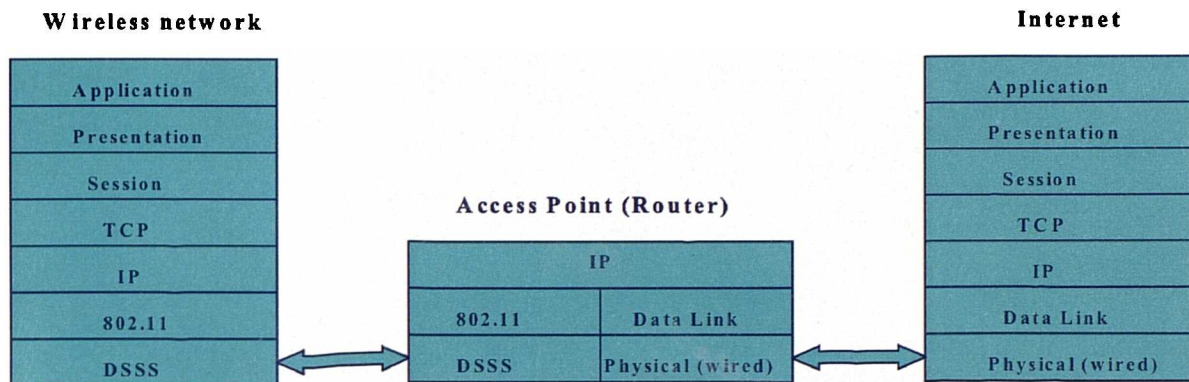


Figure 2-7: Protocol Reference Model for Wireless LANs and the Internet.

#### 2.2.2.1.2 IEEE 802.11 layers

As mentioned earlier, the IEEE 802.11 standard is limited to the PHY and the MAC layers of the network. The PHY layer handles the transmission of data between nodes using a RF technology like, DSSS, FHSS or infrared (IR) pulse position modulation. The RF transmission standards are defined for operation in the 2.4 GHz frequency band typically occupying the 83 MHz of bandwidth from 2.400 GHz to 2.483 GHz. The infrared standard operates in the 850 to 950 nm band with peak power of 2 W. The physical layer supports two data rates, 1 and 2 Mbps. The MAC layer is a bunch of protocols, which is responsible for maintaining order in the use of a shared medium. The 802.11 specifies a carrier sense multiple access with collision avoidance (CSMA/CA). This means when a node receives a packet to be transmitted, it first listens to ensure no other node is transmitting. If the channel is clear, it then transmits the packet. Otherwise, it chooses a random “backoff factor” which determines the amount of time the node must wait until it is allowed to transmit its packet. As long as the channel is busy, the backoff counter does not change. When the channel is clear, the transmitting node decrements the counter. Once it reaches zero, the packet is transmitted. Since the probability that two nodes choose the same backoff factor is small, collisions between packets are minimised.

The MAC layer operates together with the physical layer by sampling the energy over the medium transmitting data. The PHY uses a clear assessment algorithm to determine if the channel is clear.

This is accomplished by measuring the RF energy at the antenna and determining the strength of the received signal. If the latter's strength is below a specified threshold the channel is declared clear and the MAC layer is given the clear channel status for data transmission.

### ***2.2.2.1.3 IEEE 802.11 security***

The 802.11 standard provides security for both users and providers using two methods: authentication and encryption. Authentication is used to check whether a station is authorised to communicate with a second one in a given coverage area. In the Client/server mode, authentication is established between the Access Point and each station. Encryption means that data is not sent in clear. This will protect both the users and the providers from fraud. The 802.11 standard supports for instance 802.1x and WEP (Wired Equivalent Privacy) protocols to provide authentication and encryption (these protocols are further described in Chapter 6).

### ***2.2.2.1.4 The IEEE 802.11 family***

#### ***IEEE 802.11b***

IEEE 802.11b was ratified in September 1999, and is getting much attention. This standard essentially legitimises Ethernet-like speeds for Wireless LANs. Although there are advantages and disadvantages to both spread spectrum technologies (FHSS, DSSS), DSSS was chosen as the basis for the IEEE 802.11b standard. This standard operates in the 2.4 GHz and uses DSSS for low data rates (1Mb/s with BPSK and 2 Mb/s with QPSK modulation scheme). In order to achieve higher data rates a high-rate (HR) extension is included to the DSSS modulation technique (HR-DSSS). This extension provides data rates of 5.5 and 11 Mb/s if Complementary Code Keying (CCK) is employed as the modulation scheme. An optional mode is based on the use of the PBCC (Packet Binary Convolutional Code) technique is also provided by the IEEE 802.11b standard. IEEE 802.11b high-rate products started shipping in late 1999 (except Aironet which released its 11 Mb/s product in late 1998 before the standard was actually ratified). The IEEE 802.11 Working Group currently has several Task Groups working on MAC enhancements. IEEE 802.11 Task Group E is working to improve and manage QoS, while Task Group I is developing enhanced security and authentication mechanisms. Task Group G is currently debating over a technology for a higher speed extension to the IEEE 802.11b, which promises +20 Mb/s speeds.

#### ***IEEE 802.11a***

The technology behind this standard is similar to that of HiperLAN/2, with each using an Orthogonal Frequency Division Multiplexing (OFDM) for data transmission. Products will

operate in the 5 GHz band, with speeds up to 54 Mb/s. OFDM technology is a modulation scheme that supports higher bandwidth than that of DSSS technologies supported under IEEE 802.11b. Due to the fact that the IEEE 802.11a MAC layer is the same as that of the IEEE 802.11b, many vendors are tending to push this standard.

#### IEEE 802.11h

Task Group H's goal is to add features to the current 802.11 5 GHz standard so regulatory bodies outside of the United States will allow for 5 GHz 802.11 products. Currently under ETSI rules and regulations, IEEE 802.11a products are illegal in Europe. Task Group H purpose is to provide the IEEE 802.11 MAC and PHY with network management and control extensions for spectrum and transmit power management in 5 GHz license exempt band. It also plans to provide improvements in channel energy measurement and reporting, channel coverage in many regulatory domains, as well as providing Dynamic Frequency Selection (DFS) and Transmit Power Control (TPC) mechanisms. Another task is to ensure that certain elements of HiperLAN/2 are adopted by IEEE 802.11a in order to make the latter a global standard.

#### IEEE 802.11g

The recent draft standard, a compromise proposal from TI and Intersil Corp., combines elements from the final two independent proposals that were originally considered for the IEEE 802.11g standard. It offers backwards compatibility with IEEE 802.11b and a set of data-rates from 20 Mb/s to 54 Mb/s in the 2.4 GHz band. In addition, the new draft standard also contains modes of operation using PBCC and CCK-OFDM modulation techniques. This integration of modes will enable consumer products to support multiple modulations and provides a clear bridge from 11Mbps to 54 Mb/s data rates. Although, much of the controversy in the WLAN market has centred on IEEE 802.11a versus H2 in the 5 GHz spectrum, silicon vendors such as Intersil and TI recognize the cost implications of moving to 5 GHz and try to offer a fast solution in the 2.4 GHz band at cheaper prices than comparable WLAN devices operating in the 5 GHz band. However, the benefit of the 5 GHz WLAN products would be that they do not suffer from interference problems.

### **2.2.2.2 ETSI HiperLAN/2**

The demand for high speed Internet access and the emergence of multimedia applications are leading towards a new area, radio-based broadband access networks. Currently, in Europe the correspondent project is called BRAN, and the standardization process is being coordinated by ETSI. The outcome, an all-new high performance radio technology is called HiperLAN/2 and is

mainly directed for LAN environments. The profile of the ETSI HIPERLAN family throughout their development stages is shown in Figure 2-8.

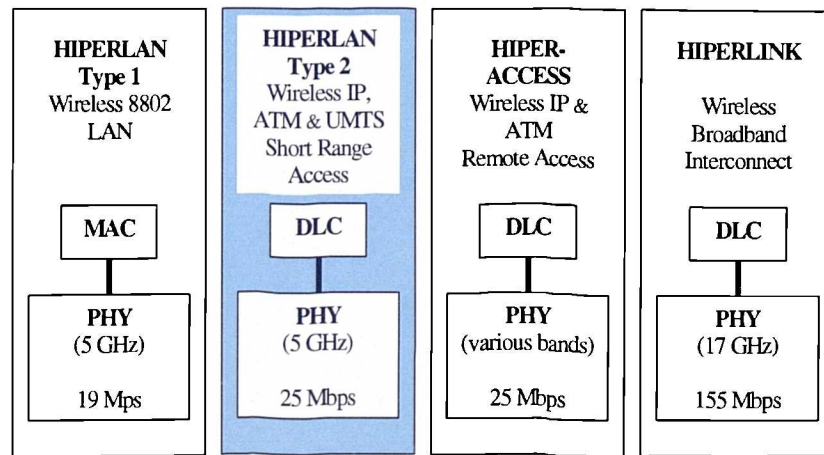


Figure 2-8: The past, the present and future of ETSI HIPER-family.

HiperLAN/2 [H/2-00] has three basic layers of its own; PHY, DLC, and the Convergence layer (CL). Again, on the top of these are the traditional higher layers.

The PHY is OFDM based and it is a burst with a preamble and a data field.

The DLC includes functions for medium access and transmission (user plane) as well as terminal/user and connection handling (control plane). Thus, the DLC layer consists of a set of sub-layers: the MAC, EC (Error Control sublayer) and RLC (with the associated signalling entities DCC, RRC and ACF) protocols.

The CL (Convergence Layer) adapts service requests from higher layers to the service offered by the DLC and converts the higher layer cells (in ATM) or packets (in Ethernet or IP) into a fixed size that is used within the DLC.

The MAC instances in the AP and the MT are responsible for controlling the access of the H2 radio interface. In this centrally controlled approach the AP assigns the radio resources within the H2 MAC frame. This assignment of resources for the individual MTs and their connections is not static, but may change from MAC frame to MAC frame with a very high dynamicity.

The fixed length H2 MAC frame ( $t_{frame} = 2 \text{ ms}$ ) consists of four major phases as shown in Figure 2-9.

- Broadcast Phase-broadcast

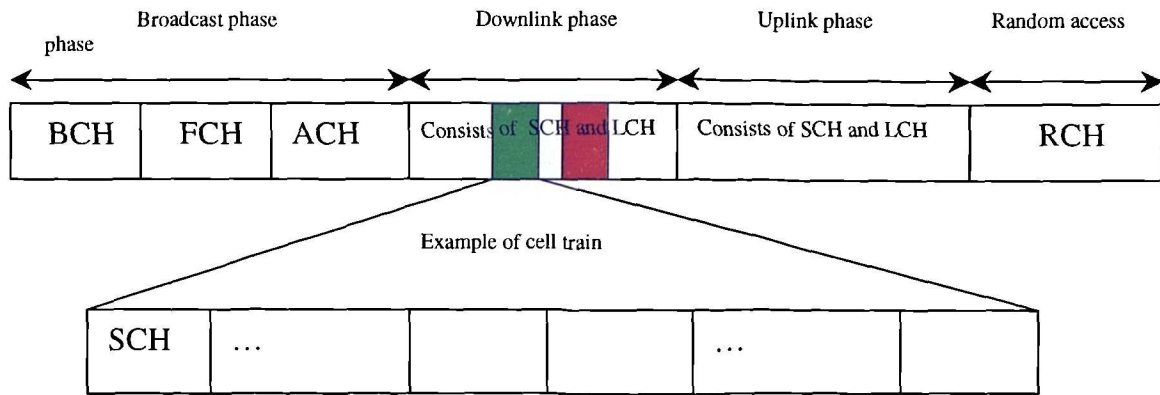


Figure 2-9: HiperLAN/2 MAC Frame.

- Downlink phase-downlink direction
- Uplink Phase-uplink direction
- Random Access Phase-uplink direction

H2 operates in the unlicensed 5 GHz frequency band, which has been allocated for wireless LANs (however, this may vary across Europe). HiperLAN/2 systems can be deployed in offices, classrooms, homes, factories, hot spot areas and potentially wherever radio transmission is an efficient alternative or a complement to wired technology. H2 makes use of a modulation method called OFDM – supporting various bit rates and modulation schemes - to transmit the analogue signals, thereby achieving higher transmission rate than IEEE 802.11 (at the physical layer up to 54 Mb/s and on the CL up to 25 Mb/s). Connections between the MTs and the APs are time-division-multiplexed over the air interface. There are two types of connections, point-to-point (bi-directional) and point-to-multipoint (unidirectional towards MT). In addition, there is also a dedicated broadcast channel through which traffic reaches all terminals transmitted from one AP. H2 enables QoS support, automatic frequency allocation, power save, security (authentication and encryption) support and is planned to include provisions for a flexible forwarding mechanism for ad-hoc networks.

Furthermore, H2 major functions can be reused for FWA systems, which is not the case for 802.11a [IST-STRIKE].

### 2.2.2.3 Regulation Status Overview

The proposals operating in the 2.4 GHz unlicensed band face no problems with regulations worldwide. This frequency band is free for everyone to implement a WLAN. The main problem for these networks is RF interference with other equipment, such as cordless phones, microwave ovens, and neighbouring WLANs.

For the 5 GHz band the picture is more confusing: there is no global agreement for the frequency sub-bands that can be used for the same purpose and with the same limitations. Especially for Europe, ETSI suggested frequency allocations but each country sets its own rules regarding licensing and usage.

Furthermore, Europe does not allow 802.11a products to be sold in Europe, unless they involve TCP and DFS. The 802.11h proposal aims at these enhancements.

## **2.3 Why HiperLAN/2?**

We decided to face the research challenges of this work by addressing the ETSI HiperLAN/2 and IEEE 802.11 systems. Despite that the IEEE 802.11 wireless Ethernet technology is a mature technology with well-known drawbacks, the H2 standard has been a major breakthrough in the European research initiatives in the WLANs arena. Motivated by these facts, we have decided to redress the balance and project and utilise the strengths, while isolating the weaknesses of both technologies, in a complementary rather than competitive manner.

Therefore, in order to catch up with the vigorous technological trends and the ever-changing environment of the telecommunications scene, and thus exploit the research outcomes, we have prioritised and divided our activities. To investigate and address research problems, solve open standardisation issues and propose innovative frameworks, we have selected the recently standardised HiperLAN/2 system. While, we address more pragmatic problems on a real life Wireless Network Test-bed (UniS WNT), which is based on the IEEE 802.11 system.

## **2.4 All-IP WLAN**

### **2.4.1 Overview**

An “all-IP” mobile/fixed converged network suitable for use with a very wide range of access technologies and including the use of ad-hoc and moving networks is the subject of exhaustive research activities in Europe and around the globe. The solution will essentially deliver IPv6 packets and provide the ancillary functions necessary to support this for a wide range of users including:

- Network discovery
- Security and Network cost and information exchange – allowing users to connect securely to networks (including ad-hoc segments) with certainty of cost and integrity of signalling and network identity

- Quality of service (meaning bandwidth, delay and packet loss)
- Mobility – allowing users to maintain connectivity – even with complex connections (e.g. WLAN downlink and GPRS uplink)

The term all-IP is much overused – in the context of this work however, this means:

- IP packets are transported using IP routing to the access points
- QoS is dealt with at the IP layer – mapping layer 2 (DLC/MAC) QoS over the wireless hop
- The business model is potentially open - and out of scope
- The coupling to higher layers and service creation is through a standardised interface
- WLAN technologies can be used to access the core network.

The Figure 2-10 shows an all-IP WLAN stack – with the interfaces included:

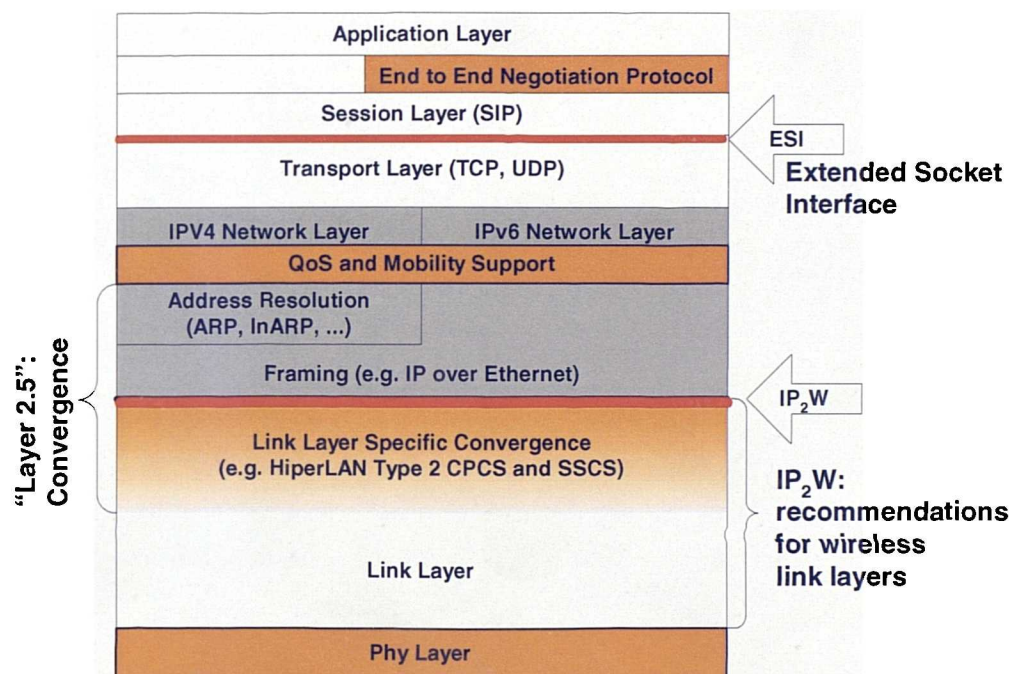


Figure 2-10: WLAN network stack.

The major issues involved in delivering the all-IP WLAN network vision are:

- Mobility management (including address management and auto-configuration)
- Quality of service (end-to-end, including specification of QoS and domain solutions)
- Security (including mutual network/user authentication)
- Business models (including network discovery and negotiation)

### 2.4.2 Mobility Management

Mobility management will become a fundamental issue in 4G Wireless LAN networks. A transparent IP-based WLAN system will provide mechanisms that enable seamless roaming between multiple, heterogeneous sub-networks. In more detail the following mobility capabilities should be provided by future mobile network architectures:

**User Mobility** or personal mobility. This is the ability of the user to access his personalized network services, while he is away from his home network. If a user can seamlessly access all services provided to him by his home domain though he is in another domain, then these domains support user mobility.

**Terminal Mobility.** This is the ability of the network to maintain continuous services even when the user's terminal changes locations. Terminal mobility becomes more challenging in case of streaming applications with strict QoS requirements and devices with multiple, heterogeneous radio access technologies. Handing of multi-homed terminal mobility will require inter-technology handovers and resources management initiated by Level 2 triggers and supported by enhanced signalling protocols.

**Network Mobility** is the ability of the network to support roaming of an entire sub-network, structured or ad-hoc. Cases of mobile networks could include ad-hoc and Personal Area Networks, consisting of network appliances like digital cameras, mobile phones, laptops/palmtops, PDAs, as well as wireless LANs that are constantly moving as part of a vehicle of some sort (e.g. airplane, train, boat, bus or taxi). In this case the core and access networks should be able to recognize the moving network as a special roaming node.

When a terminal or a network changes the point of attachment then it handovers to a new network. Handover between heterogeneous access networks may be categorised based on the type (horizontal vs. vertical handover) and the topology of the access networks (local vs. global handover). In horizontal handover, the terminal roams between wireless cells of the same type that belong to the same or different IP segments. Coverage and data rates remain the same, simplifying the roaming. In vertical handover, the terminal roams between wireless cells with different coverage and hierarchical layers. On the other hand, local handover or micro-mobility focuses on the optimization of handover over the air, and assumes that a direct radio link exists or may be established between the source and the destination base stations, while in global handover or macro-mobility, the overall network is involved in order to achieve more efficient utilisation of the network links.

Mobile IP [MIP-02] has been the protocol that initially addressed the mobility management problem. In the Mobile IP model, when a mobile terminal roams to a foreign network, an agent in

the terminal's primary IP sub-network HA is delegated to receive the data targeting the terminal, tunnelling them to an agent in the visiting network FA, who forwards them to the current position of the terminal. The main drawback of MIP is that it introduces significant network overhead. Overhead may be identified in terms of increased delay due to triangular routing, packet loss due to registration delays and increased signalling due to frequent terminal registration. Establishment of new tunnels may introduce additional delays in the handoff process, causing packet loss and delayed delivery of data to applications, while pre-established tunnels may further complicate the overall network architecture. Route optimization [MIP-RO] can improve the performance, but cannot provide acceptable service quality, especially for streaming applications like VoIP.

However, the network overhead is drastically increased and the QoS experiences noticeable degradation, in the case when the mobile hosts change frequently their point of attachment. Based on the assumption that the majority of frequent handovers take place between nodes of the same regional network, many IP micro-mobility protocols have been proposed aiming to provide fast, seamless, local mobility.

IP Micro-mobility protocols reduce the round-trip delay and eliminate the end-to-end registration signalling overhead, by handling local movements of the mobile terminals transparent from the global Mobile IP network. This is achieved by maintaining regional location databases that map mobile hosts identifiers to location information, which is used for local delivery of the packets. Moreover, in order to further reduce registration signalling, IP micro-mobility protocols introduce IP paging techniques. As the number of mobile users increases, IP paging may dramatically help in order to restrict the associated signalling overhead from rising exponentially. Just like cellular networks that decrease network overhead by paging mechanisms, micro-mobility protocols take into account the operational mode of the terminal and maintain approximate location information for the idle users. In this way, intermediate signalling messages generated by frequent local movements of an idle terminal are suppressed, with the cost of increased delay when a new call is initiated. In general, micro-mobility management protocols may be categorized into two large groups [CAMB-02].

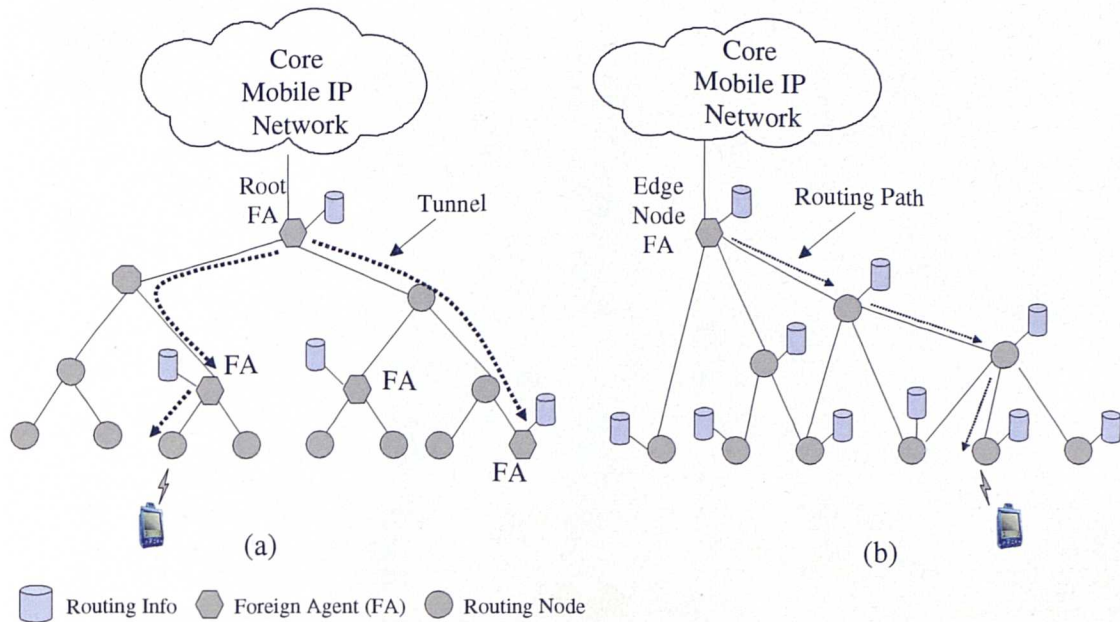


Figure 2-11: a) Hierarchical tunnelling and b) Explicit Terminal Routing Approaches.

**Explicit Terminal Routing approach (or Host-specific forwarding):** The protocols that fall in this category avoid the overhead introduced by encapsulation and tunnelling, with the cost of introducing explicit routing entries for each terminal. Instead of location database, each host keeps a (kind of) routing table that contains an index to the next-hop node in the path towards the terminal's point of attachment. Data packets that address a specific mobile terminal are delivered encapsulated from terminal HA to a FA located at the edge node of the visiting network. In the sequel, each node knows how to route the packets in order to reach the terminal's point of attachment. When a terminal moves between two points of attachment, location update information is propagated explicitly (via signalling messages) or implicitly (via data messages snooping) up to an anchor or crossover node, which is located in both the old and the new path or up to the edge FA. Examples of micro-mobility protocols that use the Explicit Terminal Routing approach are the Cellular IP, the HAWAII and the UniWA (Unified Wireless Access). Also in this category may fall protocols like "iwander" that are based on Ethernet Switching and layer 2, MAC to port number binding. These protocols do not require a hierarchical network structure, but may face scaling limitations due to the increased number of routing information especially at the nodes closer to the Edge node.

**Hierarchical Tunnelling approach:** The protocols that fall in this category incorporate a number of FAs in a tree-like structure. Each FA maintains part of a location database. The entries of the database are created and maintained by registration messages transmitted by the mobile hosts that have visited this sub-tree, and contain the address of the next lower-level FA in the path towards the mobile host. Data packets that address a specific mobile terminal are delivered encapsulated from terminal's HA to the root FA of the visiting tree network. Each FA in the lower layers of the

tree structure that is the end-point of a tunnel has to decapsulate the packets, search its location database for a corresponding entry, and then re-encapsulate and forward them through new tunnels down to the terminal's point of attachment. When a terminal moves between two leaves of the tree, it propagates location update messages up to a tunnel end-point, which is located in both the old and the new path. Examples of micro-mobility protocols that use the Hierarchical Tunnelling approach are the Hierarchical Mobile IP(v6), etc. Hierarchical tunnelling micro-mobility protocols may have to introduce new control/ signalling messages or require a hierarchical tree structure.

The last years, there has been considerable debate in IETF on suitable fast and seamless handoff extensions. However, the 4G networks architecture poses for very intelligent distribution of roaming functionality and enhancements in the existing signalling and transport protocols. In the following years, special focus will be put on many research areas including: i) Scaling over large networks (e.g. large number of terminals, large number of cells per access router), ii) signalling overhead vs. response delay time (Layer 2 vs. Layer 3 signalling vs. control protocols), iii) location update vs. IP paging, and iv) Roaming Information (Terminal velocity/direction/moving model/profile vs. Open Sessions requirements/Contract vs. Network resources/congestion).

### **2.4.3 End-to-end (ete) QoS**

The range of applications and services is constantly growing creating a huge amount of different usage scenarios: different applications on various types of terminals connected to various types of networks (which again are connected to other networks). When some applications work fine with best-effort forwarding service in the connecting network, certain applications require something better, something the application can depend on, and certainly better than best-effort.

One way to enhance the quality of IP-based services is to bring the content closer to the end-users by replicating content, for example, popular content of web sites or mission critical content of private companies, in servers geographically closer to users. These content distribution networks have the benefit of avoiding congestion and latencies in backbones and between ISPs. However, we still need mechanisms to handle the different data flows locally inside the access networks, prioritise multimedia traffic or bulk-data transfers, for example.

Currently we can identify two ways to provide differentiated treatment of flow: with DSCP or through application signalling. In the DiffServ-based solution an end host can mark the transmitted packets with proper DSCP values to trigger certain service responses. The problems of this approach are that not all network support the DSCP values or may have different meanings for the values, the initial value may not persist end-to-end, the approach does not guarantee the

service, only tells of a relative priority, and a receiving host has no mechanism to affect dynamically the code point used.

The most popular QoS signalling protocol, in terms of research interest, is the Resource Reservation Protocol and the Integrated Services specifications. Other interesting protocols are, for example YESSIR, Boomerang and INSIGNIA. All of the protocols have their own history behind the design and are targeted at certain usage scenarios and applications. In tomorrow's world of hybrid networks and services, none of these protocols can provide the QoS in all situations. An analysis of these protocols can be found in [MAN-03].

When we discuss QoS in a mobile environment, the coupling of QoS and mobility management is still one of the key problems - how to keep providing the same level of quality to the packet flow during and after a handover. For example, the GSM networks offer voice and data services and support mobility of customers – an ongoing call is very rarely dropped due to movement. We need the same level of support for mobile IP-based networks. However, no technology currently exists that would allow for fully seamless mobility and still be commercially feasible. The question is always about a balance between resource utilisation and level of service.

There have been several designs for extensions to RSVP to allow for more seamless mobility. One solution is presented in [RAA-01] which couples RSVP and the mobility management mechanisms and proposes small extensions to localize the RSVP signalling. The extension allows the mobile host to be in charge of both upstream and downstream reservations and re-setup reservations on both directions immediately following a handover. Also, context transfers studied in the IETF Seamoby WG aim at enhancing and minimizing signalling needed during handoffs.

Another group of examples are protocols based on advance reservations, where neighbouring access points keep resources reserved for mobile nodes moving to their coverage area. When a mobile node requests resources the neighbouring access points are checked too and a passive reservation is done around the mobile nodes current location. The problems with these schemes are that they require topological information of the access network and usually advance knowledge of the handover event. Furthermore, the way the resources reserved in advance are used in the neighbouring service areas is a somewhat open issue. A good overview of these different schemes can be found in [MOO-01].

What is missing is a common QoS signalling framework, one that application designers could use to make their applications QoS-aware. Currently, there exists too broad a range of QoS mechanisms, both for signalling and provisioning, which makes it impossible for an application designer to decide on a technology - moreover, because the very same signalling protocol must be supported by the network, too. A flexible way of making applications QoS-aware is to use a

middleware that can map the applications QoS requests to QoS technology used at the IP-layer. This is discussed in more detail in Chapter 4.

Still, at the end of the day, the deployment of various QoS mechanisms is a classic "chicken and egg" problem. We need new business models before any operator is willing to invest in expensive new QoS mechanisms. Take, for example, the General Packet Radio Service: the specifications included a number of different QoS parameters, as for instance, bandwidths and priorities, but nobody is currently using them.

On the other hand, we also need a clear path in deploying QoS in IP networks incrementally, network by network, application by application. Currently, most parts of the technologies needed for providing distinguished services in a best-effort IP network are already available. An architecture that combines per-application signalling and traffic classes within the backbones would be a very interesting solution [BER-00]. There are still some open issues, but the enabling technologies for QoS-enabled services are already here. The question is, which operators make the first move and decide on the migration path from best-effort to the Integrated Services network as foreseen by the IETF in the mid 1990's.

#### **2.4.4 Security**

Some of the key issues, that every all-IP based WLAN should address, are:

- Mutual authentication of local networks with user service providers
- End-to-end encryption
- Security of signalling traffic
- Interaction of security and efficiency considerations

Mutual authentication is crucial for users to trust ad-hoc and moving networks. Currently the sort of services used – Internet access and VPN corporate access – do not justify advanced security features. With advanced services, such as Video/Voice, then users need to protect their identity and authentication mechanism as well as be certain they are connected to a genuine network company. Mechanisms and solutions for mutual authentication of local networks (such as community networks or coffee shop networks, etc) – to allow users to discover, in a secure way, what aspects of the local network can be trusted.

Current mobile networks approach these issues either by providing a complete, integrated solution, as in cellular systems, or by a very low level system for Internet access, as in WLANs.

### **2.4.5 Related projects**

Various ongoing EU projects in the IST Programme deal with the evolution of UMTS, WLANs and the All-IP Networking architecture.

**EVOLUTE** (<http://evolute.intranet.gr/>) aims to design, specify and develop an all IP-based network infrastructure that will offer seamless multimedia services to users who access the network via a variety of different wireless technologies. For their importance in the near to medium term evolution of global mobile telecommunications, two types of wireless technologies have been selected, being the UMTS and the WLANs. Both real-time (e.g., mobile telephony) and non-real-time (e.g., mobile web access) services are used to test and prove the efficiency of EVOLUTE architecture.

**WSI** (<http://www.ist-wsi.org>) focuses on research capacities on strategic objectives and essential issues through active information and the formation of Innovation Cells, offers a platform that enables synergy among projects in the wireless area of the IST Programme and disseminates information to accelerate the emergence of standards.

**WINE-GLASS** (<http://domobili.cselt.it/WineGlass>) objective is to exploit enhanced and/or new IP-based techniques to support mobility and soft-guaranteed QoS in a wireless Internet architecture incorporating UMTS and WLANs, and to explore their potential in enabling location- and QoS-aware application services for wireless mobile users.

**BRAIN** (<http://www.ist-brain.org>) facilitates the development of access to existing and emerging IP-based broadband applications and services for mobile users in global markets. It proposes an open architecture for wireless broadband Internet access, which will allow an evolution from fixed Internet, emerging wireless/mobile Internet specifications and UMTS/GSM. It also facilitates new business opportunities for operators, service providers and content providers to offer high-speed (up to 20Mbps) services complementary to existing mobile services.

**MOBY DICK** (<http://www.ist-mobydick.org>) facilitates the development of seamless access to existing and emerging IP-based applications and proposes an architecture for wireless Internet access by developing new mechanisms for QoS support after and during handover and charging.

**MIND** (<http://www.ist-mind.org>) aims to extend the concepts of IP mobile networks. This includes: new network topologies - ad-hoc, self-organising and meshed networks; enhanced support for QoS, ad-hoc networks and self organisation at all layers of HiperLAN/2; QoS support in IP-based mobile networks; investigation of the spectrum requirements for systems beyond 3G.

Finally, **6WINIT** (<http://www.ist-6winit.org>) project aims to validate the introduction of the new mobile wireless Internet in Europe, based on a combination of the new Internet Protocol version 6 and the new wireless protocols (GPRS and UMTS/3GPP).

## **2.5 Summary**

In this chapter, a number of issues related to the All-IP infrastructures and the wireless/cellular technologies were presented. Emphasis was given to the basic characteristics of the IP and the different wireless/cellular schemes, and an investigation of their suitability in supporting IP traffic was attempted. The need for integration between the wireless networks and the IP was discussed. The development of a common protocol infrastructure (QoS, mobility management, security etc) for Wireless LAN based on the IP protocol was pinpointed. Furthermore, the introduction of mobility management in IP based wireless networks, the interworking among different entities in the network, and the QoS requirements of the multimedia traffic, are some of the areas that need further investigation. Highlights of the technological progress in the IP-based WLAN research area were also given in a separate section.

# Chapter 3

## 3 All-IP WLAN System Design

### 3.1 Introduction

The chapter initially discusses the system architecture of an All-IP Wireless LAN; more emphasis is given in the centralised approach (infrastructure-based WLANs). In the following section, the protocol reference model of the IP-based WLAN protocol stack is presented; the HiperLAN/2 case is considered. The chapter then continues describing the IP service definitions from ETSI as well as the ITU-T points of view. Furthermore, suitable traffic models for each service, which have been used in our simulation studies, along with an estimation of their requirements, in terms of IP-based QoS, are being presented.

### 3.2 System Architecture

In this section the architecture of the IP-based WLAN will be presented; the different entities and the protocol stack will be also discussed. As has already been discussed, two network architectures have been considered for this research and simulation activities, namely the IEEE 802.11 and the ETSI HiperLAN/2. In the first case, more realistic scenarios are examined, while the latter is used to add a research flavour in our studies. Examples of H2 DLC/MAC protocols are mainly given. While, the best suited architecture for the next generation all-IP WLANs is proposed.

#### 3.2.1 Protocol Reference Model

The physical layer of HiperLAN/2 is based on the signal modulation OFDM with several sub-carriers and forward error correction schemes, which can support variety of channel configurations [ETSI-475].

The MAC protocol currently stands as it was briefly discussed in Chapter 2, section 2.2.2.2.

The H2 Convergence layer, functions as an adapter between the DLC and higher layers (e.g. IP, application layers). Two types of CLs have been specified by ETSI BRAN, an ATM-cell based CL and a packet based CL. This research mainly focuses on the packet based CL (since we are only dealing with IP-based traffic and infrastructures), which offers services to HL (Higher

Layers) that use packets or frames of variable size or fixed size (e.g. Ethernet and/or the IP suite). The packet based CL (IP-based) is divided into two main parts, a *common part* and a *service specific part*, which consists of several Service Specific Convergence Sub-layers. The user plane of the common part is further subdivided into the Common Part Convergence Sublayer and the Segmentation and Reassembly sublayer (Figure 3-1) [ETSI-493.1]. The CL has two main functions, adapting the service request from HLs to the service offered by the DLC layer and secondly to adjust the HL variable size packets to the DLC SDU fixed size.

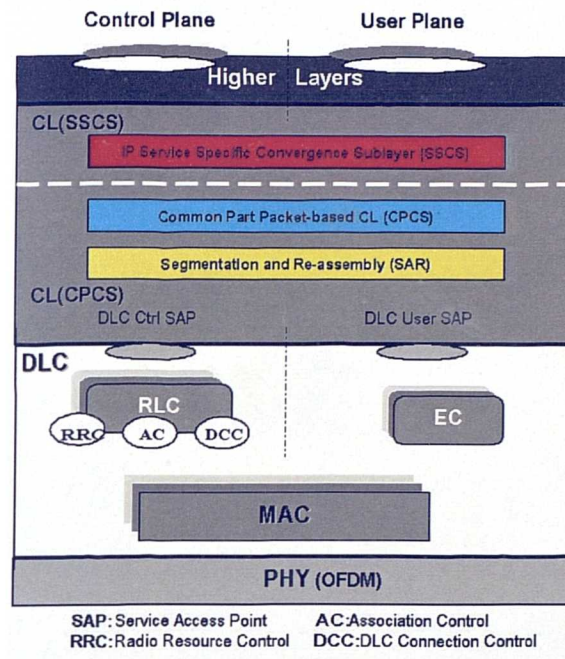


Figure 3-1: HiperLAN/2 protocol stack (AP and MT).

In the H2 cellular system, the AP is the controller for each formed radio cell that covers certain geographical areas. H2 supports transmission in the UL, DL as well as DiL - for ad hoc configuration.

### 3.2.1.1 HiperLAN/2 PHY Layer

The transmission format on the PHY layer is a burst, which consists of a preamble and a data field. OFDM has been selected as the modulation scheme due to its good performance in highly dispersive channels. Table 3-1 lists the main parameters of the standard, where the frequency sampling is 20 MHz and a 64-FFT is typically used. In order to facilitate implementation of filters and to achieve sufficient adjacent channel suppression, only 52 subcarriers are used; 48 carry data and 4 are pilots for phase tracking. This allows coded data rates from 6 to 54 Mb/s using variable modulation types from BPSK to 64-QAM. Convolutional coding is used with the rate 1/2, while this rate may be increased to 9/16 or 3/4 by puncturing the coded output bits. As a result, seven data rates are specified [ETSI-475].

A general block diagram of a H2 transceiver is shown in Figure 3-2. In the transmission chain, binary input data is convolutionally encoded. After interleaving, the binary values are converted into QAM values. To facilitate coherent reception, 4 pilot values are added to each 48 data values, so a total of 52 QAM values is formed per OFDM symbol. Then, these are modulated onto 52 subcarriers by applying the 64-IFFT (Inverse FFT). To make the system robust against multipath propagation, a cyclic prefix is also added.

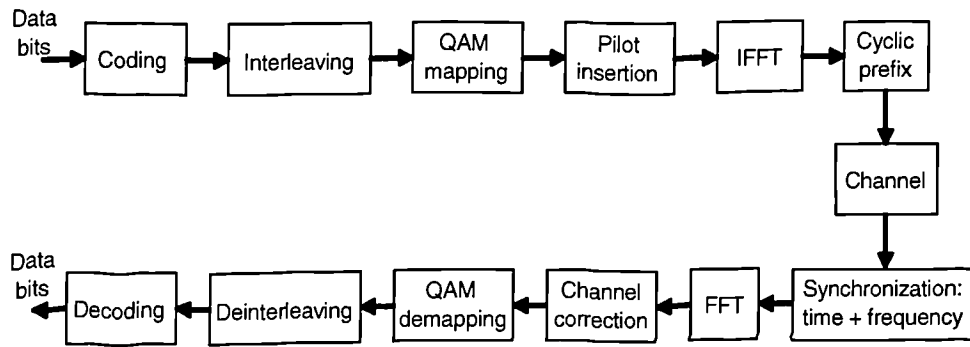


Figure 3-2: Block diagram of a HiperLAN/2 transceiver.

Parameter	Value
Sampling rate	20 MHz
Data rate (Mb/s)	6,9,12,18,24,36,48,54
Modulation	BPSK, QPSK, 16-QAM, 64-QAM
Coding rate	1/2, 9/16, 3/4
No of subcarriers	52
No of pilots	4
OFDM symbol duration	4 $\mu$ s (80 samples)

Table 3-1: OFDM modulation main parameters.

The OFDM receiver basically performs the reverse operations of the transmitter, together with additional training tasks. First, the receiver has to estimate the frequency offset and the symbol timing, using special training symbols in the broadcast preamble. Then, it can apply an FFT to every symbol to recover the 52 QAM values of all subcarriers. The training symbols and pilot subcarriers are used to correct the channel response as well as the remaining phase drift. The QAM values are then de-mapped into binary values, after which a Viterbi decoder is used to decode the information bits.

Independently of the PHY burst type, each burst consists of two sections, preamble and payload. Therefore, each burst is started with a preamble section, which is followed by a payload section.

These preambles are essential to perform packet detection, automatic gain control, time synchronization, frequency estimation, and channel estimation.

Next, the most important of the preambles, which is used for synchronisation in the broadcast burst, is presented<sup>1</sup>. Note that, the term "short OFDM symbol" refers only to its length, which is 16 samples instead of a regular OFDM symbol of 64 samples.

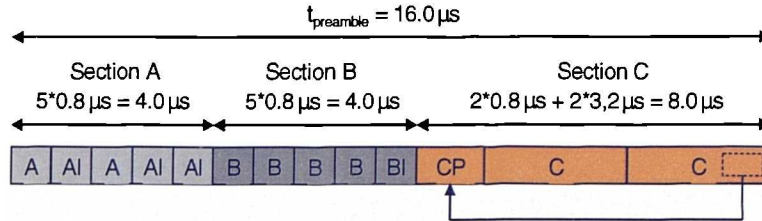


Figure 3-3: Broadcast burst preamble.

The broadcast burst preamble is composed of three sections: Section A, Section B, and Section C. Section A consists of 5 specific short OFDM symbols that are denoted in Figure 3-3 by A and AI. The first 4 short OFDM symbols in Section A (A, AI, A, AI) form a regular OFDM symbol consisting of 12 loaded sub-carriers ( $\pm 2$ ,  $\pm 6$ ,  $\pm 10$ ,  $\pm 14$ ,  $\pm 18$ , and  $\pm 22$ ) given by the frequency-domain sequence SA,

$$SA_{-26 \dots 26} = \sqrt{\frac{13}{6}} * [0, 0, 0, 0, -1 + j, 0, 0, 0, 1 + j, 0, 0, 0, 0, 1 - j, 0, 0, 0, -1 - j, 0, 0, 0, -1 + j, 0, 0, 0, 0, -1 - j, 0, 0, 0, -1 + j, 0, 0, 0, 1 - j, 0, 0, 0, 1 + j, 0, 0, 0, 0] \quad \text{Equation 3-1}$$

The last short symbol in Section A (AI) is a repetition of the preceding 16 time-domain samples. Section B consists of 5 specific short OFDM symbols B and BI. The first 4 short OFDM symbols in Section B (B, B, B, B) form a regular OFDM symbol consisting of 12 loaded subcarriers ( $\pm 4$ ,  $\pm 8$ ,  $\pm 12$ ,  $\pm 16$ ,  $\pm 20$ , and  $\pm 24$ ) given by the frequency-domain sequence SB,

$$SB_{-26 \dots 26} = \sqrt{\frac{13}{6}} * [0, 0, 1 + j, 0, 0, 0, -1 - j, 0, 0, 0, 1 + j, 0, 0, 0, 0, 1 + j, 0, 0, 0, -1 - j, 0, 0, 0, -1 - j, 0, 0, 0, 1 + j, 0, 0, 0, 0, -1 - j, 0, 0, 0, 1 + j, 0, 0, 0, 1 + j, 0, 0, 0, 0] \quad \text{Equation 3-2}$$

The last short symbol in Section B (BI) is a sign-inverted copy of the preceding short symbol B, i.e.  $BI = -B$ .

Section C consists of two OFDM symbols (C) of normal length preceded by a CP of the symbols. All the 52 subcarriers are in use and they are modulated by the elements of the frequency-domain sequence SC given by,

<sup>1</sup> This analysis is given here, due to the importance of the broadcast preamble in the link adaptation and MAC researches [POL-ALM] (Chapter 4).

$$SC_{26..26} = \{1, 1, -1, -1, 1, 1, -1, 1, -1, 1, 1, 1, 1, 1, -1, -1, 1, 1, -1, 1, -1, 1, 1, 1, 0, 1, -1, -1, 1, 1, -1, 1, -1, -1, -1, -1, -1, 1, 1, -1, -1, 1, 1, 1, 1\}$$
Equation 3-3

The cyclic prefix CP is a copy of the last 32 samples of the C symbols and is thus double in length compared to the cyclic prefix of the normal data symbols. Finally, the broadcast burst is created by concatenating the preamble with the data payload.

### 3.2.1.2 HiperLAN/2 MAC Protocol

The H2 DLC layer consists of the MAC function, the EC function and the RLC function. It is divided into the user data transport functions and the control functions. For the H2, a centrally controlled DLC layer is selected, in other words the access point decides how the resources are allocated in a MAC frame.

The MAC protocol is a centrally scheduled TDMA/TDD scheme. Each MAC frame has a fix duration of 2ms and consists of several phases. Important entities of the MAC frame are the logical and transport channels. A logical channel is a generic term for any distinct data path and it is mapped onto different transport channels. The transport channels are the basic elements that construct the PDU trains. Transport channels describe the basic message format. The MAC decomposition to its child entities is bullet-listed here and a visual representation is shown in Figure 3-4 [ETSI-761.1].

- **Broadcast phase:** This phase carries the BCCH and the FCCH, which are mapped onto the BCH and FCH respectively. The BCH transmits control information in each MAC frame and to all MTs. The FCH carries an exact description of how the current MAC resources have been allocated to the UL/DL. This phase also includes the ACH, which provides information on access attempts made by MTs in the RCH of the previous MAC frame.
- **Downlink phase:** It consists of SCH and LCH. During this phase, control information and user data are transmitted from the AP to MTs. The information is assembled in groups of variable length, called cell trains. Each cell train is unique for a correspondent MT.
- **Uplink phase:** In this phase (same structure as DL), control information and user data are transmitted from the MTs to the AP.
- **Random Access phase:** This phase carries the RCHs.

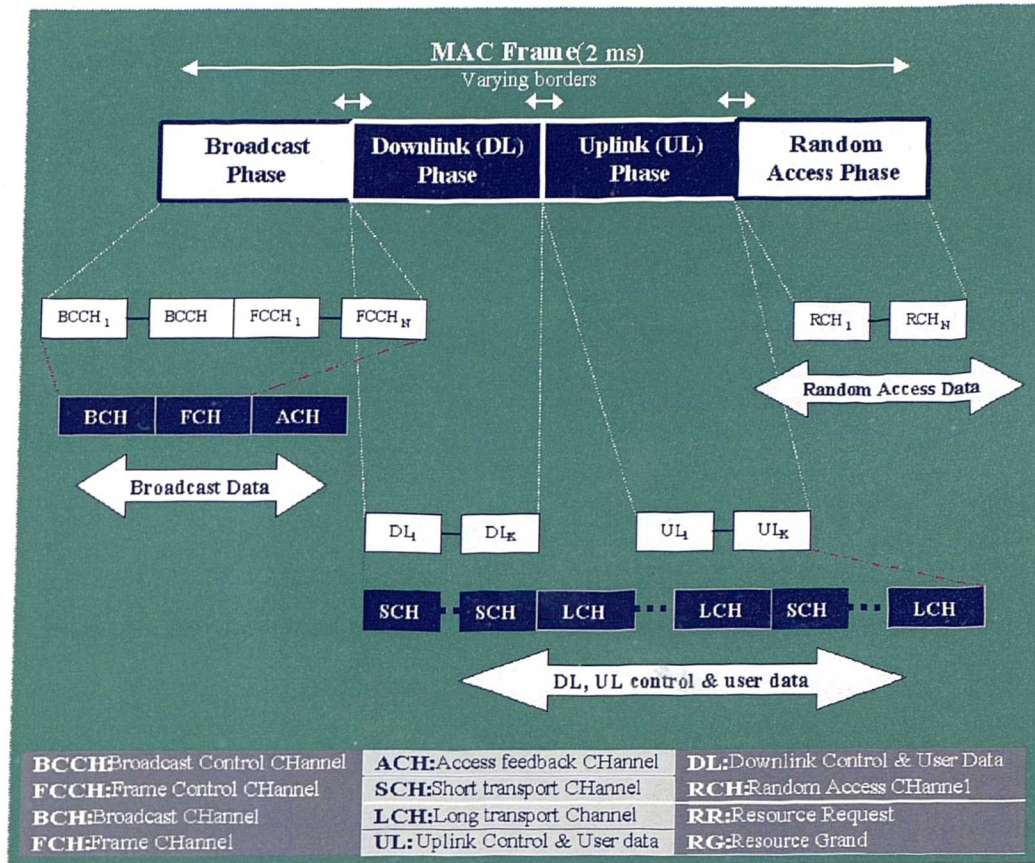


Figure 3-4: MAC frame for HiperLAN/2.

The DL, UL phases consist of two types of DLC PDUs (see Figure 3-4): long PDUs (LCH) and short PDUs (SCH). The LCH PDUs carry *mainly* the payload of the connections. Their size is 54 bytes, whereby 48 bytes are allocated for the payload and the rest (6 bytes) are used for DLC headers. At the same time, the SCH PDUs contain *generally* control data (e.g. ARQ) and their size is 9 bytes. SCH PDUs are also used by MTs to request resources (RR) for a particular connection in the next MAC frame. The RR can either be transmitted during the UL phase in SCH PDUs or in the RCH in competition with other MTs.

The structure of each cell train is cached in the FCH, which serves as a directory for UL and DL phases. The FCH PDU carries a list of *Resource Grant Information Elements*, each defining a cell train of one MT, either UL or DL. The RG-IE is organized in blocks of 8 bytes. The block consists of the IE addressing the particular MT by its MAC-id and determining the location of the cell train within the MAC frame by the start pointer field. The FCH PDUs are segmented in groups of constant length each protected individually by a CRC, in order for the receiver to decode early the relevant information contained in them.

A small introduction to the H2 CL for IP follows.

### 3.2.1.3 Convergence Layer for IP support

The H2 Convergence layer functions as an adapter between DLC and higher layers (e.g. the Internet). Two types of CLs have been specified by ETSI BRAN, a cell based CL and a packet based CL. The cell based CL layer offers services to HL that use the fixed size ATM cell as the transfer unit. Concurrently, the packet based CL offers services to HL that use packets or frames of variable size or fixed size that are larger than the ATM cell size (e.g. Ethernet and/ or the IP suite). This research, however considers mainly the packet based CL.

The CL control plane interacts with the control plane of the DLC layer and the following functions: RRC, ACF, DCC (see also Figure 3-1). The common part of the control plane is transparent to the upper and lower layers [ETSI-493.1]. The mapping of the HL packets (e.g. IP packets) down to the PHY layer bursts is shown in the Figure 3-5.

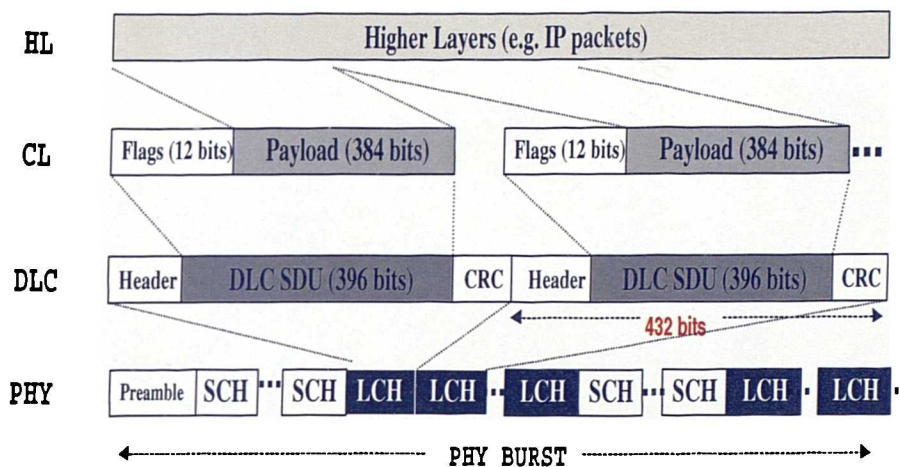


Figure 3-5: Mapping of packets to the H2 layers.

In the IP SSCL (Service Specific Convergence Layer) case, IP packets will be transferred to the CL layer through the CL SAP. At this point, packets will be “stamped” with information about their length, source and destination addresses and their QoS parameters. For that reason, QoS provisioning and address management for H2 DLC/MAC and CL layers are important research topics.

The target of the address management function in the H2 CL is to maintain a mapping between the local link layer addresses and the addresses sent by the IP wireless interface. This can be accomplished by using other wireless interfaces’ addresses (i.e. IEEE802) as static hardware addresses [BON-00].

At the same time, implementing end-to-end QoS requires cooperation at all protocol levels, however in this research the mapping of the IP (network) layer QoS to the DLC QoS is the main objective. The DLC entity of interest is the MAC one. Thus, suitable QoS techniques should be

introduced for the management of the scarce radio resources (bandwidth) available by the air interface for a cellular system such as H2. An effective way of doing that is by enhancing the proposed MAC protocol for H2 by ETSI.

### 3.2.2 H2 Resource Management Analysis

As it was discussed in the previous sections, the H2 MAC protocol has a dynamic nature, since it can support a variety of actual data rates from individual MTs by defining the allocated number of PDUs and their PHY mode. In order to succeed in that, the MAC protocol introduces additional overhead, which affect the air interface resources. A mathematical approach for the resources managed by the H2 MAC protocol is given here [POL-01], which is an enhancement to that suggested in [KAD-99].

(a)↓			
Modulation	Code rate	Capacity of one OFDM symbol (bytes)	Transmission rate (Mb/s)
BPSK	$\frac{1}{2}$	3	6
BPSK	$\frac{3}{4}$	4.5	9
QPSK	$\frac{1}{2}$	6	12
QPSK	$\frac{3}{4}$	9	18
16QAM	$\frac{9}{16}$	13.5	27
16QAM	$\frac{3}{4}$	18	36
64QAM (optional)	$\frac{3}{4}$	27	54
(b)↓			
Transport channel	Direction	PHY mode	Length (octets)
BCH	DL	$\frac{1}{2}$ BPSK	15
FCH	DL	$\frac{1}{2}$ BPSK	Multiple of 27
SCH	DL/UL/DiL	Set in FCCH	9
LCH	DL/UL/DiL	Set in FCCH	54
ACH	DL	$\frac{1}{2}$ BPSK	9
RCH	UL	$\frac{1}{2}$ BPSK	9

Table 3-2: (a) HiperLAN/2 transmission rates (after coding) [ETSI-475], (b) Characteristics of H2 transport channels [ETSI-761.1].

A single H2 radio cell is considered, where a number of MT ( $n_{MT}$ ) are currently active ( $n_{act}$ ) (connected/talking to the AP). The selected PHY mode and the dynamic and static length of the control channels must also be considered. All the specifications are taken from [ETSI-761.1]. The length of the MAC frame phases (BCH, FCH, ACH, DL, UL, RCH) in a number of OFDM symbols is given by:

**BCH:** It has a fixed length of 15 octets and uses binary PSK with code rate  $\frac{1}{2}$  (i.e. 3 octets, as it is shown on Table 3-2) . The BCH comes after the MAC preamble. Thus its length is given,

$$\Lambda_{BCH} = \frac{15}{3} + P_{BCH} = 5 + P_{BCH} \quad \text{Equation 3-4}$$

where  $P_{BCH}$  is the BCH preamble, which we consider to be equal with the MAC preamble,  $P_{MAC}$ .

**FCH:** Its length is variable and is a multiple of 27 octets. The FCH shall be built of fixed size IE blocks (every IE block contains 3 IEs, each with a length of 8 octets), and a CRC of length 24 bits (3 octets). FCH shall be transmitted using binary PSK and coding rate  $\frac{1}{2}$ . The total length for this analysis is given by a RG (Resource Grant) for the BCH and one RG for each active MT ( $n_{act}$ ). Thus the total length of the FCH is,

$$\Lambda_{FCH} = \left[ \left( n_{act} \cdot \frac{8}{24} \right) \cdot \left( \frac{27}{R_{FCH}} \right) \right] = n_{act} \cdot \left( \frac{9}{R_{FCH}} \right) \quad \text{Equation 3-5}$$

where  $R_{FCH}$  is the net data rate on top of PHY layer in Megabits per second (Mb/s), which is determined by the chosen PHY mode.

**ACH:** It has a fixed length of 9 octets and uses binary PSK with code rate  $\frac{1}{2}$ . The total length of the ACH is given by,

$$\Lambda_{ACH} = \frac{9}{3} = 3 \quad \text{Equation 3-6}$$

**DL and UL:** The DL mainly consists of LCH, if the ARQ is neglected. Hence, except the DL preamble, no other overhead is introduced by the downlink phase. However, in the UL one SCH per active MT is introduced (for piggybacking RRs). Finally, a PHY preamble is heading each cell train. Then, the total length of UL overhead is,

$$\Lambda_{UL} = n_{MT} \cdot \left[ P_{PHY} + n_{act} \cdot \frac{9}{R_{SCH}} \right] \quad \text{Equation 3-7}$$

where  $R_{SCH} = R_{FCH}$ .

**RCH:** It has identical format to the SCH (9 octets). It has an extra PHY preamble ( $P_{PHY}$ ) and the number of slots ( $N_{slot/RCH}$ ), it is divided into, determines its size. It uses binary PSK with code rate  $\frac{1}{2}$  (i.e. 3 octets, as it is shown on Table 3-2). The total length of RCH is given by,

$$\Lambda_{RCH} = N_{slot/RCH} \cdot \left( \frac{9}{3} + P_{PHY} \right) = N_{slot/RCH} \cdot (3 + P_{PHY}) \quad \text{Equation 3-8}$$

The MAC frame length is  $t_{MAC} = 2ms$  and  $S_{OFDM}$  is the number of OFDM symbols, which fit into one MAC frame. From Table 3-1, the length of one OFDM symbol is  $t_{symbol} = 4\mu s$  (80 samples) and thus  $S_{OFDM}$  is:

$$S_{OFDM} = \frac{t_{mac}}{t_{symbol}} = \frac{2 \cdot 10^{-3}}{4 \cdot 10^{-6}} = 500 \quad \text{Equation 3-9i}$$

Therefore, 500 OFDM symbols fit into one MAC frame.

Setting  $N_{slo/RCH} = 3$  and  $P_{PHY} = 4$  ( $=P_{BCH} = P_{MAC}$ ) [ETSI-761.1]. Then, the number of free OFDM symbols for LCH is given by Equation 3-9ii and by doing the maths there (substituting Equations 3-4,..., 3-9i in 3-9ii),

$$\begin{aligned} \Lambda_{LCH} &= S_{OFDM} - \Lambda_{BCH} - \Lambda_{FCH} - \Lambda_{ACH} - \Lambda_{UL} - \Lambda_{RCH} = 500 - 9 - \left(\frac{9 \cdot n_{act}}{R_{FCH}}\right) - 3 - n_{MT} \cdot \left(4 + \frac{n_{act} \cdot 9}{R_{SCH}}\right) - 21 = \\ &= 471 - \left(4 + \frac{9 \cdot n_{act}}{R_{SCH}}\right) - n_{MT} \cdot \left(4 + \frac{9 \cdot n_{act}}{R_{SCH}}\right) = 471 - (1 + n_{MT}) \cdot \left(4 + \frac{9 \cdot n_{act}}{R_{SCH}}\right) \end{aligned}$$

$$\text{Equation 3-9ii}$$

The number of symbols per LCH PDU ( $S_{LCH}$ ) is equal to,

$$S_{LCH} = \frac{54}{R_{LCH}} \quad \text{Equation 3-10i}$$

where  $R_{LCH}$  is data rate of the LCH PDUs which also depends on the PHY mode.

Dividing  $\Lambda_{LCH}$  (Equation 3-9ii) by the Equation 3-10i and calculating the number of 48 octets payload fields per MAC frame ( $\rho_{field/MAC}$ ) can find the user data rate on top of the MAC layer ( $R_{MAC}$ ),

$$\begin{aligned} Throughput = R_{MAC} &= \left[ \frac{\Lambda_{LCH}}{S_{LCH}} \right] \cdot \rho_{field/MAC} = \left( \frac{\Lambda_{LCH} \cdot R_{LCH}}{54} \right) \cdot \left( \frac{48 \cdot 8}{t_{MAC}} \right) \Leftrightarrow \\ &\Leftrightarrow R_{MAC} = \left[ 471 - (1 + n_{MT}) \cdot \left( 4 + \frac{9 \cdot n_{act}}{R_{SCH}} \right) \right] \cdot \left( \frac{7.1 \cdot R_{LCH}}{t_{MAC}} \right) \end{aligned} \quad \text{Equation 3-10ii}$$

Equation 3-10ii shows that the MAC layer throughput (this is the global throughput) is closely related with the selected PHY layer mode for the LCH PDU, SCH PDU and CL (S/PDU trains), therefore *QoS provisioning for WLAN systems (and more specifically for H2) is multilayer coordination and cooperation* (refer to Chapter 4 for more details).

### 3.3 Offered Services and Traffic Requirements

In this section, we describe the definition of service and application given by ITU, and then outline typical services and applications that are used throughout this study.

**Service definition:**

- This term represents telecom capabilities that the costumer buys or leases from a service provider. Service is an abstraction of the network-oriented or equipment-oriented view. Identical services can be provided by different network elements, and different services can be provided by the same network elements.
- A set of functions and facilities offered to a user by a provider.

**Application definition:**

- An application is a program that directly interacts with the user. An application utilizes services and might incorporate modules to fulfil its tasks. The application is not restricted to a special environment to run in (cf. services run in server environments).
- The word application is used as the generic term to represent the set of features, combining communication and document processing, on which end users may perform operations. Applications may depend on the working methods and on the allowed processing of documents, co-operative working, etc.
- A set of activities performed to respond to the needs of users in a given situation, for purposes such as business, education, personal communication or entertainment. It implies that software and hardware utilisation could be performed in a fully or partially automatic way and could be accessed locally or remotely. In the last case, it requests the use of telecommunication services.

In the following paragraphs, we briefly address broadband services and/or applications expected to reach SOHO, corporate and residential customers.

### **3.3.1 Services**

Services play a key role in the success of wireless broadband networks. New applications and contents must be deployed in order to make the broadband connection attractive. The core networks such as IP will interlink the diverse access networks for providing ubiquitous multimedia services. Due to the rapid growth of wireless networks, there is a need for QoS extension to services that are available in the wired domain with minimal degradation, to meet end-to-end QoS.

#### **3.3.1.1 Internet Access**

With this application the user can explore/navigate all Internet information using specific programs (browsers). Users can access information with diverse formats (text, video, audio,

pictures, motion video), can download information in order to be consulted afterwards, receive and send e-mail or consult issues accessing to specific searching Internet pages.

Internet access is the most common use of residential and SOHO networks.

### **3.3.1.2 Video-telephony or videoconferencing**

Video-telephony provides real time exchange of audio and video between two or multiple users located in separated sites. This traffic is bi-directional and can also include data information in the form of text or graphics. There are professional equipments for companies that allow multiple users communication with good quality of service, using various ISDN or ADSL lines.

A difference with classical mobile or fixed telephony, is that users must agree on a date and an hour to meet, because there is no way to call another user, as they do not have a call number.

### **3.3.1.3 VoIP telephony**

VoIP telephony is the home version of the professional video telephony. It is an application with increasing use, in form of the so-called messengers. With this application, users from different continents can establish international calls, using VoIP to communicate in the public Internet, at local telephone rates. The quality of service of this type of communication is not guaranteed and usually suffers from jitter and communication drops, but it is cost effective and does not need as large bandwidth as professional video telephony.

### **3.3.1.4 Streaming video and broadcast**

These applications enable the user to have real time access to multiple TV channels, radio and data programming. They use UDP packets in contrary to the usual TCP/IP packets used by the majority of Internet applications, erroneous packets are not retransmitted in order to increase packet speed and reduce protocol overhead. When the user asks for some contents to a streaming video server, the video is sent from the start, but with the broadcast server the user will receive the contents that has transmitted at that time. Broadcast includes applications such as conventional open access TV, pay-per-view and subscription TV.

### **3.3.1.5 Other applications**

The previous list of applications is not comprehensive, there are more applications not included in this study and should be considered for the next generation WLANs, like telemedicine, movies on demand, virtual CD-ROM, urban guidance, tourist information, etc.

## **3.4 Quality of Service parameters**

In the scope of this research, the relevant QoS parameters are the following:

- **Throughput** (global throughput), is the amount of data transferred from one place to another, specifically is the bit rate that the connection can achieve.
- **Latency or Delay**, in networking, is the amount of time it takes for a packet to travel from its source to its destination.
- **Jitter**, is the delay variation experienced by data frames, with respect to the average delay.
- **Frame Packet Loss ratio**, is defined as the percentage of frames, with non-recovered errors, with respect to the total frames received. A frame is a SDU received by a high layer.
- **Bit/Frame Error Rate**, is the percentage of bits/frames, with non-recovered errors, divided by the total number of bits/frames that have been transmitted, received or processed over a given time period.

## **3.5 Broadband WLANs QoS classes**

The QoS mechanisms provided in the radio network have to be robust and capable of providing reasonable QoS resolution, taking into account the air interface characteristics in contradiction to fixed networks.

Following ITU classes of services definition [M-1079] there are four different QoS classes (or traffic classes):

- Conversational class
- Streaming class
- Interactive class
- Background class

The main distinguishing factor between these classes is how delay sensitive they are.

Conversational services, like video telephony, are the most delay sensitive applications and those data streams should be carried with the lowest delay. Streaming class is less sensitive to the delay than conversational class, both must support real-time traffic flows.

Interactive and background classes are used by traditional Internet applications like WWW, email, Telnet, FTP and news. Interactive class is mainly used by interactive applications, e.g. interactive email or interactive web browsing. Background class is meant for background traffic, e.g. background download of emails or background file downloading, the end user is typically a computer, which automatically receives information from a remote server.

### **3.5.1 Conversational class**

The most well known use of this scheme is telephony speech. But with Internet and multimedia, a number of new applications will require this scheme, for example:

- Voice telephony
- Voice over IP
- Video conference
- Other delay sensitive applications (interactive games)

With this scheme, the required characteristics are strictly given by human perception. Therefore, the limit for acceptable transfer delay is very strict, as failure to provide low enough transfer delay will result in unacceptable lack of quality. The transfer delay requirement is therefore both significantly lower and more stringent than the round trip delay of the streaming traffic case. Furthermore, it cannot use large data buffers to compensate data rate variations, as for streaming, due to the conversation net delay increase that it would impose.

Real time conversation - fundamental characteristics for QoS:

- Preserve time relation (variation) between information entities of the stream
- Conversational pattern (stringent and low delay)

### **3.5.2 Streaming class**

When the user is looking at (listening to) real time video (audio) the scheme of real time streams applies. The real time data flow is always aiming at a living (human) destination. Initially data buffers must be refilled; this supposes an initial delay in the information to play. The use of data buffers can compensate some amount of jitter or even temporary network data rate drops.

us scheme is characterised by the fact that the delay variation (jitter) between information entities within a flow must be limited, although it does not have any requirements on low transfer delay.

Real time streams - fundamental characteristics for QoS:

- Preserve delay variation/jitter between information entities of the stream
- Low packet delay (Tolerant packet loss)

### 3.5.3 Interactive class

This scheme applies when the end-user, that is either a machine or a human, is on-line requesting data (either real time video or audio) from remote equipment (e.g. a server).

Examples of human interaction with the remote equipment are:

- Web browsing,
- Data base retrieval,
- Servers access,
- Video on demand.

Examples of machines interaction with remote equipment are:

- Polling for measurement records,
- Automatic data base enquiries (tele-machines).

Interactive traffic is the other classical data communication scheme that on an overall level is characterised by the request response pattern of the end-user. At the message destination there is an entity expecting the response to a *information request, within a certain time*. Round trip delay time is therefore one of the key attributes. Another characteristic is that the content of the packets must be transparently transferred (with low bit error rate).

Interactive traffic - fundamental characteristics for QoS:

- Low request-response time (low round trip delay time).
- Payload content without errors (low FER).

### 3.5.4 Background class

In this scheme the end-user, that typically is a computer, sends / receives data-files automatically in the background. This scheme is the least delivery time sensitive. Some examples are:

- Background delivery of e-mails.
- Download of databases.
- Reception of measurement records.

Background traffic - fundamental characteristics for QoS:

- Delivery time insensitive (can accept high delays).
- Payload contents without errors (low FER).

### 3.5.5 QoS requirements

For the most critical services, delay/latency, upper and lower limits of throughput (UL and DL), jitter, BER (after error correction) and FER (at application level) are shown [WIND-99], [STR-2.1] in the following Tables.

Application	Delay/Latency	UL BW	DL BW	Jitter	BER	FER
Interactive games	50 ms	> 80 kb/s	512 kb/s to 6 Mb/s	5 ms	$10^{-5}$	0
Service type	CO (Conversational)					

Table 3-3: Interactive Games.

Application	Delay/Latency	UL BW	DL BW	Jitter	BER	FER
Video conference H.261/H.263	Preferred < 150 ms limit < 400 ms	64 kb/s to 2 Mb/s	64 kb/s to 2 Mb/s	1 ms	$4 \cdot 10^{-11}$ to $10^{-6}$	<1%
Service type	CO (Conversational)					

Table 3-4: Videoconference.

Application	Delay/Latency	UL BW	DL BW	Jitter	BER	FER
Fast Internet Access	Preferred 100 ms Acceptable 1s	6-64 kb/s	64-512 kb/s	n.a.	$10^{-8}$ (data) $10^{-6}$ (audio), $10^{-9}$ - $10^{-4}$ (video)	0 (data) <1% (audio, video)
		64-512 kb/s	512 kb/s - 2 Mb/s			
		512 kb/s – 2 Mb/s	2 Mb/s – 8 Mb/s			
Service type	IN (Interactive)					

Table 3-5: Broadband Internet Access.

Application	Delay/Latency	UL BW	DL BW	Jitter	BER	FER
High quality streaming	delay play-start 1s	16 kb/s to 64 kb/s (pay TV)	<10 Mb/s	0.5 ms	$10^{-11}$	<1%
Service type	ST (Streaming)					

Table 3-6: High quality streaming.

## 5.6 Summary of service requirements

This section presents a summary of significant service classes to be supported by the WLAN systems. From literature and other sources (e.g. EU projects), an end-to-end summary of service classes and parameters values is presented in Table 3-8. Service classes are defined as conversational (CO), streaming (ST), interactive (IN), and background (BG). For each type of service the most stringent application is selected with its associated QoS parameters. The QoS parameters are intended to be end-to-end. This classification is more comprehensive than the ETSI BRAN classification, since it includes UL and DL data rate, BER and FER.

Service	Max Delay (ms) including jitter, for at least 95% of the packets	Max Jitter (ms) for at least 95% of the packets	Packet Error Ratio	Data rates
Voice	20	10	3 %	< 64 kb/s
Time critical data	40	20	0.2%	< 400 kb/s
Streaming	100	50	1 %	< 1 Mb/s MPEG4 < 4 Mb/s for full screen video Video conferencing with H.263 < 100 kb/s
Non time critical data	n.a.	n.a.	0.1%	n.a.

Note: The QoS parameters are specified between two CLs on a single radio link

Table 3-7: HiperLAN/2 services QoS requirements.

	Application	Delay /Latency	Jitter	FER	BER	UL BW	DL BW
CO	Network Games	50 ms	5 ms	0	$10^{-5}$	80 kb/s	6 Mb/s
IN	Internet Access	100 ms	10 ms	0	$10^{-8}$	2 Mb/s	8 Mb/s
ST	High quality streaming	1000 ms	0.5 ms	<1%	$10^{-11}$	64 kb/s	10 Mb/s
BG	e-mail	5000 ms	n.a.	0	$10^{-6}$	64 kb/s	64 kb/s

Table 3-8: End-to-end services QoS requirements.

## 3.6 Traffic Models

### 3.6.1 Internet (UMTS 30.03 model)

The Internet represents the largest existing WAN offering a broad spectrum of information resources in the form of technical reports, product information, software, images, audio and video sources and electronic commerce services. The popularity among larger user groups to use these

sources was driven by the introduction of the WWW. In the WWW, global naming conventions, protocols and object formats are used. An object can be a text file, an audio source, an image file or even the code of a program that can be interpreted and executed on the user terminal (Java applet).

World Wide Web browsing packet data traffic is characterized to appear during so-called browsing sessions as illustrated in Figure 3-6. Session arrivals are modelled by a Poisson process, similar to the arrivals of speech/voice calls. During a session, the user will initiate several packet call requests. The number of packet calls within a session is a geometrically distributed random variable  $N_{pc}$  with an average value of  $\bar{N}_{pc} = 5$  [STR-2.1]. This means, an average of five packet calls will appear within one session.

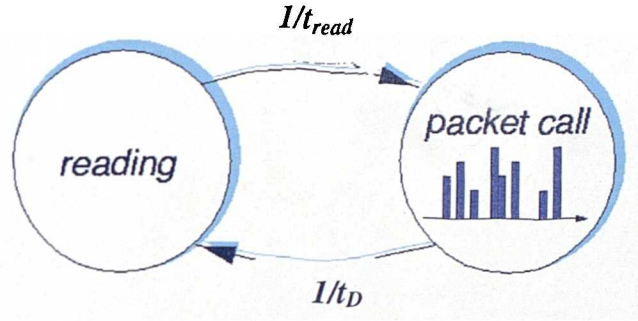


Figure 3-6: Traffic model for packet data services.

The reading time between consecutive packet calls  $t_{read}$  is a geometrically distributed random variable with mean  $\bar{t}_{read} = 12s$ . After this pause, a new packet call will be initiated. The reading time starts after the session is completely received and hence, the exact value of  $t_{read}$  will be generated after the last transmitted packet is received and triggers the next packet call. Note that the mean duration of the session is  $t_D = 48s$ .

The average requested file size is  $\bar{S}_{pc} = 12\text{kbyte}$ , spread over several packets with a mean value of 480 byte data. The packets represent elements of web pages requested by the user. However, this model completely neglects signalling and acknowledgment traffic, while needs a more detailed implementation for future investigations of packet data traffic.

The parameters for web browsing packet data services are summarized in Table 3-9.

Parameter	WWW
$\bar{N}_{pc}$	5
$\bar{t}_{read} [s]$	12
$t_D [s]$	48

$\bar{S}_{pc}$ [kbit]	96
-----------------------	----

Table 3-9: Parameters for traffic models.

## 3.2 Internet (Choi's model)

WWW browser displays WWW pages that are composed of one or more WWW objects. Figure 7 shows a WWW page displayed by a WWW browser. It contains several different objects such as text and images. Within one page *links* can be defined that are connected to an identifier of a WWW resource, a URL. When the user chooses a link, a new object or page is displayed. By the direct specification of a URL in the text field of the WWW browser, a specific page or object can be chosen by the user (see Figure 3-7). Every object of the WWW has a unique URL.

The presentation of a WWW page is described by the HTML. In an HTML file the appearance and the URL of the contained objects are specified. An HTML page can also contain text that is displayed at the related position of the page. An HTML page itself is also a WWW object that is analysed by the browser. Further WWW objects are specified in the HTML file as part of the page that are requested and transmitted.

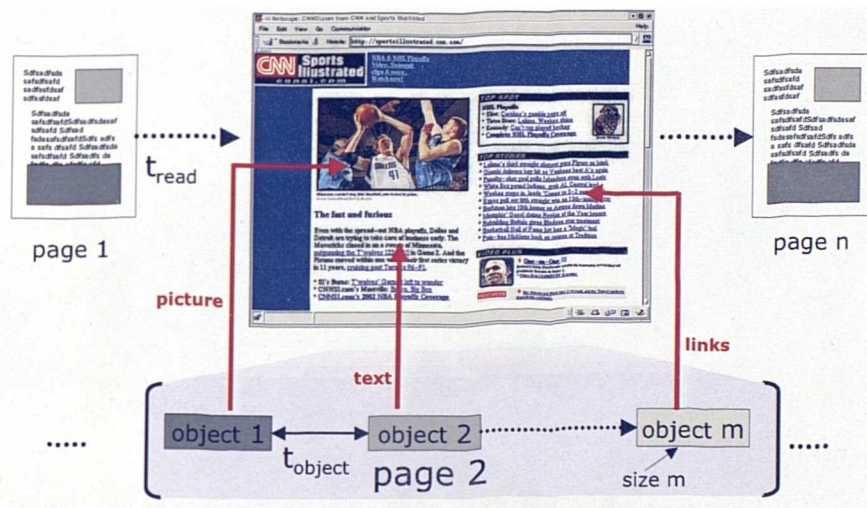


Figure 3-7: Structure of a WWW page [STR-2.1].

The WWW is characterized by client-server architecture. A WWW browser represents a client that poses requests to a WWW server concerning the transmission of WWW objects that are stored on the server. The communication between client and server is controlled by HTTP. HTTP is an application-oriented protocol that uses the TCP for end-to-end transport. In the initial version of HTTP a new TCP connection is set-up for each requested object. The client transmits a request for a certain object to the server. The request contains the URL of the object, data format parameters and parameters for access control. The server processes this request and transmits the

requested object to the client. Then the TCP connection is released. In newer HTTP versions, e.g., HTTP version 1.1, TCP connections can be reused by the following objects (persistent TCP connections). Several TCP connections can be set-up in parallel and are used for the pipelined transmission of several different objects (pipelined TCP connections). Figure 3-8 depicts the sequence of a WWW session.

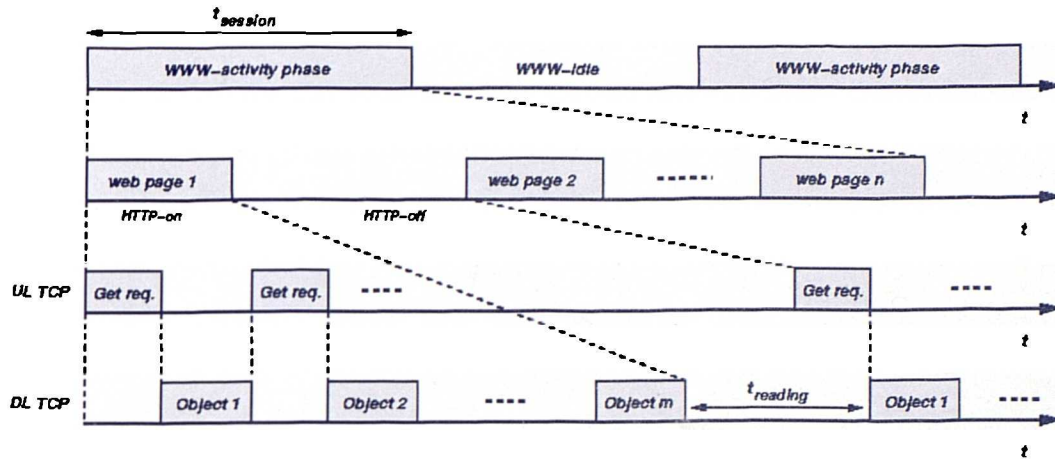


Figure 3-8: Sequence and parameters for a WWW session.

Choi's behavioural model is representative for WWW browsing performed on 2 PC connected to the fixed Internet. As mentioned above WWW sessions consist of requests for a number of pages. These pages consist of objects, each with a certain object size. Another characteristic parameter is the delay between two pages depending on the user's behaviour to surf the Web [CHO-99]. Table 3-10 gives an overview of the Mosaic WWW traffic parameters.

Choi's traffic model is an on/off-model with alternating phases of packet generation and silence (see Figure 3-9) [CHO-99]. An on-phase starts after the arrival and acceptance of a web request. During this phase, the objects of a WWW page are requested. The off-phase represents a silence period after all objects have been retrieved. Thus, the on and off-phases equal the page loading times and page viewing times, respectively. During the on-phase the page's objects are downloaded.

WWW Parameter	Distribution	Mean	Variance
Viewing time [s]	Weibull	39.5	$8.57 \cdot 10^3$
No. of inline objects per page	Gamma	5.55	130.0
Size of main object [kB]	Log-Normal	10	625.0
Size of inline objects [kB]	Log-Normal	7.7	$1.59 \cdot 10^4$

Table 3-10: Choi WWW traffic model parameters [CHO-99].

We distinguish two types of objects: the main object containing the document's HTML code and inline objects, such as linked objects, images or Java applets. To fetch all those inline objects modern browsers open several TCP connections in parallel after the successful retrieval of the

main object. In Table 3-10 the random variables used by Choi to describe the object sizes, the number of inline objects and the length of the viewing time are listed.

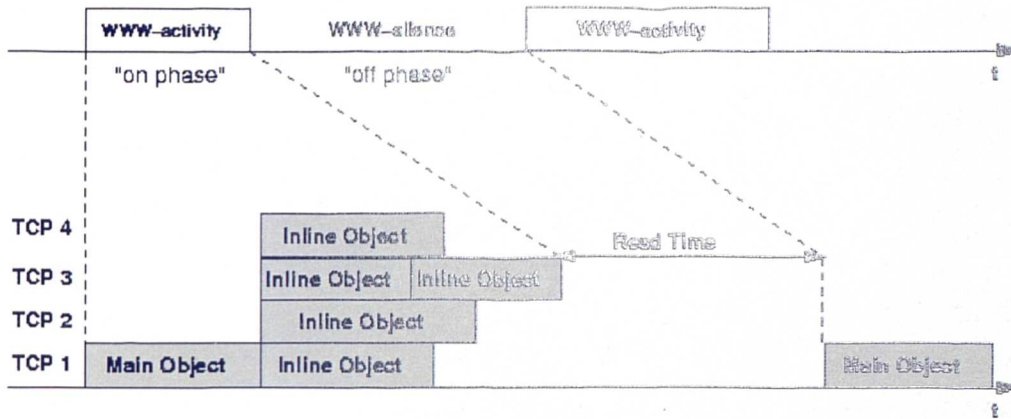


Figure 3-9: Choi's Behavioural Model.

The amount of data  $S$  that is generated during the on-phase is given by

$$S = s_{\text{Main}} + \sum_{i=0}^{n_{\text{Inline}}} s_{\text{Inline},i} \quad \text{Equation 3-11}$$

which is a combination of three random variables. The random variable  $S$  represents the payload to be transmitted by the HTTP protocol.

### 3.6.3 Voice [STR-2.1]

A speech connection is characterised by an alternation between active and silent periods. An event-driven state model is applied to define these random periods (see Figure 3-10). Both, active and silent times are negative exponentially distributed random numbers. The mean active and silent times are  $T_{\text{active}} + T_{\text{silent}} = 3\text{s}$ .  $T_{\text{active}}$  defines the mean time the model is in the active state and transmitting on the uplink.  $T_{\text{silent}}$  is the mean time the model is in the silent state and receiving on downlink, respectively. This results in a channel activity of 50%. The stochastic generators of all mobiles work independently from each other.

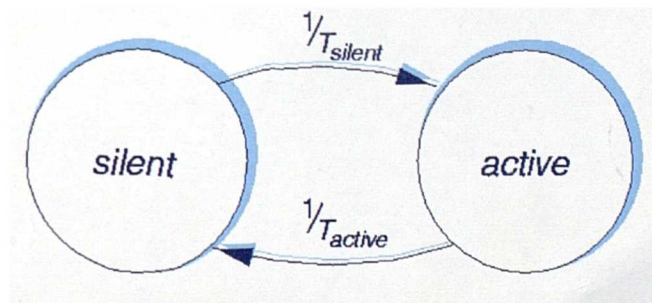


Figure 3-10: Traffic model for speech service.

The mean call duration for a speech service is  $t_{D,Speech} = 120s$ . The call duration is negative exponentially distributed. For the call duration time, the resources for the connection should be allocated exclusively.

### 3.6.4 VoIP Model

The transfer of voice traffic over packet networks, and especially VoIP, is rapidly gaining acceptance. While many corporations have been using voice over CS networks to save money, the dominance of IP has increasingly shifted the attention from CS to PS VoIP. VoIP transfer can significantly reduce the per-minute cost, resulting in reduced long-distance bills. In fact, many dial-around-calling schemes available today already rely on VoIP backbones to transfer voice, passing some of the cost savings to the customer.

But along with initial excitement, customers are worried over possible degradation in voice quality when voice is carried over these packet networks. Whether these concerns are based on experience with the early Internet telephony applications, or whether they are based on understanding the nature of packet networks, voice quality is a critical parameter in acceptance of VoIP services. Therefore, it is crucial to understand the factors affecting VoIP transmission, as well as to obtain the tools to measure and optimise them.

#### 3.6.4.1 Audio Codecs

Voice channels occupy 64 kb/s using PCM coding when carried over CS links. Compression techniques (Codecs) were developed allowing a reduction in the required bandwidth, while preserving voice quality.

#### 3.6.4.2 Codec (G.723)

The G.723 [ITU-723] is used for speech transmission in videoconferencing systems and is used in the videoconference standard H.323 [ITU-323] and SIP MGCP (IETF MGCP/MEGACO protocol). The coder has two possible bit rates:

- 5.3 kb/s
- 6.3 kb/s

It encodes speech or other audio signals in frames using linear predictive analysis by synthesis coding. The excitation signal for the high rate coder is *Multi-Pulse Maximum Likelihood Quantization* (MP-MLQ) and for the low rate coder *Algebraic Codebook Excited Linear Predictive* (ACELP). The recommendation foresees a frame size of 30 ms in which the speech or

other audio signals are encoded. Both rates are a mandatory part of the encoder and decoder. It is possible to switch between the two rates at any 30 ms frame boundary. An option for variable rate operation using discontinuous transmission and noise fill during non-speech intervals is also possible. In addition, there is a look ahead of 7.5 ms, which gives a coding delay of 37.5 ms. All additional delays in the implementation and operation of this coder are due to:

- Actual time spent processing the data in the encoder and decoder;
- Transmission time on the communication link;
- Additional buffering delay for the multiplexing protocol.

The output of speech coder during periods is depicted in Figure 3-11 generating speech packets of 24 bytes at high rate and 20 bytes when transmitting at low rate. Silence compression techniques like *Voice Activity Detection* (VAD) and *Comfort Noise Generator* (CNG) reduce transmitted data volume during silence intervals of speech. The purpose of the CNG algorithm is to create a noise that matches the actual background noise with a global transmission cost as low as possible. At the transmitting end, the CNG algorithm uses the activity information given by the VAD for each frame, and then computes the encoded parameters needed to synthesise the artificial noise at the receiving end. These encoded parameters compose the *Silence Insertion Descriptor* (SID) frames, which require fewer bits than the active speech frames and are transmitted during inactive periods.

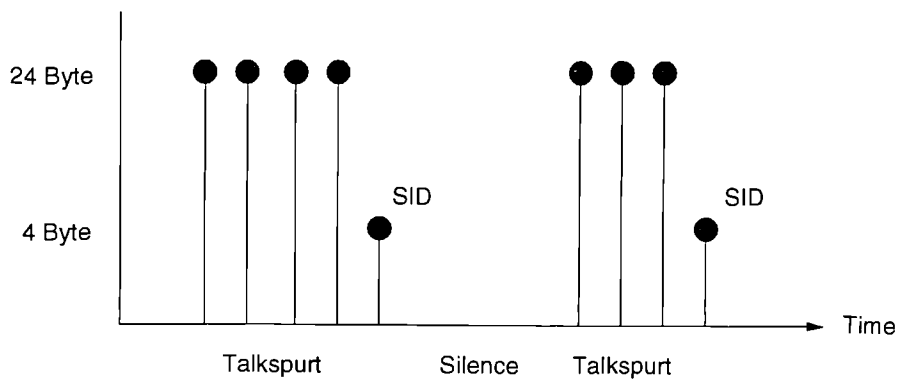


Figure 3-11: Output of Speech Coder G.723 - High Rate (6.3 kb/s).

### 3.6.4.3 Performance Issues

Delay in telecommunication systems is incurred by two activities - echo and talker overlap. Echo is caused by the signal reflections of the speaker's voice from far end telephone equipment back into the speaker's ear. Echo becomes a significant problem when the round trip delay becomes greater than 50 ms [STR-2.1]. Since echo is perceived as a significant quality problem, Voice over Packet systems must address the need for echo control and implement some means of echo cancellation.

Talker overlap (or the problem of one talker stepping on the other talker's speech) becomes significant, if the one-way delay becomes greater than 250 ms [STR-2.1]. The end-to-end delay budget is therefore the major constraint and driving requirement for reducing delay through a packet network.

**Accumulation Delay:** This delay (sometimes called algorithmic delay) is caused by the need to collect a frame of voice samples to be processed by the voice coder. It is related to the type of voice coder used and varies from a single sample time (0.125 ms) to many milliseconds, e.g., 30 ms for the G.723 codec.

**Processing Delay:** This delay is caused by the actual process of encoding and collecting the encoded samples into a packet for transmission over the network. The encoding delay is a function of both the processor execution time and the type of algorithm used. Often, multiple voice coder frames will be collected in a single packet to reduce the packet overhead.

**Network Delay:** This delay is caused by the physical medium and protocols used to transmit the voice data, and by the buffers used to remove packet jitter on the receive side. Network delay is a function of the capacity of the links in the network and the processing that occurs as the packet transit the network. The jitter buffers add delay, which is used to remove the packet delay variation that each packet is subjected to as it transits the packet network. This delay can be a significant part of the overall delay since packet delay variations can be as high as 70-100 ms in IP networks [STR-2.1].

The delay problem is compounded by the need to remove jitter; a variable inter-packet timing caused by the network a packet traverses. Removing jitter requires collecting packets and holding them a long time, enough to allow the slowest packets to arrive in time to be played in the correct sequence. This causes additional delay.

The two conflicting goals of minimizing delay and removing jitter have engendered various schemes to adapt the jitter buffer size to match the time varying requirements of network jitter removal. This adaptation has the explicit goal of minimising the size and delay of the jitter buffer, while at the same time preventing buffer underflow caused by jitter.

#### 3.6.4.4 Voice Traffic Characteristics

Voice activities can be considered as alternating between two states: talkspurt and silent. Data is generated during talkspurt only, and no data is transmitted during silence, thereby making statistical multiplex gain possible. Paul T. Brady discovered that both talkspurt and silence periods of digitised voice are exponentially distributed [BRA-69].

**Single Speaker Model:** The commonly accepted model for a speaker in a voice call is a continuous-time, discrete-state Markov chain. The holding time in each state is exponentially distributed with mean  $1/\lambda$  and  $1/\mu$ , respectively (see Figure 3-12).

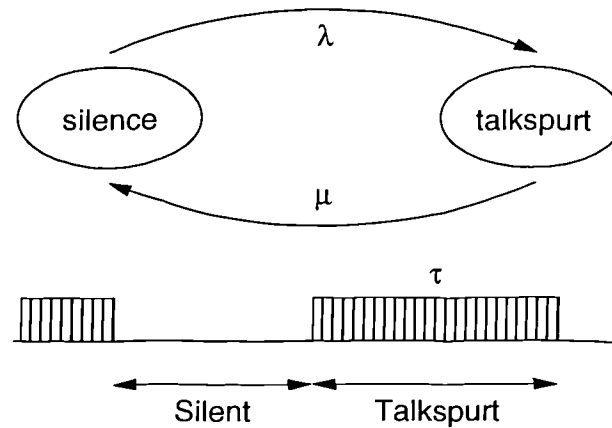


Figure 3-12: Single Speaker Model.

The commonly used values are  $1/\lambda = 650$  ms and  $1/\mu = 352$  ms. These values are often referred to as the *Traditional Packet Voice Model*. Values of  $1/\lambda = 742$  ms and  $1/\mu = 435$  ms were measured from actual telephone traffic with silence detector of high sensitivity [YAT-82].

The fundamental basis of all packet voice performance studies for the past is the classic model of exponential distributions by Brady. This model, however, was based on digitised voice statistics. Its applicability to packet voice with different audio content has not been examined before.

Packetisation may change the traffic characteristics of packet voice, since both the techniques and implementations are different. In addition, while digitised voice in the traditional model was mainly concerned with telephone conversations, packet voice will be used for a wider range of new applications that may affect the traffic model too. For instance, a significant amount of applications will be for information or entertainment distribution.

**Packet Voice Traffic Model:** The proposed traffic model for VoIP is based on the Single Speaker Model (see Table 3-11) and it was used in our simulations.

During talkspurt periods, the codec generates a CBR of 24 bytes every 30 ms (based on the high data rate of the G.723, Section 3.6.4.2), respectively 20 byte at low rate. Directly after the transition to silent state a 4 bytes frame for the CNG is sent.

	Single-Speaker model for digitised speech	Traditional Packet Voice	Packet Voice
Talkspurt	1.34s	0.352s	7.24s
Silence	1.67s	0.65s	5.69s

Table 3-11: Parameters for Voice modelling with Exponential Distribution.

### 3.6.5 A Study on Video Streaming (and VoD)

Video streaming (and Video on Demand) applications are usually based on MPEG standard (MPEG-2). MPEG video is broken up into a hierarchy of layers to help with error handling, random search and editing, and synchronization. From the top level, the first layer is known as the video sequence layer, and is a self-contained bit stream, for example a coded movie or advertisement. The second layer down is the *Group of Pictures* (GOP), which is composed of one or more groups of intra (I) frames and/or non-intra (P and/or B) pictures that will be defined later. Of course the third layer down is the picture layer itself, and the next layer beneath it is called the slice layer. Each slice is a contiguous sequence of raster ordered macro-blocks, most often on a row basis in typical video applications, but not limited to this by the specification. Each slice consists of macro-blocks, which are 16 x 16 arrays of luminance pixels, or picture data elements, with two 8 x 8 arrays of associated chrominance pixels.

#### *Intra Frame Coding Techniques*

The term intra coding refers to the fact that the various loss less and lossy compression techniques are performed relative to the information that is contained only within the current frame, and not relative to any other frame in the video sequence. In other words, no temporal processing is performed outside of the current picture or frame. It turns out that the coding is very similar to that of a JPEG still image video encoder, with only slight implementation detail differences. The basic processing parts are the video filter, *Discrete Cosine Transform* (DCT), DCT coefficient quantiser, and run-length amplitude *Variable Length Coder* (VLC).

#### *Predicted Frame Coding*

Starting with an intra, or I frame, the encoder can forward predict a future frame. This is commonly referred to as a P frame, and it may also be predicted from other P frames, although only in a forward time manner. As an example, consider a GOP that lasts for 6 frames. In this case, the frame ordering is given by I P P P P P I P P P P... . Each P frame in this sequence is the difference between the prediction and the real picture information.

#### *Bi-directional Frame Coding*

The encoder also has the option of using forward/backward interpolated prediction. These frames are commonly referred to as bi-directional interpolated prediction frames, or B frames for short. As an example of the usage of I, P, and B frames, consider a GOP that lasts for 6~frames, and is given as I B P B P B I B P B P B... . As in the previous example with altering I frames and P frames only, I frames are coded spatially only and the P frames are forward predicted based on previous I and P frames. The B frames however, are coded based on a forward prediction from a previous I or P frame, as well as a backward prediction from a succeeding I or P frame. As such,

the example sequence is processed by the encoder such that the first B frame is predicted from the first I frame and first P frame, the second B frame is predicted from the second and third P frames, and the third B frame is predicted from the third P frame and the first I frame of the next GOP. From this example, it can be seen that backward prediction requires that the future frames that are to be used for backward prediction be encoded and transmitted first, out of their usual order. Most broadcast quality applications however, have tended to use 2 consecutive B frames (I B B P B B P...) as the ideal trade-off between compression efficiency and video quality.

The main advantage of the usage of B frames is the coding efficiency. In most cases, B frames will result in less bits being coded overall. Quality can also be improved in the case of moving objects that reveal hidden areas within a video sequence. Backward prediction in this case allows the encoder to make more intelligent decisions on how to encode the video within these areas. Moreover, since B frames are not used to predict future frames, errors generated will not be propagated further within the sequence.

### Group of Pictures

I are the frames with the most information in them and thus, they have the largest frame size. The P frames contain less information and have an average size of about a quarter of the I frame. This compression is achieved by the motion compensation. The lowest frame sizes are the B frames, which have an average size of about half the P frames. It can be seen that a considerable compression can be achieved by not just sending all I frames, but instead sending a mix of the different frame types. This mix of frames is called the *Group of Pictures* (GOP), which is usually periodic between two I frames.

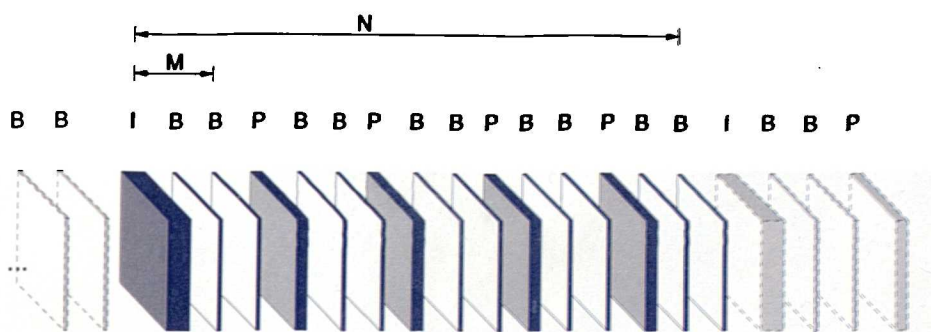


Figure 3-13: Encoded (15, 3) Group of Pictures, hatched frames belong to other Groups of Pictures.

The sequence is decided in advance and is therefore deterministic. Quite often the fixed GOP pattern is regular in the sense that the number of B frames between two reference frames (I or P frame) is fixed. Such a GOP pattern can be characterized by two parameters: the I-to-I frame distance  $N$  (GOP length) and the I-to-P frame distance  $M$ . If no B frames are used (e.g. in case of video conferencing), then  $M = 1$  applies. A typical (15, 3) GOP with  $N = 15$  and  $M = 3$  might be

(I B B P B B P B B P B B P B B) and then repeats again which is shown in Figure 3-13. Since most of MPEG sequences are coded with regular GOP patterns (often to simplify the codec design), we limit our treatment to regular GOP.

GOP formats	Sequence
(15, 3)	I B B P B B P B B P B B P B B
(15, 1)	I P P P P P P P P P P P P P P P
(12, 3)	I B B P B B P B B P B B
(12, 4)	I B B B P B B B P B B B
(10, 2)	I B B P B B P B B P

Table 3-12: Usual GOP patterns.

#### *Stochastic Model*

Based on statistical surveys of Krunz [KRU-95], the statistical characteristics of a scene-based video streaming traffic generator [KRU-97] are introduced in this section.

Compression schemes are employed in order to reduce the amount of bits contained in video streams. The size of a compressed video frame varies depending on the scene activity and the type of compression involved. Hence, the output of a video codec is a *Variable Bit Rate* stream. Several traffic models have been proposed to characterize compressed video streams [LOM-98]. The parameters of these models were obtained by matching certain statistical characteristics of an actual video sequence and the model under consideration. Particular emphasis was on matching the mean, variance, and more importantly the correlation structure of the bit-per-frame sequence. Correlations arise naturally in compressed video traffic and their patterns are largely determined by the type of video, as well as the compression algorithm used. This work deals with statistics of long video sequences (up to several hours) to study the statistical characteristics of MPEG-I traffic.

The distribution of the frame size is very close to a Pearson type V density function [KRU-97]. It is not purposive to develop a model based on a distribution for all frames, since the type of each frame would not be captured by such a model. Hence, it is more appropriate to study the frame size distribution for each frame type.

Frame type	Mean [kbit]	Std. [kbit]
I frame	197.1	63.8
P frame	58.0	37.3
B frame	19.6	5.7

Table 3 13 Statistics (MPEG 640 x 480 pixel)

For each of the frame types, I, P and B frames, a lognormal probability density function is found to provide the best fit for the frame size histogram. In Table 3-13, typical statistical values for I, P and B frame types for modelling MPEG-I (640 x 480 pixels) are shown.

### 3.6.6 Streaming audio [MENA-00]

Another typical application for IP-based WLANs could be the streaming audio. In this study, we have considered the model proposed by MENA et al. The two traffic sources of the model are characterised at the application level by using the ON-OFF models.

- ON period: Within a session the data blocks are emitted based on an ON/OFF scheme. The duration of the ON period is exponential with parameter  $\lambda = 0.2s$ .
- OFF period: The distribution of the OFF duration is deterministic with mean  $m = 1.8s$ .

Furthermore, this application is asymmetric with few numbers of packet requests in the uplink direction. While, for the downlink the following parameters should be considered:

- Session duration: Pareto distribution with shape parameter  $\alpha = 1.6$  and mean  $m = 40 * 60 = 2.4 (s)$ ;
- Session inter-arrival time: Exponential distribution with parameter  $\lambda = \frac{3.6}{0.16 * K}$ , where  $K$  is the average number of audio connections performed by a user;
- Session bit rate is 20 kb/s (in the case with audio music traffic);
- Packet length is 500 bytes;

## 3.7 Summary

In this chapter various issues related to the IP-based WLAN (HiperLAN/2) architecture and protocol stack have been presented and some considerations and potential solutions have been discussed. Furthermore, an analytical approach was followed to identify the interworking requirements between various H2 layers (PHY-DLC/MAC-CL). The IP services and traffic related sections offered an insight into the expected applications an All-IP WLAN could/will support. Finally, the selected traffic models for each service class, which were used in our simulation studies, were presented and wherever possible mathematically analysed.

# Chapter 4

## 4 QoS Framework for All-IP WLAN

### 4.1 Introduction

The motivation for the work described in this chapter was the anticipated internetworking between the Internet (IP-based core) and the new WLAN specification based on the ETSI HiperLAN/2 standard (both physical and MAC layers). By studying this wireless protocol its advantages and disadvantages can be related to the IP multimedia traffic characteristics and QoS, and help us create a more detailed set of requirements (than the one presented in Chapter 3) from a MAC protocol, designed specifically for the transmission of HiperLAN/2 cells. The chapter starts with a detailed analysis of the possible enhancements to ETSI HiperLAN/2 PHY and MAC protocols, in order to support IP traffic both on the control as well as the data planes. A complete QoS framework and the interworking and interoperability of various layers of the WLAN system are presented. On the PHY level, a link adaptation technique is evaluated in terms of PER and throughput. While, on the MAC level various classifying, queuing and scheduling algorithms are simulated and evaluated. A connection admission control mechanism is also proposed; all simulation models were developed in OPNET. Finally, the chapter concludes with a discussion on the proposed multilayer and multi-protocol QoS framework for the next generation IP-based WLAN.

### 4.2 Enhancements to HiperLAN/2 PHY for QoS Support

*In this section, novel and effective techniques, which improve the performance of the IP-based HiperLAN/2 networks, are introduced. This is what we call, “QoS support for WLAN PHY”. In order to achieve a high throughput over fading/dispersive channels, the deployment of techniques such as link adaptation and adaptive methods for adjustment of the modulation alphabet and the coding complexity are used. While, their performance is evaluated by simulation means [POL-ALM].*

### 4.2.1 Link Adaptation Technique

Link Adaptation techniques help the maximization of spectral efficiency and data throughput per subscriber under given channel conditions. In case of poor link quality the modulation and coding scheme chosen for the transmission of PDUs can be adapted to a more robust one. However, PDUs coded with a more robust PHY mode require more capacity in the MAC frame.

Despite the ability of H2 to select the appropriate modulation level via LA, this level is identical for all subcarriers. However, an OFDM system is able to switch modulation schemes individually for each subcarrier. This is performed by allocating bits to subcarriers with high SNR values, while in subcarriers with low SNR only a number of (or no) bits is allocated. This technique is referred to in the literature as adaptive modulation. The allocation of bits to subcarriers is known as the loading technique. In the literature a *number of loading algorithm techniques* are reported [CHOW-95], [FISC-96] with the most well known being that of Chow et al.

In this study, adaptive modulation and channel coding were employed and the performance of the LA technique is compared with the case of fixed modulation. Here, the problem is how a better spectral efficiency  $\Gamma$  is achieved whilst ensuring lowest PER. From [MAR-98], it is known:

$$R_c = \frac{R_b}{\rho} \quad \text{Equation 4-1}$$

where  $\rho$  is the coding rate and  $R_b$  and  $R_c$  are the channel bit rates before and after coding respectively. Furthermore, the modulation level  $m$  and the used bandwidth  $B$  can be expressed as:

$$m = \frac{1}{N_{sc}} * \sum_{j=1}^{N_{sc}-1} m_j = \frac{B}{R_b} \quad \text{Equation 4-2}$$

where  $N_{sc}$  is the number of OFDM subcarriers. By combining 4-1 and 4-2, the spectral efficiency  $\Gamma$  can be expressed as,

$$B = R_c * \Gamma \Leftrightarrow \Gamma = \rho * m \quad \text{Equation 4-3}$$

$\Gamma$ (spectral efficiency)	$\rho$ (coding rate)	$m$ (modulation level)	H2 PHY mode
1	1/2	2 ( $2^2$ )	(1/2)QPSK
3	3/4	4 ( $2^4$ )	(3/4)16QAM

Table 4-1: Spectral efficiencies and corresponding H2 PHY modes.

Here, the cases of  $\Gamma = 1$  and  $\Gamma = 3$  (bit(s) per subcarrier) are chosen and extensively tested. These values correspond to two existing/ standardized PHY modes and are considered to offer the best results compared with the uncoded (no coding) case [BING-00]. In Table 4-1 these two critical

values are shown against the selected PHY modes. The calculation of PER is based on the transmission of LCH PDUs. In order to test the performance of the LA, the algorithm of Chow et al was implemented, using the data of Table 4-1 and Table 3-2.

#### 4.2.2 H2 LA Performance Evaluation

The link adaptation algorithm was simulated and tested in line with ETSI specifications [ETSI-475]. This section presents the link adaptation results. Here, it must be stressed that all the simulation parameters were selected according to ETSI comparison criteria listed in [30701F]. Furthermore, the link adaptation technique has not been investigated in the recent ETSI specification reports [ETSI-475, 761.1]. Hence, here as a complement or/and extension to similar research initiatives [BOL-00], [RAPP-00], the results for the LA technique are presented.

Each H2 MAC frame has a length of 2 ms (40,000 samples) and comprises a broadcast burst with preamble and data ( $N_B$ ), an uplink burst with preamble and data ( $N_U$ ) and random data ( $N_R$  OFDM symbols). The multipath channel of [MED-98], [30701F] was implemented. It is called channel model E. It mainly corresponds to typical large open space environments for nLOS conditions and different average rms (root mean square) delay spreads. For model E, the average delay spread is 250 ns. This channel model makes use of 18 paths, with minimum delay of 10 ns between them. Each path was modelled as a Rayleigh distribution with a Doppler spectrum corresponding to a terminal speed of 3 m/s, and with an average power that decay exponentially with time.

The seven PHY modes of the ETSI H2 standard were initially simulated under the conditions of channel E. The LCH (mainly user data PDU trains) PER statistics and the channel throughput were obtained. The loading algorithm of Chow et al (see Table 4-1) was introduced into two existing PHY modes (1/2QPSK and 3/4 16QAM). The results for PER for all PHY modes and the two selected LAs are shown in Figure 4-1 and Figure 4-2 respectively. While, the results for the throughputs for fixed and adaptive modulation are shown in Figure 4-3 and Figure 4-4 respectively.

Comparing the 1/2LA1 hybrid modulation with the mother one (1/2QPSK) and for  $PER = 10^{-3}$ , there is 0.5-1 dB gain. While for the 3/4LA3 and the mother of 3/4-16QAM the gain is 2.0-2.5 dB (Figure 4-1 and Figure 4-2). Furthermore, in the Figure 4-3 and Figure 4-4, there is an increase in the throughput for the LA1 in the order of 0.5-1 Mb/s. While for the LA3 the increase is about 1.5-2.0 Mb/s.

It must be noted that only the coded cases (for fixed and adaptive modulation) were evaluated in this study. Furthermore, the data rate range was varied from 6 to 54 Mb/s, by using various signal alphabets for modulating the OFDM subcarriers and by applying different puncturing patterns to a mother convolutional code of 1/2.

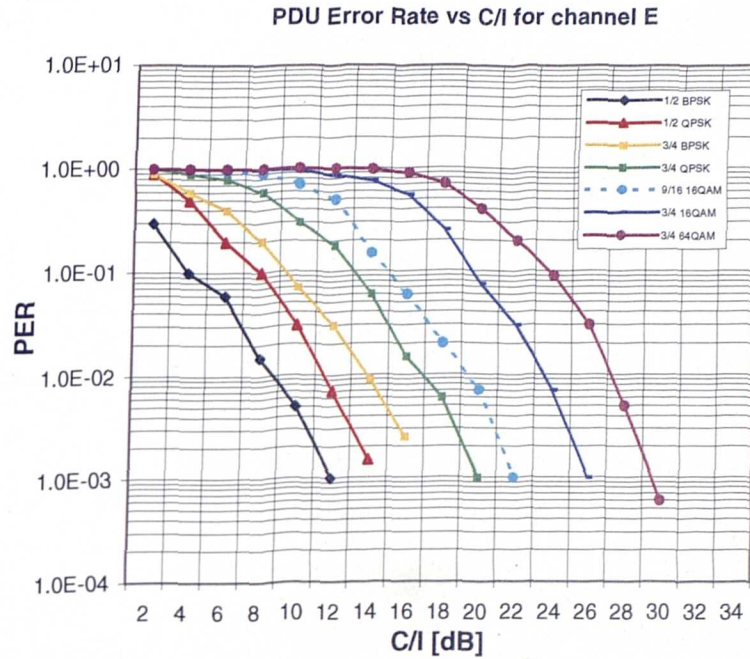
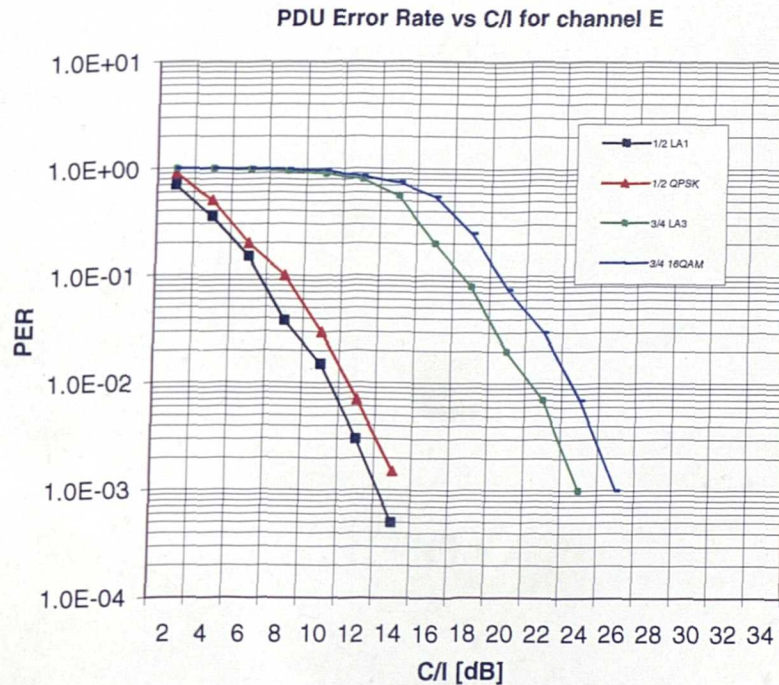


Figure 4-1: PER vs. C/I for all PHY modes (channel E).

Figure 4-2: PER vs. C/I for  $\frac{1}{2}$  and  $\frac{3}{4}$  LA.

Since end-to-end QoS in higher layers (i.e. IP, application level) can only be guaranteed if also the lower layers (i.e. PHY, DLC) are able to provide the required QoS. LA and adaptive modulation play an important role in achieving that. Here, by applying a simple but effective LA technique, the spectral efficiency has been increased helping the overall system performance to significantly

improve. Hence by introducing these enhancements to the PHY of HiperLAN/2, the QoS contracts assigned by higher layers can be maintained and guaranteed.

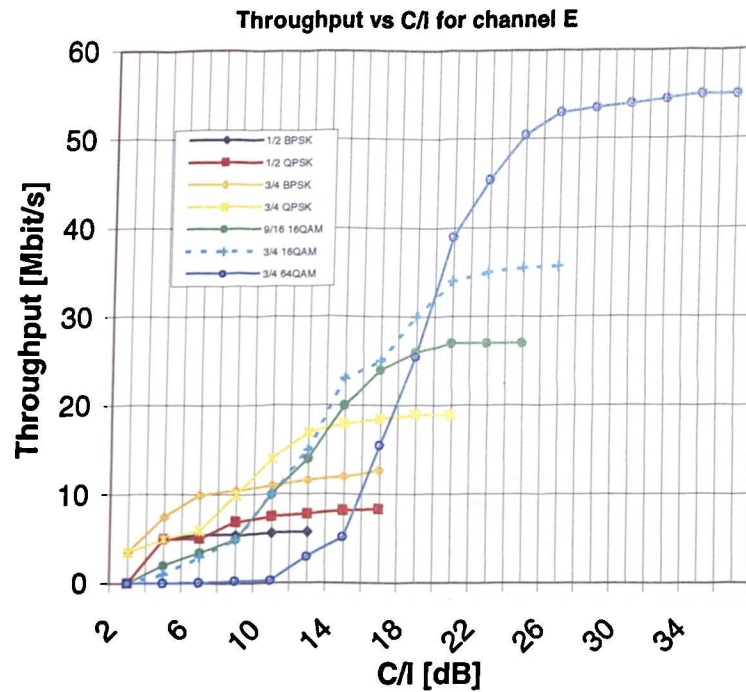


Figure 4-3: Throughput vs. C/I for all PHY modes (channel E).

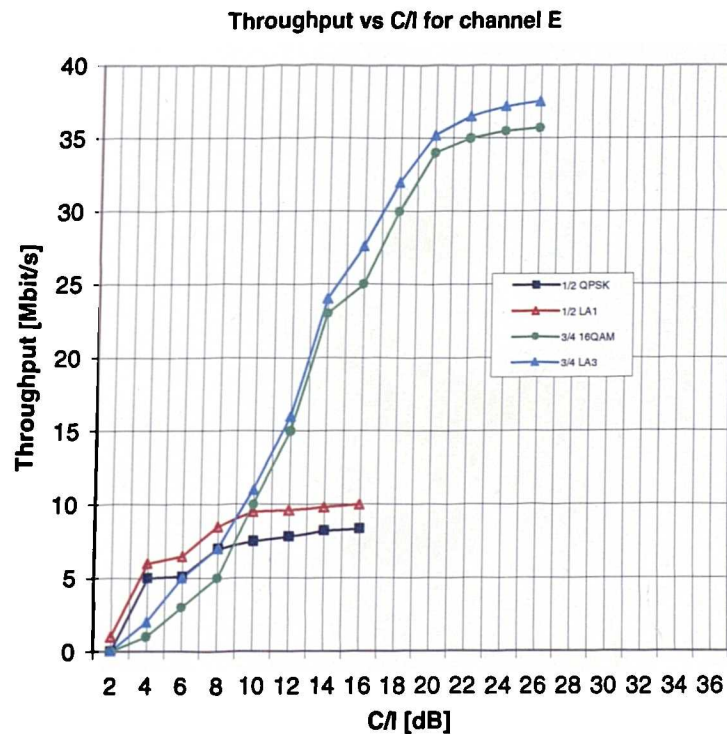


Figure 4-4: Throughput vs. C/I for 1/2 and 3/4 LA (channel E).

### 4.3 Enhancements to HiperLAN/2 MAC Protocol for QoS Support – Cross Layer Cooperation

The following few sections are dedicated to the design and specification of a framework for supporting QoS contracts set by various layers of an IP-based WLAN, e.g. PHY, DLC/MAC or higher layers (IP).

#### 4.3.1 IP Interface Definition

The IP layer should support a full service differentiation node. This implies a packet classifier, a packet scheduler, policy and admission control, and a mapping between the MT and the corresponding IP addresses. Multiple queues will have to be supported, according with the desired levels of QoS. An output queue is required for the exchange of data to the radio layer, but these packets will now be added with the adequate MT identification number. Optionally, header compression can be used (it is not used in this study). The Downlink IP Layer structure is shown in Figure 4-5.

Associated policy and admission control entities may impose further restrictions to the classifier, which will be discussed further on.

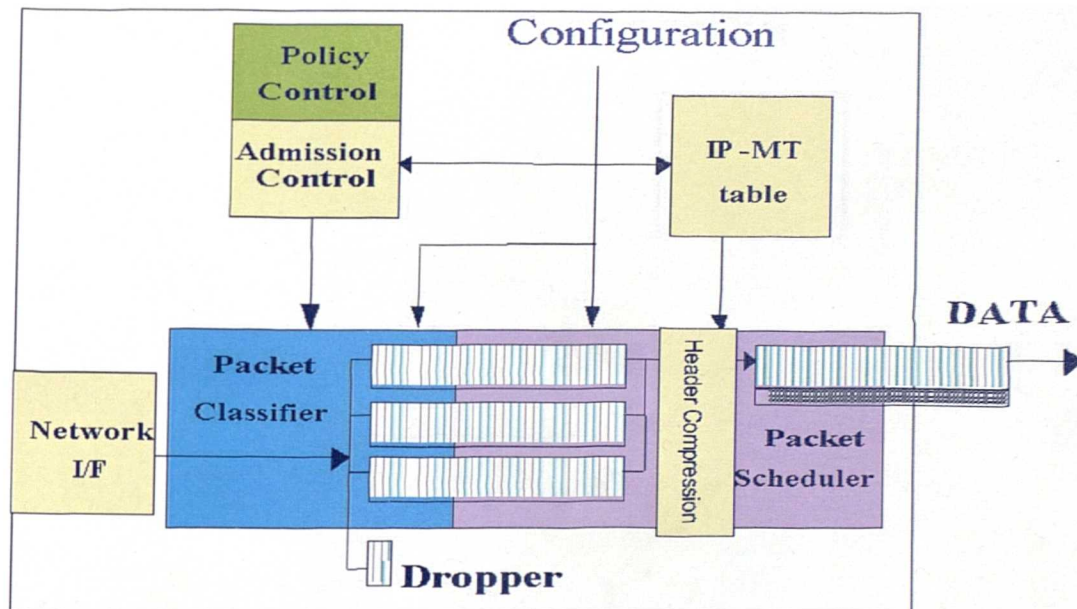


Figure 4-5: DL IP Layer structure.

The Radio layer is quite different. Given the packets sequentially fetched from the IP layer, it has to put them in the proper radio communication channel. This implies that radio channels will have to be open for this action to take place. Thus, the radio layer will have two different key entities: a DFS mechanism and a channel creation unit (responsible for the creation and deletion of logical

channels, each one with different end point and different QoS associated). Associated units required are a service classifier, which will provide information to the DFS and a channel mapper (mapping each packet in the proper radio channel, and performing the eventually required segmentation of the packets). The reference for these channels can be conceptually built with the MT identifier, associated with the type of requested service. The DFS (Dynamic Frequency Selection) will use this logical reference to index a table, and retrieve the appropriate information concerning the existing physical channel. The PHY architecture is shown in Figure 4-6

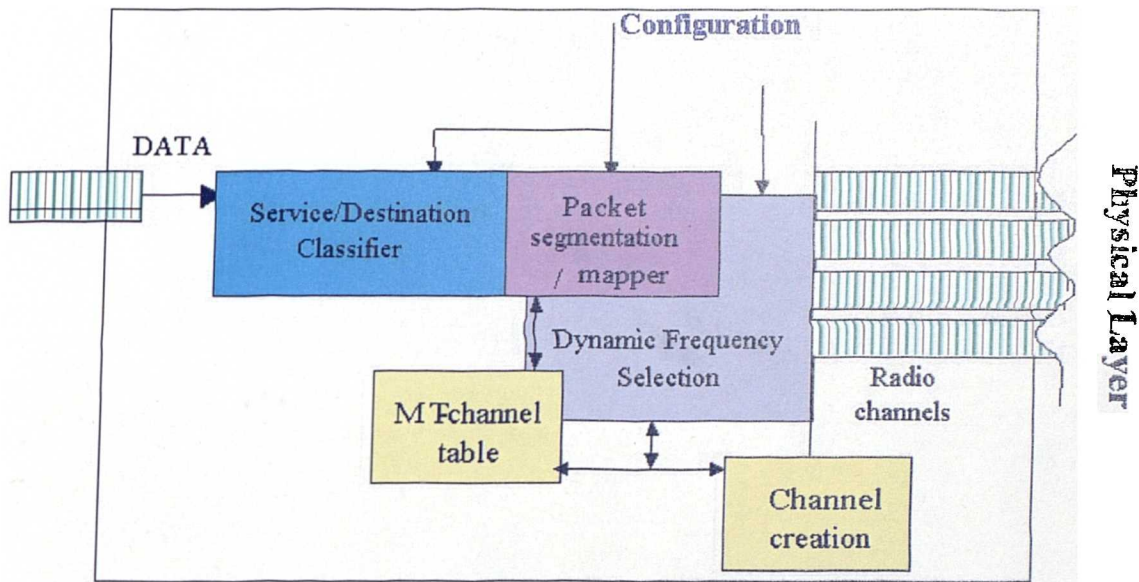


Figure 4-6: DL PHY structure.

In the uplink case, the system is somewhat simpler. There is no need for a packet scheduler on the IP layer, as the PHY should have larger bandwidths than the wireless layer, but a classifier should be used nevertheless, as shown in Figure 4-7.

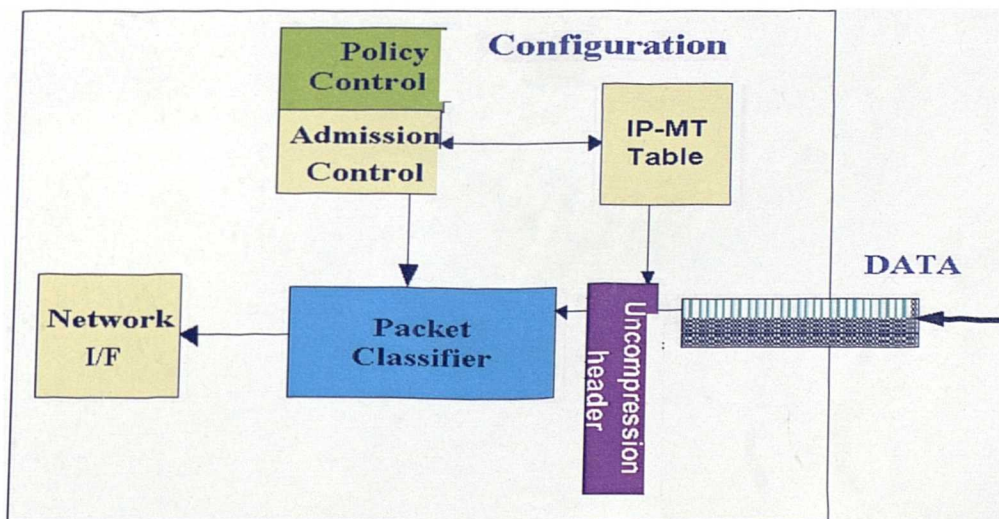


Figure 4-7: UL IP Layer Structure.

On the other hand, the radio link is not much simpler, and the DFA (Dynamic Frequency Allocation) will have to be able to manage both UL and DL coherently. Furthermore, while DL traffic can be a function of the incoming packet traffic, UL traffic has to be signalled initially to the AP for channel/frequency allocation to be performed. Channel destruction, however, can be simply handled based on existing traffic in that channel. Nevertheless, a radio link layer protocol will be required for these channel requests. UL PHY is shown in Figure 4-8.

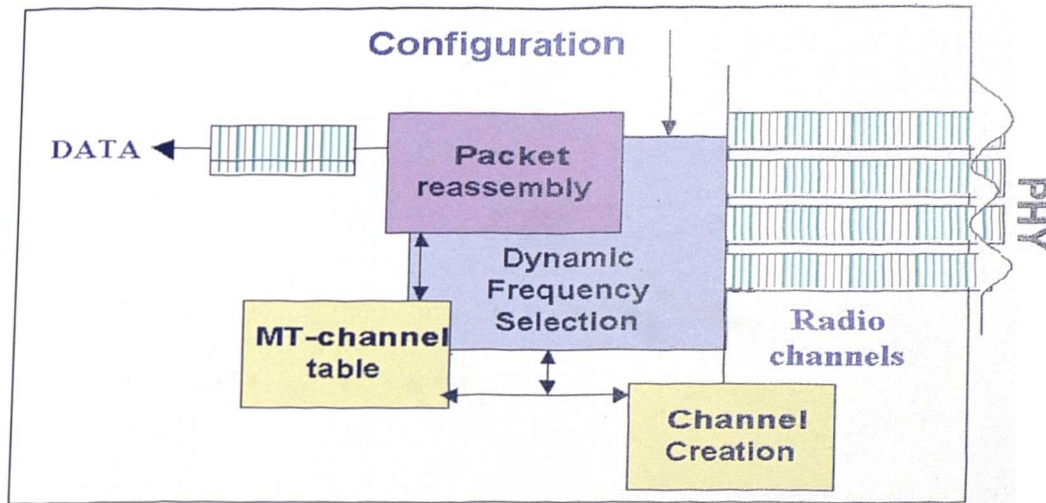


Figure 4-8:UL Radio Layer.

### 4.3.2 QoS Support and Scheduling

The basic way to provide QoS is by giving one user priority over another, and/or to control the bandwidth usage of certain users so that a pre-assigned QoS can be assured. Furthermore, in order to support the variety of services, it is necessary to develop a QoS mechanism/framework that can support multiple services and traffic classes. These traffic classes will classify packet streams according to their real-time needs. Four traffic classes are selected, in close relation to the UMTS QoS Classes defined in [TS-23.107] and used in [WIS-01], the following (for more details on these classes refer to Chapter 3, section 3.5):

- Conversational (CO),
- Streaming (ST),
- Interactive (IN),
- Background (BG or BE).

The main distinguishing factor between these QoS classes is how delay sensitive the traffic is: Conversational class is meant for traffic which is very delay sensitive while Background class is

the most delay insensitive traffic class. Conversational and Streaming classes are mainly intended for carrying real-time traffic flows. The main divider between them is how delay sensitive the traffic is. Furthermore, Conversational class supports mainly bi-directional traffic with uplink and downlink QoS provisioning. Examples of Conversational real-time services are, video telephony or voice over IP.

Interactive and Background classes are mainly used by traditional Internet applications like WWW, Email, Telnet and FTP. The main difference between Interactive and Background class is that Interactive class is mainly used by interactive applications, e.g. interactive Email or interactive Web browsing, while Background class is used for background traffic. Traffic in the Interactive class has higher priority in scheduling than Background class traffic, so background applications use transmission resources only when interactive applications do not need them. This is very important in a wireless environment where the bandwidth is low compared to fixed networks.

At the same time, there are two commonly used models for end-to-end QoS, which are integrated services and differentiated services. In this study, two service models are supported, the IntServ model and the DiffServ model. In the Integrated Services domain [RFC-2210, 2211, 2212], network resources are explicitly identified and reserved per flow. Each network node classifies incoming packets and uses the reservations to provide QoS. While in the Differentiated Services domain [RFC-2474], the resources are reserved per aggregate of flows. The traffic is differentiated into a set of classes; while the network nodes provide priority-based treatment based on these classes (Figure 4-9).

Nevertheless, the QoS mechanism must obey certain rules, according to selected QoS metrics. Then, the revised structure of MAC protocol should adapt these metrics, in order to establish radio links with specified QoS. The metrics can be expressed as a set of parameters, such as delay, jitter, PER and throughput. Depending on the amount of traffic to be delivered on the radio channel, QoS contracts assigned by higher layers have to be classified, prioritised and handled at the MAC layer (by the classifier and scheduler).

The enhancements in the H2 MAC protocol were introduced in order to support three “nodal points/areas” and two service models (see Figure 4-9). While the control and data domains as well as the interworking/cooperation of various layers (and sub-layers) of the IP-based H2 network (i.e. PHY, DLC, CL, IP) are shown in Figure 4-10 and Figure 4-11 respectively.

In the first domain, the traffic is classified (1<sup>st</sup> focal point) by source and destination address, type of service or/and other fields (per flow treatment). Then, two queuing (2<sup>nd</sup> focal point) and a scheduling (3<sup>rd</sup> focal point) schemes are used, namely the Weighted Fair Queuing and the Class-Based Queuing. Concurrently, a reserved path, which is RSVP enhanced, is introduced in the

IntServ domain. This mainly supports conversational and streaming traffic classes and multimedia services (e.g., VoIP, video conferencing/streaming); while background BE traffic is also allowed to flow (a simulation assumption).

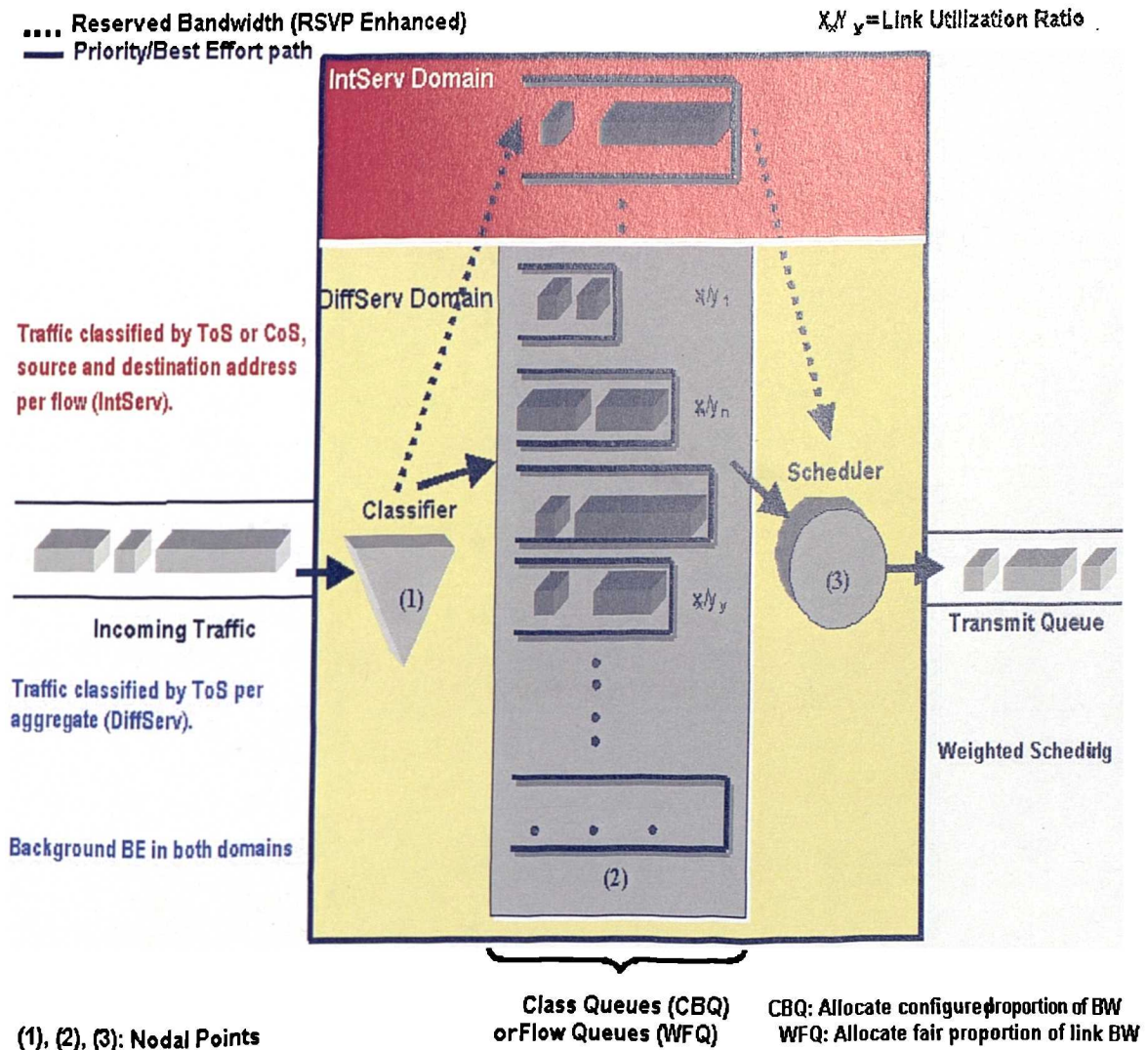


Figure 4-9. QoS Provisioning for all-IP HiperLAN/2.

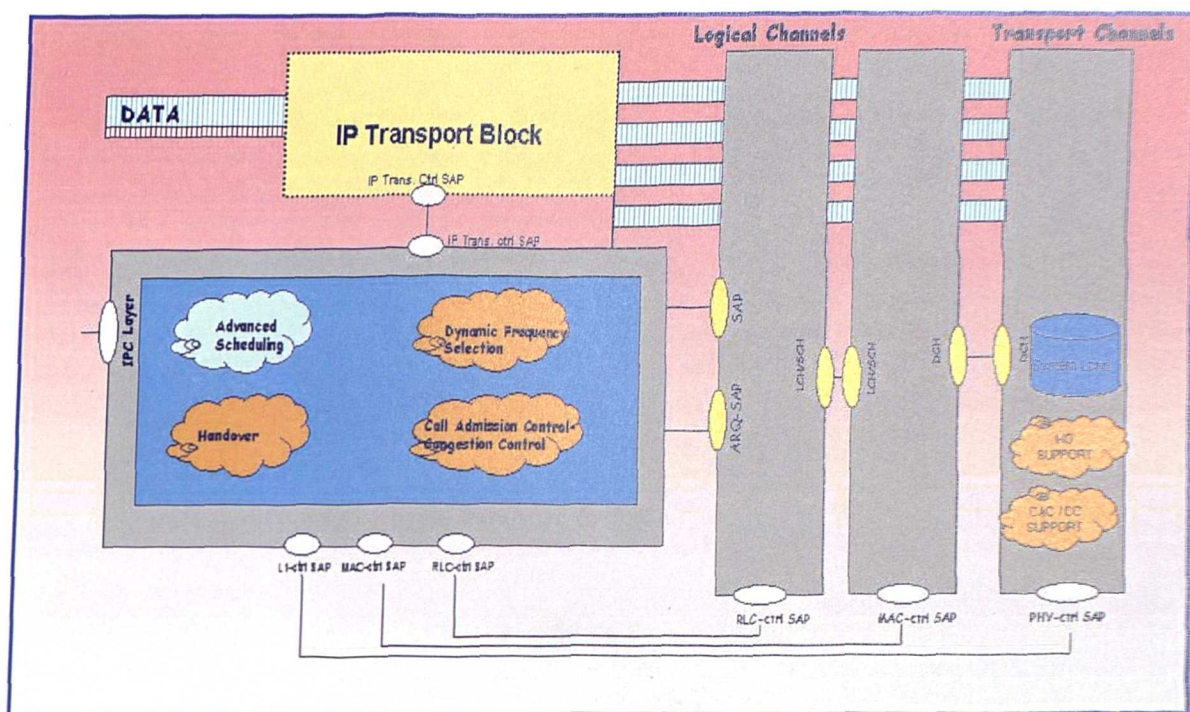


Figure 4-10: IP-based H2 control plane (layers interworking/cooperation).

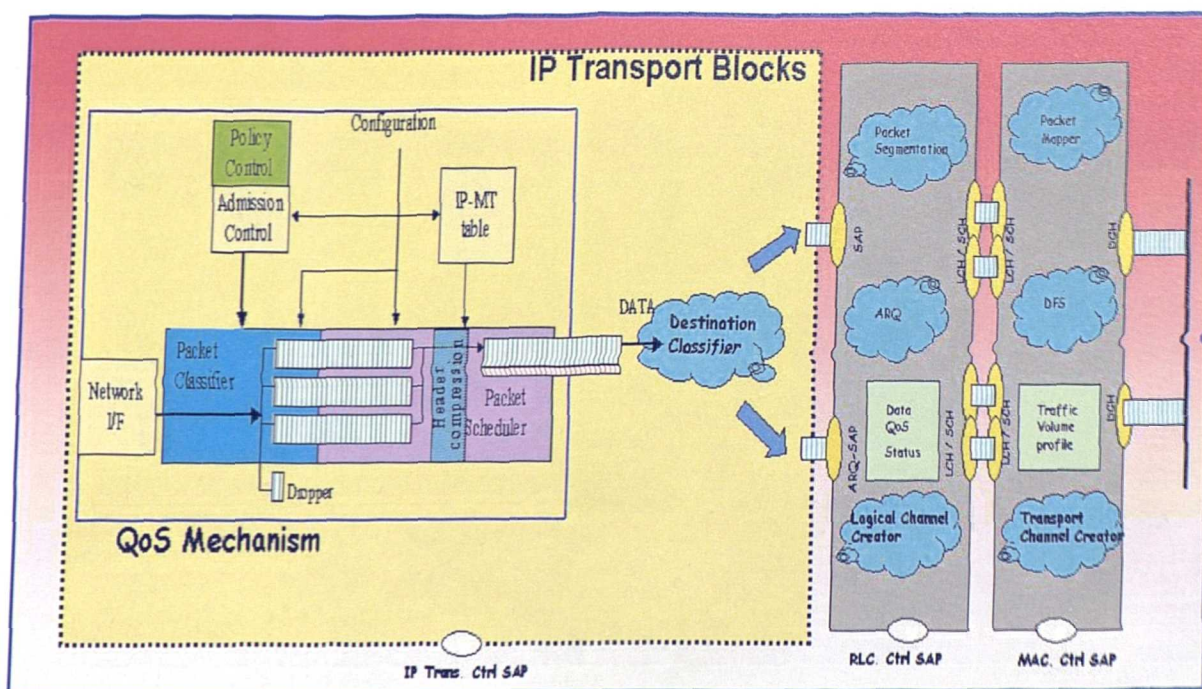


Figure 4-11: IP-based H2 data plane (layers interworking/cooperation).

On the other hand, in the DiffServ domain the classification is handled by the ToS per aggregate in the HiperLAN/2 AP, the rest of the focal points remain the same. Next, a brief introduction to WFQ and CBQ algorithms is given.

**Weighted fair queuing** [KES-91] prioritises traffic flows while ensuring that lower priority flows receive an acceptable level of service. Traffic is classified and put into queues that are assigned

relative weights/priorities. The queues are serviced on a round-robin basis and the amount of data removed from each queue is in proportion to the relative weight of the queue. Higher priority queues get a larger portion of the bandwidth, but even the lowest priority queue always receives some bandwidth. This is an important feature in the bandwidth limited wireless environments.

**Class-based queuing** [WAK-95] goes beyond prioritisation by providing assured bandwidth to traffic flows. Traffic is first classified and put into queues. Each queue is assigned a number of bytes that will be forwarded each time the queue is serviced. The queues are serviced on a round-robin basis and the assigned amount of data is forwarded from each queue. Each traffic class is therefore allocated a percentage of bandwidth on the outbound link and may burst above their allocated bandwidth if other traffic classes are not using the bandwidth. These characteristics make this algorithm less “fair” in distributing the resources in bandwidth and interference limited environments and when it serves multimedia or mixed traffic applications.

### **4.3.3 Simulation models and Performance analysis**

This study is performed and supported by means of computer simulations, with each simulation experiment consisting of a network scenario with variables:

- ✓ Five load conditions and changing each time the two traffic controllers (either WFQ or CBQ),

and a traffic scenario with variables:

- ✓ Two difference service domains (DiffServ or IntServ) and changing each time the services (HTTP or VoIP or VoD).

Overall sixty experiments were performed, giving the sequence of results shown in the following subsections.

#### **4.3.3.1 Network Related Aspects**

The network topology consists of a wireless and a wired part. The wired part is using full duplex links (10Base-T, 100Base-T or PPP-DSX) interconnecting high performance gateway routers with multimedia servers/clients. The wireless part (WLAN) can support data rates up to 12 Mb/s. In order to model the wireless channel characteristics, a Gilbert model [GILB-60] was incorporated to introduce bursty bit errors in the transmitted packets.

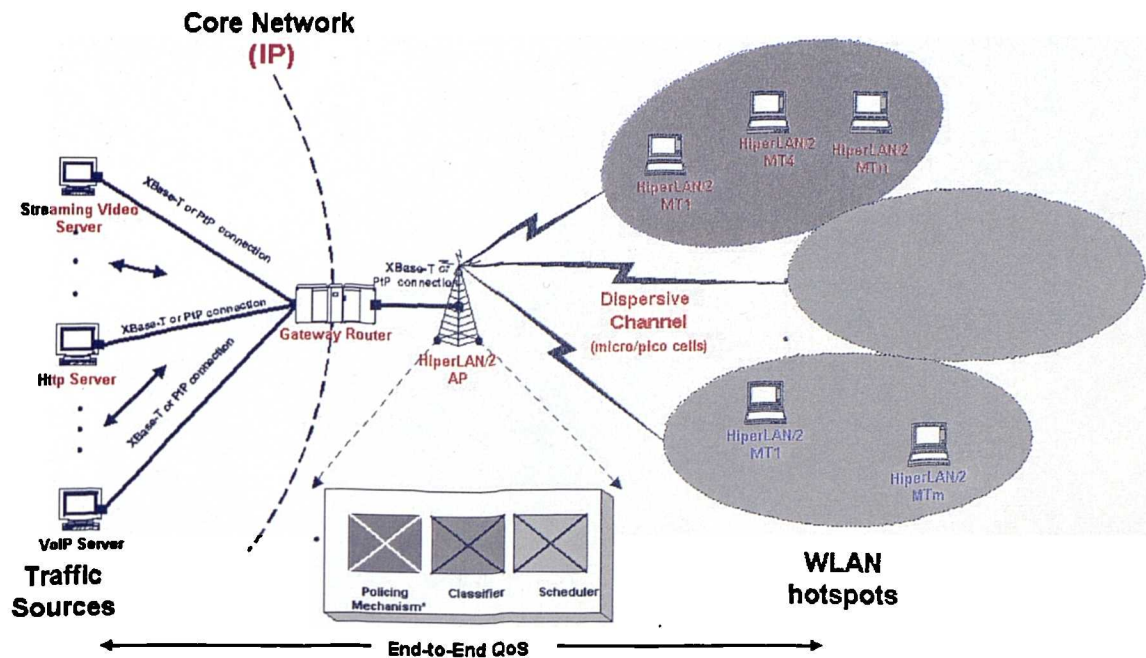


Figure 4-12: Simulation scenario(s).

Network scenarios are also considered with the four aforementioned classes of service, whereby each Class of Service is represented by a separate output queue in the AP. At the same time, single traffic types are considered, in network topologies with single and multiple traffic classes [KAD-00]. As it can be seen from Figure 4-12, one or many traffic sources are identified (generating traffic of a given type) as the originating source and a number of mobile users (from 1 to 20) as the target sources. The classifying-queuing-scheduling is performed at the access point. Then, a number of QoS parameters can be measured, including the end-to-end delay, the throughput of a given system load and for various classes and services, at the two service domains.

#### 4.3.3.2 Traffic Generation Related Aspects

In order to produce meaningful results, particular attention is paid to the model(s) used in generating each type of traffic. This study is limited to voice (VoIP), video (streaming, VoD) and data (http/WWW) traffic.

- **Voice:** For the modelling and description of the VoIP traffic refer to chapter 3, section 3.6.4.
- **Video:** For the modelling and description of the streaming video traffic refer to chapter 3, section 3.6.5 (where an extended study of MPEG modelling is given).
- **Data:** For the modelling and description of the Internet traffic (WWW) refer to chapter 3, section 3.6.1 and 3.6.2.

### 4.3.3.3 Performance Related Aspects [POL-02]

In order to evaluate the performance of the system, we have selected a number of QoS metrics. These metrics can provide a clear idea of the system performance, the traffic requirements and the status of the QoS contracts. At the same time, packet loss may be due to, in the case of real time applications, the ete-delay became larger than the specified maximum delay requirements. Hence, it is important to derive the probability that the ete-delay  $D_{ETE}$  exceeds a threshold  $D_{THR}$ ,  $P_r(D_{ETE} > D_{THR})$ . Since packet loss may be tolerable up to a given maximum delay  $Dr_{(max)}$ , then the comparison between  $D_{ETE}$  (where condition the  $P_r(D_{ETE} > D_{THR})$  must always be true) and  $Dr_{(max)}$  sets the grounds on whether an application can be accepted or not.

The principal requirement for real-time voice and video communication is the bounded delay; if either voice or video samples are not delivered to the receiver on time, then they will have to be discarded. Following ITU-T recommendation G.114 [ITU-96] as well as the Tables in Chapter 3 sections 3.5.5 and 3.5.6, we consider a 100 ms ete-delay for VoIP traffic, from which we allocate 10 ms bound for network delay,  $D_{NET}$ . Furthermore, according to IEEE 802.1D recommendation [KAR-00] and Tables in Chapter 3 sections 3.5.5 and 3.5.6, we consider an ete-delay requirement of 200 ms for video (streaming) traffic, from which we allocate a  $D_{NET} = 100$  ms bound for network delay. In equation 4-4, the conditions for delays bounds are mathematically given:

$$\begin{aligned} P_r\{(D_1 = D_{ETE(vo)} < D_{THR(vo)}) \cup (D_2 = D_{THR(vo)} \leq Dr_{(max)})\} = \\ = \sum_{i=1}^2 P_r(D_i) \text{ where } D_{ETE(vo)} = D + \sum_{i=g}^r d_i + \sum_{j=g}^r P_j + T_r \end{aligned} \quad \text{Equation 4-4a}$$

and for video the delay conditions are:

$$\begin{aligned} P_r\{(D_1 = D_{ETE(vi)} < D_{THR(vi)}) \cup (D_2 = D_{THR(vi)} \leq Dr_{(max)})\} = \\ = \sum_j P_r(D_j) \end{aligned} \quad \text{Equation 4-4b}$$

Delay Requirements → Priorities and Classes ↓	$D_{NET}$ (ms)	$D_{THR}$ (ms)	$Dr_{(max)}$ (ms)
Conversational RT	10	100	150
Streaming RT	100	200	400
Interactive BE	100	1000	n.a.
Background BE	n.a.	n.a.	n.a.

Table 4-2: Delay requirements for voice and video traffic.

All these variables are delay components and are defined accordingly:  $D$  is the duration of the time sequence,  $d_i$  is the time to send/receive a packet to/from the network,  $P_j$  is the time elapsed between the sound sequence and the time the packet sent/received to/by the network, and finally  $T_r$  is the transmission delay. Furthermore, these conditions/requirements are summarised in Table 4-2.

For HTTP traffic, the burstiness is in the order of hundreds. For this type of traffic, throughput is a more appropriate QoS metric than end-to-end delay and jitters. The throughput is defined as the rate (in bit/sec) of data packets arriving at the MT. Finally, in this study; we have set the guaranteed bit rate to 1024 kb/s and the maximum bit rate to 6 (or 12) Mb/s (depending on the modulation used).

#### 4.3.3.4 Performance Evaluation Analysis

In this section, we investigate the behaviour of the H2 QoS mechanism, for each traffic type and in two different domains: IntServ and DiffServ.

**Data.** In this scenario, multiple http sources are multiplexed into the same output queue (see Figure 4-12). They are given different priority according to their CoS/ToS classification (interactive or background BE). The results for the system throughput for both domains are shown in Figure 4-13 and Figure 4-14. In Figure 4-13 and for 80% load (i.e. 16 users), the WFQ algorithm gives 5 kb/s increase in the system throughput, while in Figure 4-14 the absolute increase between the two algorithms remains the same, but the RSVP-enhanced domain provides a slight improvement (5-8 kb/s). Therefore, in this case the domains behave in an analogous manner towards http traffic. It must be noted, that for low load (4-8 users) in both domains, the performance of the CBQ algorithm is almost the same with that of WFQ. Finally, the performance of the system has not been significantly improved in the IntServ domain, as it was expected, due to the traffic type (WWW).

**Voice.** The same experiment is repeated for VoIP traffic. The priorities are given here according to ToS (VoIP and BE traffic) and CoS (CO, BE). As, it can be seen from Figure 4-15 and Figure 4-16 the lower delay bounds are kept for both algorithms and domains. For an 80% load the WFQ algorithm offers a better performance in the order of 4 ms for IntServ and 6 ms for DiffServ. The performance of both algorithms remains strictly between the delay bounds/thresholds, even in a congested H2 network. Moreover, the RSVP-enabled domain provides a significant improvement to the overall system performance, with the drawback being the increased control and signalling load (scalability). This is since RSVP is based on soft states, each connection needs periodic signalling activity, therefore the work required for the same amount of streams is much higher.

*Video.* In this last scenario, multiple multiplexed video streams are injected into the network. Again the priorities are given according to ToS (video streaming and BE traffic) and CoS (ST, BG). As expected, the resulting delay behaviour is mostly influenced by the encoding control scheme. Again for both algorithms and domains the delay requirements were well preserved. In Figure 4-17 and Figure 4-18 and for an 80% load the DiffServ WFQ gives 20 ms improvement to the system performance, while this absolute difference is maintained for all the specified load conditions.

In the IntServ domain the CBQ drops the (algorithms) marginal difference to approximately 15 ms. The RSVP deployment provides, as for the voice case, an improvement in the order of 30 to 35 ms. It must be stated that both algorithms and domains behave well in all kinds of traffic loads.

The ETSI standards have not explicitly defined a framework for QoS provisioning in all-IP WLANs (H2). In order to fill in this requirement, we have set up a QoS framework with respect to the investigation of MAC classifiers and schedulers for supporting direct IP traffic in different service domains. The H2 MAC protocol was built and simulated, and the aforementioned enhancements were introduced. These modifications could be considered as possible enhancements to the MAC protocol for an IP-based HiperLAN/2 [POL-02].

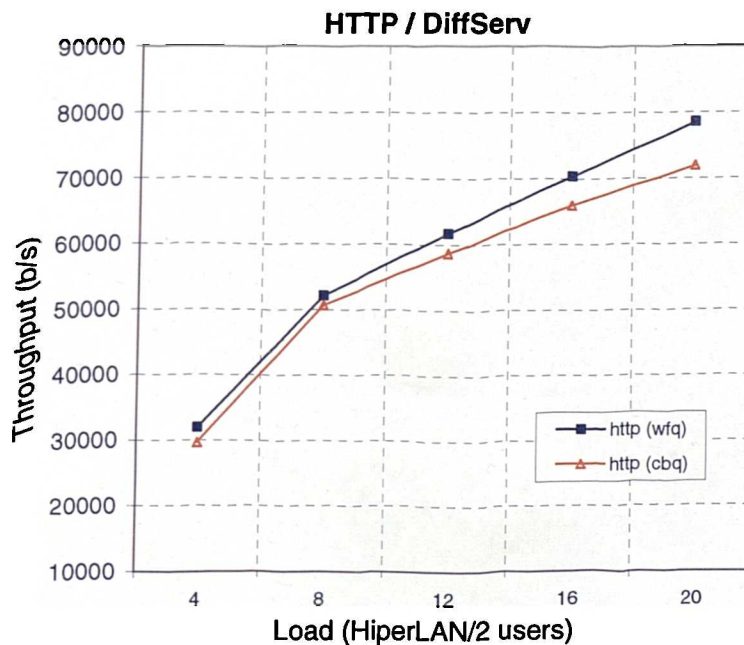


Figure 4-13: Throughput vs. load for http traffic (DiffServ).

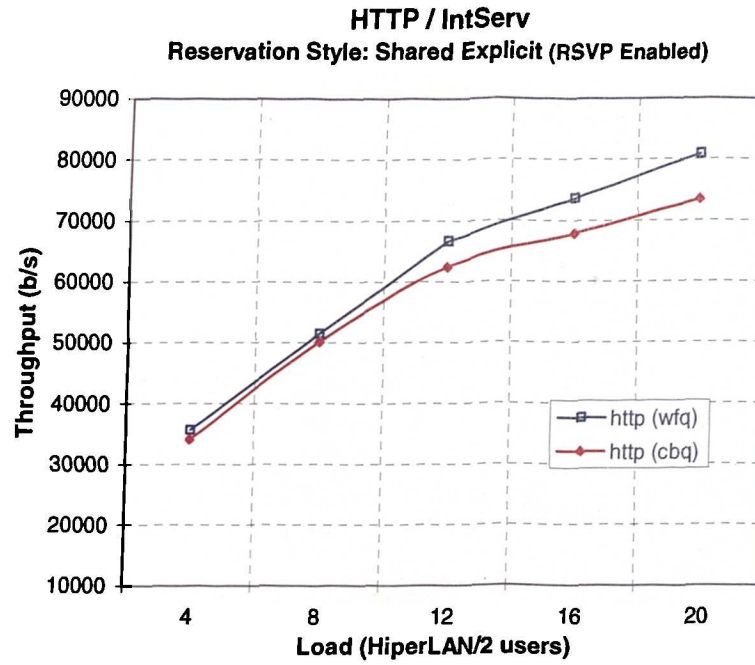


Figure 4-14: Throughput vs. load for http traffic (IntServ).

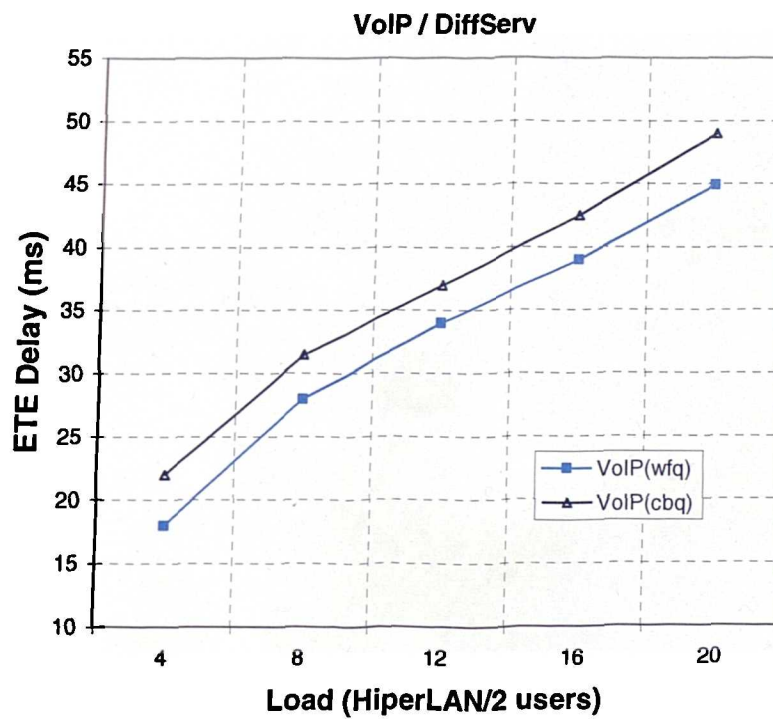


Figure 4-15: Ete-delay vs. load for VoIP (DiffServ).

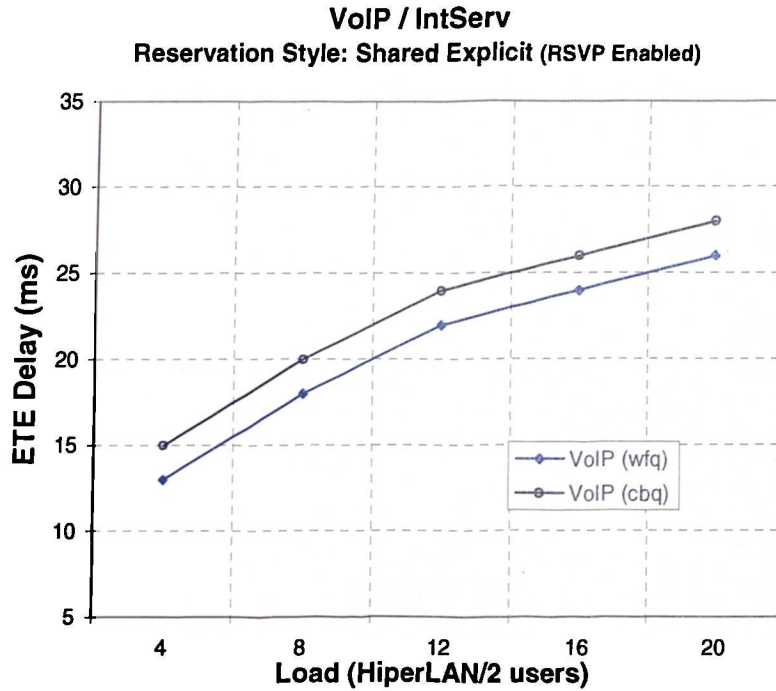


Figure 4-16: Ete-delay vs. load for VoIP (IntServ).

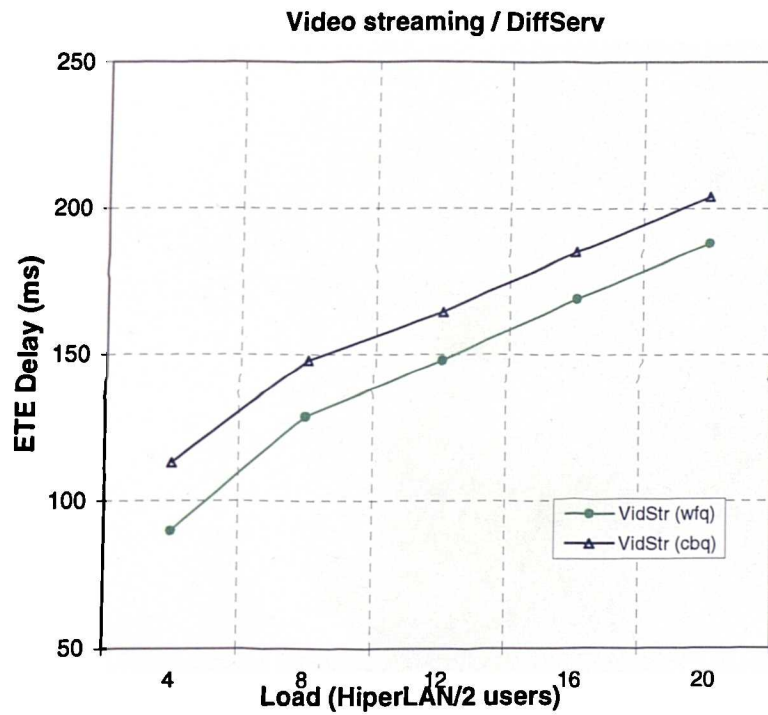


Figure 4-17: Ete-delay vs. load for video streaming (DiffServ).

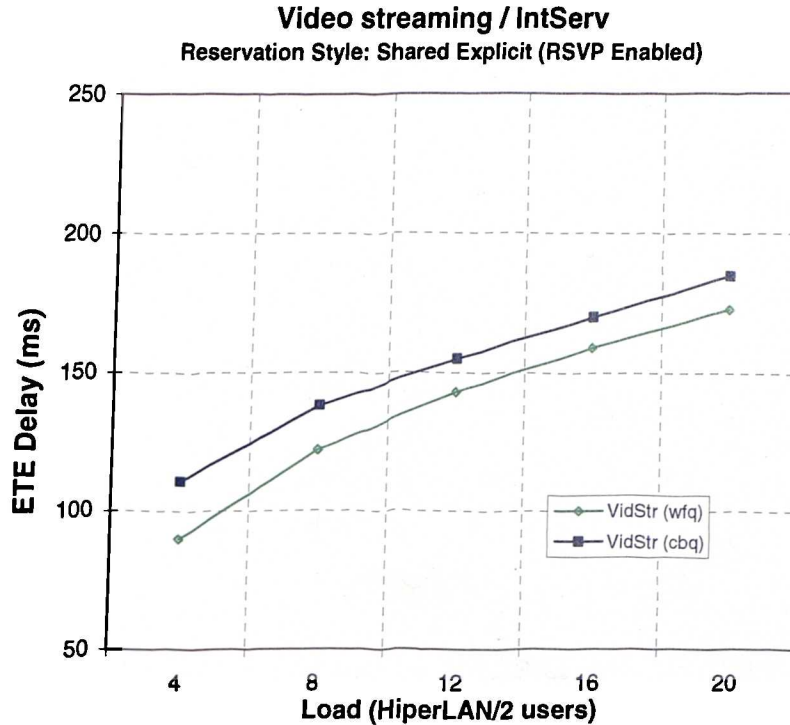


Figure 4-18: Ete-delay vs. load for video streaming (IntServ).

## 4.4 Connection Admission Control for IP-Based HiperLAN/2

This section discusses the enhancements to the QoS framework proposed in the previous section, in terms of the connection admission control algorithms. Since, radio resources are typically scarce and packet loss may be extensive, which inflicts fluctuations in the QoS. Therefore, a CAC mechanism is examined and adjusted to this framework, while real-time (VoIP) multimedia services are used in order to test the QoS support capabilities of the whole WLAN system.

### 4.4.1 Admission Control Algorithm

The Protocol Functional Entities of interest are the DLC and the CL layers. In the DLC, the Radio Link Control's functions of Congestion Control and Connection Admission Control are the main target. These are situated in the control plane (see Figure 4-10). Furthermore, an IP service specific convergence sublayer is used as part of the CL layer. The complete structure of the H2 PFEs is shown in Figure 4-19. The QoS framework proposed in [POL-02] is enhanced by introducing CAC and CC mechanisms. In the framework, the CC management was based on the WFQ or CBQ algorithms and the packet flows were treated in an aggregate (DiffServ) or per flow (IntServ) basis. The same exactly CC mechanisms are used and compared here, but enhanced by the CAC mechanism.

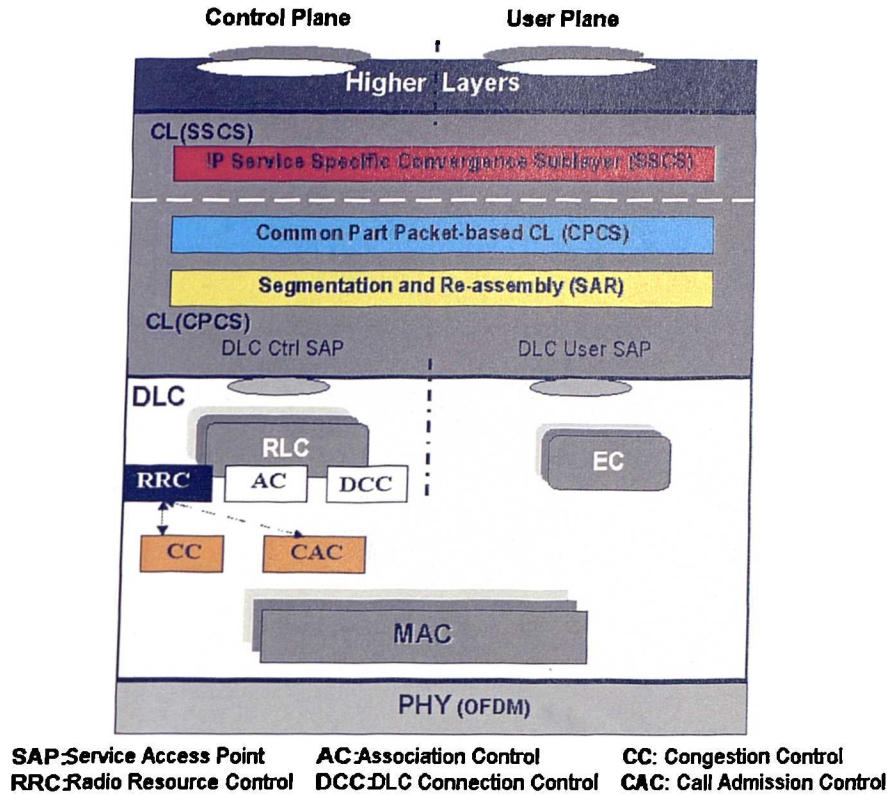


Figure 4-19: HiperLAN/2 Protocol Functional Entities.

The CAC entity is important, because it is the main decision maker, since QoS includes the specific sub-problem of connection admission control i.e. how to decide whether to admit a request (i.e. a user).

A simple CAC algorithm is set up in line with ETSI specifications [ETSI-761.1]. This CAC function determines which DLC connection will be accepted or rejected. The decision depends on the QoS metrics, requirements and traffic contracts and on how a given channel access attempt will be made depending on whether the channel is busy or idle and what priority level the attempt will be given.

The message sequence chart of Figure 4-20, describes how this algorithm works and shows the interaction between different PFEs. The case here is that the H2 MT requests resources, and therefore different layers from the H2 protocol stack (see Figure 4-19) are involved, i.e. MT's IP SSCS communicates with the AP's IP SSCS via the DLC layer to request resources (CAC and CC). There is also a traffic table (TT), which is regularly updated [OIK-01].

The *Grade of Service* is used to determine the maximum load that can be carried and can be theoretically expressed as:

$$GoS = \frac{B_{calls} + 10 \cdot L_{calls}}{N_{totalCalls}} \quad \text{Equation 4-5}$$

where  $B_{calls}$  are the blocked calls,  $L_{calls}$  are the lost calls and  $N_{totalCalls}$  is the total number of calls.

Furthermore, the probability  $P_k$  that  $k$  users out of  $m$  are active (each user having an activity factor  $a$ ) in the network for an M/M/m/m Engset loss system (number of servers is the same as the number of clients/users) is given by:

$$P_k = \binom{m}{k} \cdot a^k \cdot (1-a)^{m-k} \quad \text{Equation 4-6}$$

which means that the number of active users is binomially distributed.

At call set-up CAC has to determine the availability of requested resources. CAC must be prepared to release pre-allocated resources upon request of resource control because serving of existing calls should have higher priority than satisfying call set-up requests. Furthermore, in order to guarantee a certain degree of mobility, CAC has to reserve some of the available resources for possible handovers. A blocked handover results in an interrupted call and will occur under certain circumstances e.g. the bandwidth required by a connection cannot be supplied in the new cell location. The optimum balance between set-up blocking probability and handover blocking probability is determined by the total expected terminal mobility. The handover blocking probability may be requested at connection set-up time as a new QoS parameter. QoS renegotiation might be a possibility.

At the same time, CAC is based on assumptions about:

- A Resource Request (RR) and Resource Grant (RG) polling algorithm (see Figure 4-20). The RRs are transmitted in the SCH PDU or if there are several mobile terminals competing for resources in the RCH phase. The information for a RR is cached in the FCH for both UL and DL. While, RGs are transmitted in the next MAC frame on the SCH PDU and ACH phases and are cached in the FCH. Then, the AP knows the exact bandwidth that has to reserve (if any) and accordingly allows another call or service to be accepted or rejected or even an existing one to be blocked (these thresholds are computed in our simulations by using Equation 4-5). Furthermore, this decision making is based on the QoS classes (QoS framework) defined in section 4.3.2.
- Resources blocked by existing calls;
- Available physical layer radio resources for the future;
- Expected behaviour of the new call to accept, which is based on and influenced by the user mobility pattern (the information could be stored on FCH phase and transmitted on the next MAC frame in the ACH and SCH phases);
- Expected mobility requirements (same as above) for the MT issuing the call request;

- Expected mobility (same as above) of MTs in neighbouring cells.

CAC should admit new calls with respect to a certain threshold for cell resource utilisation. Threshold adjustment must be made by resource control depending on a statistical analysis of current traffic mix per AP, the number of MTs served by the AP under consideration, their virtual channel quality and the expected call duration per DLCC [EURE-00].

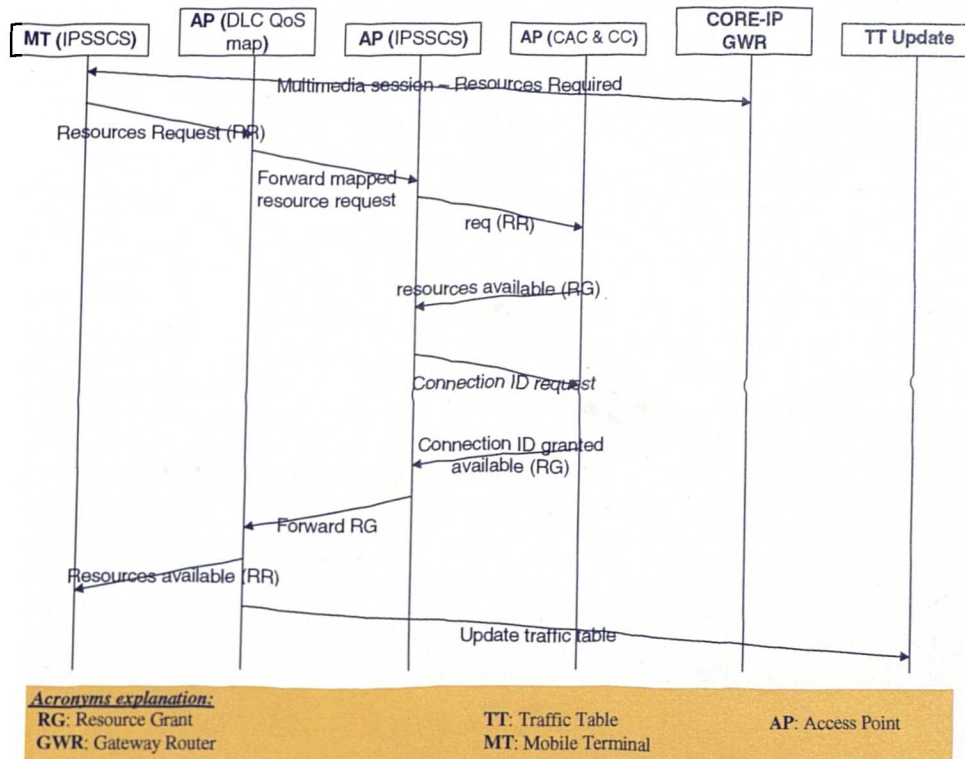


Figure 4-20: H2 MT requests resources [POL-TAF].

#### 4.4.2 Numerical Results and Analysis

A critical QoS metric for testing and analysing CAC algorithms is *blocking probability* ( $P_b$ ), which is defined as follows. Since cell capacity is limited, some calls attempted may be blocked. The probability that a new call is not admitted into the system is called the blocking probability.

Figure 4-21 shows the system performance as a function of load, when CC algorithm is the WFQ. It can be clearly seen that the DiffServ domain performs better under loaded conditions than the IntServ domain. This is due to IntServ domain scalability issues caused by RSVP's periodic signalling activity. For a load of 15 Erlangs or 75% utilization the  $P_b$  is 8% difference to the favour of the DiffServ domain.

Concurrently, Figure 4-22 shows the system performance as a function of load, when CC algorithm is the CBQ. Similarly to the previous case, the DiffServ domain performs better under

loaded conditions than the IntServ domain (signalling load only). For a load of 15 Erlangs or 75% utilization the Pb is 12% difference to the favour of the DiffServ domain. Comparing the two Figures, it is also apparent that WFQ congests the network less than the CBQ does.

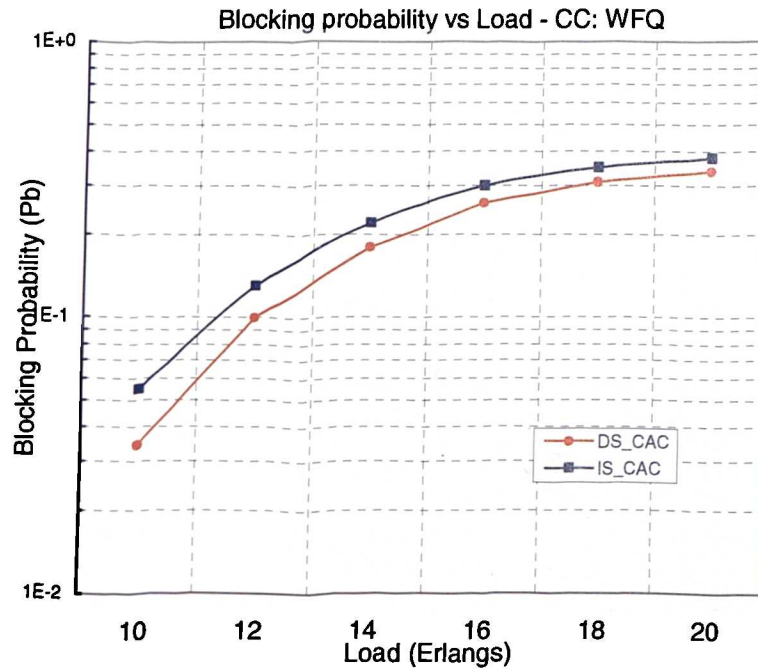


Figure 4-21: Comparison of blocking probability as load varies (CC: WFQ).

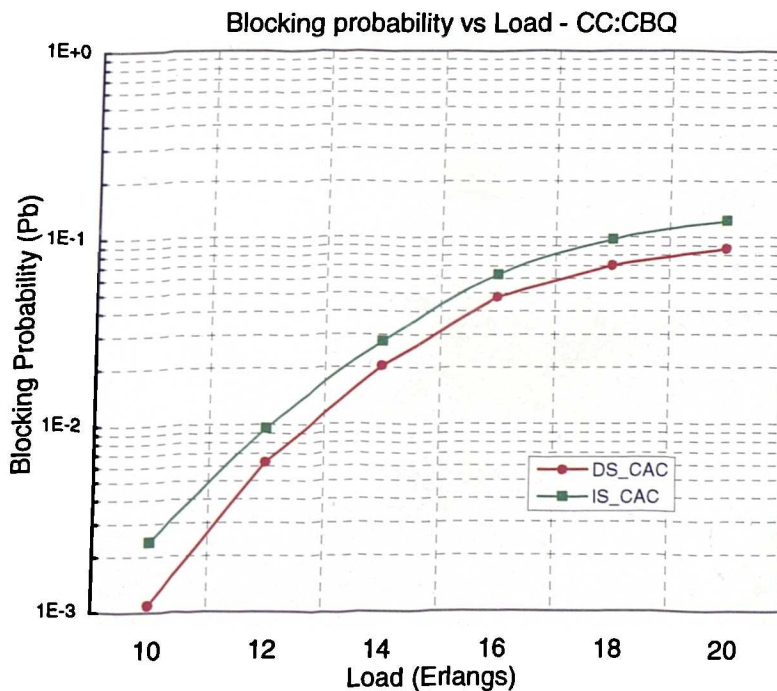


Figure 4-22: Comparison of block probability as load varies (CC: CBQ).

## 4.5 Discussion and Summary

With the radio link being the most vulnerable part of a wireless network, suitable techniques and algorithms should be introduced, to maintain the requested or agreed QoS levels. In doing so, this study suggested the mapping of the IP layer QoS to the MAC layer QoS as well as the QoS “interworking” for different WLAN layers and protocols (PHY-DLC-CL-IP) for both control and data planes.

On the PHY level, link adaptation techniques have been studied and an algorithm proposed to enhance and potentially improve the QoS support/guarantees (in terms of gain and throughput). The enhanced system was compared to the existing and the overall system performance showed significant improvement, in the order of approximately 1 to 3 magnitude units.

On the MAC level, various traffic types and classes in three nodal areas (classifying, queuing and scheduling) and for two-service domains (IntServ and DiffServ) were investigated under different network loads and topologies. Furthermore, *different QoS metrics* were considered, in order to allow IP connections according to their specified QoS requirements/contracts. Two well-known “nodal” algorithms were selected to fill in the H2 standardisation gap: WFQ and CBQ. In the IntServ domain, RSVP signalling was used. The performance evaluation for three types of services, namely VoIP, video streaming (VoD) and HTTP were selected. The WFQ for both domains and for all traffic types performs better than CBQ, while the IntServ domain proves its superiority (for real-time services) against DiffServ, with the drawback being that of scalability (due to signalling overheads). Furthermore, both (WFQ & CBQ) algorithms were shown to be able to support the specified QoS contracts, in terms of throughput and delay requirements. The two algorithms can be potential candidates for supporting QoS in the next generation IP-based HiperLAN/2. Similarly, RSVP proved to perform well with HiperLAN/2, by improving significantly the performance of the whole system.

Finally, cross-layer and cross-protocol QoS issues were identified both on data as well as control planes, while the QoS framework was topped up by introducing congestion (based on the CBQ and WFQ traffic controllers) and connection admission (based on a resource-polling algorithm, according to the ETSI H2 standardisation) control algorithms.

# Chapter 5

## 5 Mobility Management Framework for IP-Based WLAN

### 5.1 Introduction

One of the main requirements of B3G networks is the mobility provisioning in hot-spot areas and correct utilisation of the available protocols for the prompt and seamless delivery of the requested services to end-users. This chapter addresses these issues by selecting two dominant micro-mobility protocols, namely Cellular IP and Hierarchical Mobile IP, while their operational suitability to manage network mobility in IP-based Wireless LAN domains is evaluated.

Furthermore, in this chapter, we present an end system mobility solution that aims to provide seamless mobility in an IP-based WLAN environment. The proposal is built upon a domain-based approach with SIP (Session Initiation Protocol) providing support for terminal mobility and interworking with either a protocol like DHCP or the nominated micro-mobility solution to handle local mobility. We thus compare, test and evaluate two end-system mobility solutions, namely SIP/DHCP and SIP/CIP or HMIP. In the SIP/DHCP scheme, we have *configured the SIP protocol* to interoperate with the DHCP functions, whereas in the SIP/CIP or HMIP, CIP or HMIP functionalities have been extended to interwork with SIP-based components.

Finally and in order to complete the mobility management framework, we enhance the SIP/Cellular IP-based mobility management architecture by introducing the Mobile IP protocol (support for macro and global mobility). This “hybrid” configuration allows/provides mobility in two different layers namely the application and network layers and over heterogeneous networks (3G and WLANs), and thus our proposal is so-called multilayered, multi-protocol and hierarchical [EVO-D2.1]. The proposed architecture is evaluated in terms of mobility-related metrics and for two types of traffic, real time and non-real time.

## 5.2 Mobility Management in All-IP Networks

### 5.2.1 Basics on mobility management

As IP routing capability is being pushed towards the edge of networks, it is less cost effective to continue to support mobility management at layer 2 as in today's cellular networks. An integrated solution, whereby mobility management is an integral part of the IP layer routing protocol, would be more advantageous for both the operator and the end user. Adopting a common protocol, i.e. IP, will enable independence of the core and access network technologies and support a common solution for fixed and wireless network integration.

Mobility management for All-IP network infrastructures should provide means to handle several types of mobility, being the following:

- *Terminal mobility*: refers to an end user's ability to use her/his own terminal in any location and the ability of the network to maintain the user's ongoing communication as she/he roams across radio cells within the same subnet, subnets within the same administrative or different administrative domains.
- *Service Mobility*: refers to the end user's ability to maintain ongoing sessions and obtain services in a transparent manner. The service mobility includes the ability of the home service provider to either maintain control of the services it provides to the user in the visited network, or transfer their control to the visited network.
- *Personal Mobility*: refers to the ability of end users to originate and receive calls and access subscribed network services on any terminal in any location in a transparent manner, and the ability of the network to identify end users as they move across administrative domains.

Conventionally, mobility management consists of two components, namely *location management* and handoff or handover management.

Location Management is a two-stage process that enables the network to discover the current attachment point of the mobile user for call, or packet, delivery. The first stage is location update. In this stage, the MT periodically notifies the network of its new AP, allowing the network to revise the user's location profile. The second stage is paging. Here, the network is queried for the user location profile and the current position of the MT is identified.

Handoff management enables the network to maintain user's connection as the mobile terminal continues to move and change its AP to the network. The three-state process for handoff first involves initiation, where the user, the network or the changing network condition process itself,

identifies the need for handoff. Operations at this stage take place in the physical layer. The second stage is new connection generation, where the network must find new resources for the handoff and perform any additional routing operation. This involves link and network layers. The final stage is data-flow control, where the delivery of the data is diverted or extended, from the old path, to the new path.

Two levels of handoff may be identified<sup>1</sup>:

- *Cell handoff (or micro-mobility)*: It allows a MT to move from one AP to another in a subnet within an administrative domain.
- *Subnet handoff (or macro-mobility)*: It allows a MT to move from one AP within a subnet to an adjacent one within another subnet that belongs to the same or different (this is also called global mobility) administrative domains.

In general, the cell handoff is more frequent than subnet handoff. The handoff process is transparent to the MTs.

### 5.2.2 Review of the existing IP micro-mobility approaches

Traditionally, mobility management functions are implemented at the link layer. However, IP-based mobility procedures can take advantage of other network-layer facilities such as multicast, QoS guarantees support, security. In addition, mobility management at the network layer could enable seamless handoff among heterogeneous access networks. Mobile-IP has been optimised for macro-level mobility and relatively slow moving mobile hosts. It is not suitable for an environment where fast moving mobile terminals often migrate from one cell to another during active data transfer. Hence, *Mobile IP is not adequate for mobile users in a cellular network*, where the changes of serving AP could be of high frequency. Neither it is adequate for handling user mobility within a LAN or PAN network. Mobile IP needs to be complemented by another mechanism to ensure fast handoff(s), while it needs to be enhanced by more distributed mechanisms to provide mobility within a network domain. Mobile IP also lacks of real-time location tracking and the paging mechanism is not addressed at all. IETF Mobile IP does not differentiate between active and inactive mobiles. In a scenario where IETF Mobile IP is widely deployed, inactive users generate the same signalling traffic as active users, although it could be more efficient to locate them by searching (paging) from a database. In the case of QoS (e.g. RSVP) enabled networks, acquiring a new CoA upon a handoff triggers the establishment of new

---

<sup>1</sup> Note that the terms micro-mobility, intra-domain mobility and local mobility are equivalent as well as the terms macro-mobility, inter-domain mobility and global mobility are equivalent and can/will be used in an interchangeable manner for the remainder of this thesis.

QoS reservation from the HA to FA even though most of the old path remains unchanged. This further introduces more signalling overhead and latency for reservation restoration.

In this work, the term micro-mobility is referred to as the mobility where the user's routable IP address does not change. It is often used interchangeably with local mobility, where the mobility is within the same subnet. In other words, it is the ability of a mobile node to move without notifying its home agent. Macro-mobility is the movement across different domains of a network. The mobile node's home agent is notified of any macro-mobility movement. In MIP, a MT registers with its home agent each time it changes its point of attachment. Each location change requires signalling between the mobile user and its home network, which could possibly be far away and the latency of such an operation can be high. Therefore, MIP does not scale well to serve large number of mobile users moving frequently between small cells. Since mobility of users is mostly localised, location management and handoff algorithms should exploit the locality of user mobility by restricting the processing to the vicinity of a MT. This could dramatically reduce the handoff latency and the load in the core network.

Several micro-mobility management schemes have been proposed to augment MIP and to provide faster and smoother handoff. Hierarchical Mobile IP [SOL-00], Cellular IP [CAMP-00], HAWAII [RAM-00] and MER-TORA [RAM-00] are some of them. All these proposals agree that MIP is suitable to handle macro-mobility but not the micro-mobility. They generally exploit the locality of user mobility, and try to restrict handoff or location update execution to the vicinity of the mobile terminal. Local handoffs are managed locally and transparently to a MT's correspondent hosts and home agent, whereas global mobility could be managed by MIP (see Figure 5-1). This thus reduces handoff latency and the load in the Internet.

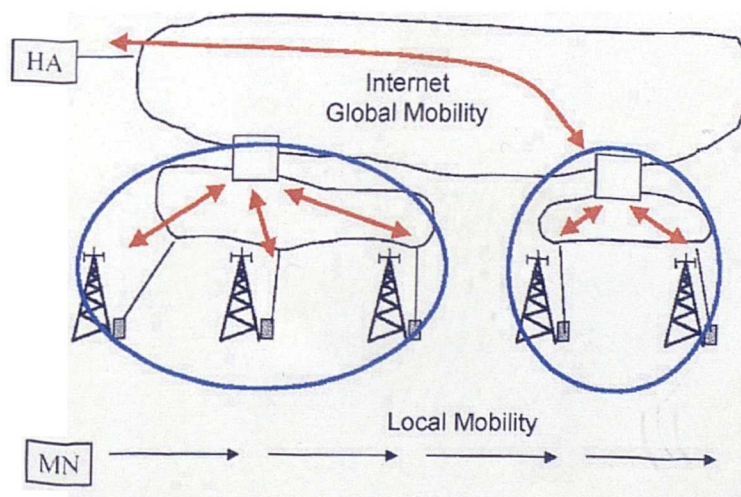


Figure 5-1: Hierarchical decomposition of mobility management.

Currently, the IETF Mobile IP WG [MOB-IP] is working on developing micro-mobility protocols, which are based on MIP (v4 and v6) [MIP-02]. The most representative proposal is the Hierarchical Mobile IP. However, the IETF Mobile IP WG was getting diluted with non-mobile IP micro-mobility protocols, such as CIP, HAWAII, EMA, etc. Hence, another working group, the Seamoby WG [SMBY] was formed to tackle the problem of seamless handoff - which is the basis of micro-mobility. Non-mobile IP protocols, such as HAWAII, CIP and EMA are shifted to for discussion in this working group.

### 5.2.3 State-of-the-Art of micro-mobility schemes

IP micro-mobility protocols assume that all mobile terminals access a network using IP routing to the access point. Seamless handoff is a central feature. A mechanism is required to ensure that as MT moves, a new path could be setup quickly, and the old path be deleted fast. Consequently, the packet loss and delivery delay due to MT's movement could be minimised. It generally requires maintaining per host location information and updating per host location information when hosts move. Furthermore, the scalability of the protocols is limited. Several proposals address the problems of handoff latency by introducing hierarchical architectures and separate inter-domain (macro-) from intra-domain (micro-) mobility. *Mobile IP with Route Optimisation* extension, *Hierarchical Mobile IP with Regional Tunnel Management*, *Universal Mobile IP*, *Mobile Enhanced Routing* protocol etc. are some of them. These proposals however do not address location management efficiently. *Cellular IP* and *HAWAII* include not only handoff solutions but also location management with paging capabilities. Hence, they are more suitable for cellular networks. Both CIP and HAWAII are distributed micro-mobility management schemes. The IP address of mobile user remains unchanged while changing its point of attachment. Nevertheless, both CIP and HAWAII can be integrated with MIP for supporting global mobility.

For any micro-mobility management scheme, cellular-network-like signalling (for registration, location updating and paging) transportable by IP is required. In addition, an efficient handoff is required for real time services to ensure low latency in packet delivery and minimal packet loss. Multicasting, buffering and other similar techniques could be exploited for this purpose. Most of the proposals require network nodes that handle user mobility, to be connected in a hierarchical manner. Some (e.g. Cellular IP, HMIP) even require a tree hierarchy. The information needed to route or to forward packets to a MT is distributed among these network nodes. The root of the hierarchy supports Mobile IP to handle global mobility, while the leaves handle mainly local mobility. Hierarchical mobility management schemes are not new but had been proposed to reduce signalling load in connection-oriented networks. GSM uses HLR and VLR for user registration and tracking. Gustafsson [GUS-00] proposed using an anchor VLR for reducing the location updates latency. Caceres's work in [CAC-98] demonstrated that hierarchical topology is

very effective in supporting fast and scalable handoff for real time applications in WLANs. This chapter does not intend to extensively describe all the proposals appeared in IETF Mobile IP and Seamless Mobility (SeaMoby) WGs. Instead, we categorize some of the most popular protocols, summarizing their main features and making a useful comparison for selecting the best suited for IP-based WLANs.

### 5.2.3.1 Micro-mobility Categories

Micro-mobility proposals can be broadly classified into three categories:

1. **Hierarchical Tunnelling**
2. **Host-specific Forwarding**
3. **Mobile-Enhanced Routing** (Ad-hoc networks)

The last one is not part of this research.

### 5.2.3.2 Hierarchical tunnelling

Hierarchical tunnelling schemes are basically extension of Mobile IP. It is characterized by their reliance on a tree-like structure of FAs. Encapsulated traffic from the HA is delivered to a root FA, which decapsulates it and re-encapsulates it for one of its children FAs based on the MT's current point of attachment (Figure 5-2). As the MT moves from leaf to leaf in the tree, location updates are made to the least common ancestor node so that traffic is always tunnelled properly (Figure 5-3). As a MT moves between different access points, location updates are made at the optimal point on the tree, tunnelling traffic to the new access point. By keeping most location updates within the local access network, the cross-Internet signalling and latency are reduced substantially. These proposals sometimes require the MT to send new types of messages or to be otherwise aware that a hierarchical tunnelling protocol is in use.

Examples of micro-mobility protocols that use hierarchical tunnelling include *Mobile IP with Hierarchical Foreign Agents* [PERK-96], *Mobile IP Regional Tunnelling* [GUS-00], *Hierarchical Mobile IPv6* [SOL-00], *Mobile IP with fast handoff* [MAL-00]; *Mobile IP with proactive handoff* [CALH-00], which requires network controlled handoff and was designed with cellular network in mind, and *TeleMIP* [DAS-00] which is also closely related to cellular network.

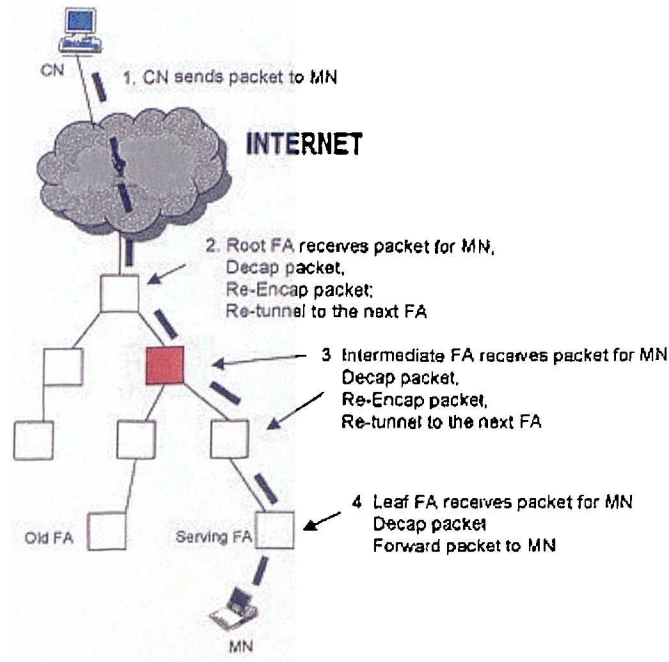


Figure 5-2: Packet delivery under Hierarchical Tunnelling schemes.

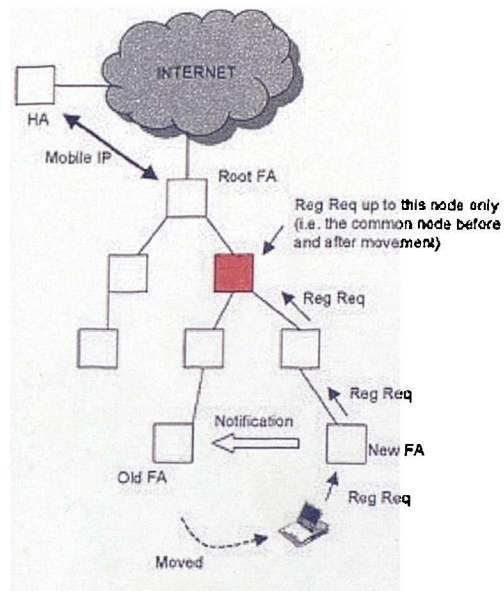


Figure 5-3: Mobile terminal executes registration with FAs.

### 5.2.3.3 Host-specific forwarding

The second category of micro-mobility protocols takes up *routing update* approaches. The overhead introduced by de-capsulation and re-encapsulation of traffic at each FA, as in the case with hierarchical tunnelling approaches, is avoided. In these proposals, host-based forwarding (instead of the conventional network-id-based routing) is used to route packets toward the mobile's point of attachment. The forwarding database is distributed among different network

nodes, which are hierarchically connected. These schemes typically introduce implicit (e.g., based on snooping data) or explicit signalling to update host-specific forwarding table. Mobility signalling messages trigger the creation or deletion of host-specific forwarding entries in the IP routing table of mobility-enabled nodes. These proposals usually require the use of a collocated care-of address, which will be used as the basis for forwarding packets.

MTs attached to a local network use either (a) the IP address of the gateway (which is topologically equivalent to the root FA in Hierarchical Tunnelling method) as their MIP CoA or (b) obtain a co-located care-of address. The gateway forwards packets toward MT's point of attachment. Within the local network, MTs are identified by their home address and data packets are routed using host-specific forwarding without tunnelling or address conversion. Examples of micro-mobility protocols that use mobile-specific routing include CIP and HAWAII.

#### **5.2.3.4 The comparison**

The previous section described the general operation of different categories of micro-mobility management protocols. Here, a summary of their characteristics is listed in Table 5-1 merely for comparison purposes. Micro-mobility protocols based on Hierarchical Tunnelling and Host-specific Forwarding are fundamentally similar in principle. Both default to MIP as the macro-mobility protocol, could be configured to support paging, require a hierarchical topology in tree form and maintain a host-specific entry for MTs that move. The difference is mainly on the technical (software) implementation, e.g. soft state expiry or explicit signalling to delete mobility management state, ways of defining paging area, packet snooping or explicit signalling to create mobility management state, end point of signalling (at the cross-over router or at the old AP).

We believe that the difference of all these schemes might not be significant in terms of HO performance, particular in the speed of handoff completion and number of packet losses. The choice of adoption should be based on other issues such as flexibility, scalability, ease of supporting QoS, efficiency in address handling, and fault tolerance, ease of configuration and management.

More information on the location management and paging can be found:

- Cellular IP [ZAC-00]
- Hierarchical Mobile IP [SARI-00], [HAV-00]

While the routing strategy is described in detail:

- Cellular IP [ZAC-00]
- Hierarchical Mobile IP [SOL-00]

Micro-mobility Protocols Characteristics / Properties	Hierarchical Tunnelling	Host Specific Forwarding
Micro-mobility protocols (instances)	<ul style="list-style-type: none"> <li>• MIP Regional Tunnelling</li> <li>• Hierarchical MIP</li> <li>• TeleMIP, etc</li> </ul>	<ul style="list-style-type: none"> <li>• HAWAII</li> <li>• Cellular IP, etc</li> </ul>
MIP as macro-mobility protocol	Yes	Yes
Paging support	Yes <sup>2</sup>	Yes
Support prefixed-based routing	No	No
Change CoA during HO	Yes	No
Packet redirection during HO	Tunnelling	Host-based forwarding
Type of updating message	Explicit signalling with MIP extension	Explicit signalling or data packet snooping
Topology	Hierarchical, Tree	Hierarchical, Tree
Scalability	Poor	Poor
Reliability	Average	Average/Poor

Table 5-1: Comparison of micro-mobility schemes [EVO-D2.2].

A characterisation and comparison of the two aforementioned micro-mobility categories (i.e., hierarchical tunnelling and host specific forwarding) is outlined in Table 5-1. While, Table 5-2 provides a list for comparing mobility protocols including discovery, paging support, location update and handover. These protocols are Cellular IP, Hierarchical Mobile IP. The table also contains some proposals on alternative solutions.

Micro-mobility Protocols / Properties	Cellular IP	HMIP	Desired Solution
Discovery Mechanism	Layer 3 beaconing	MIP Agent Advertisement	Not MIP Advertisement
Paging Area Configuration	Static configuration (Paging area set in advance)	Possible	Dynamic
Paging Method	Broadcast in paging area, or whole network	IPv6 multicast in paging area	Possible
Paging Cache Placement	Located in selected paging nodes	Absolutely centralised	Distributed
Time Based Paging Update Initiated by Mobile Terminal?	Yes	In response to paging router advertisement	Not absolutely necessary
Location Update	On crossing paging area only	Registration on paging area crossing	Possible
Paging Cache Update	By all uplink update	By regional binding	Necessary

<sup>2</sup> [SARI-00] proposes that a paging mechanism could be incorporated into Hierarchical Mobile IP.

Mechanism	packets and data packets	update	
Fast Handoff Support	Yes	Yes	Mandate
Multicasting for IP Traffic Delivery	No	No	Possible
IP Multicasting for Handover Packet Recovery	No, only packets duplication during semi-soft handover	No	Possible
Soft-State Support for Routing Entries	Yes	Yes	Flexible
Expansion of the Network in the Internet	Very easy	Easy	Very easy
Interoperability with MIP?	Inter-network, almost completely	Yes, completely	Possible

Table 5-2: Comparison of IP-based mobility protocols.

Finally, Figure 5-4 provides a chart with the classification of different mobility management protocols and proposals.

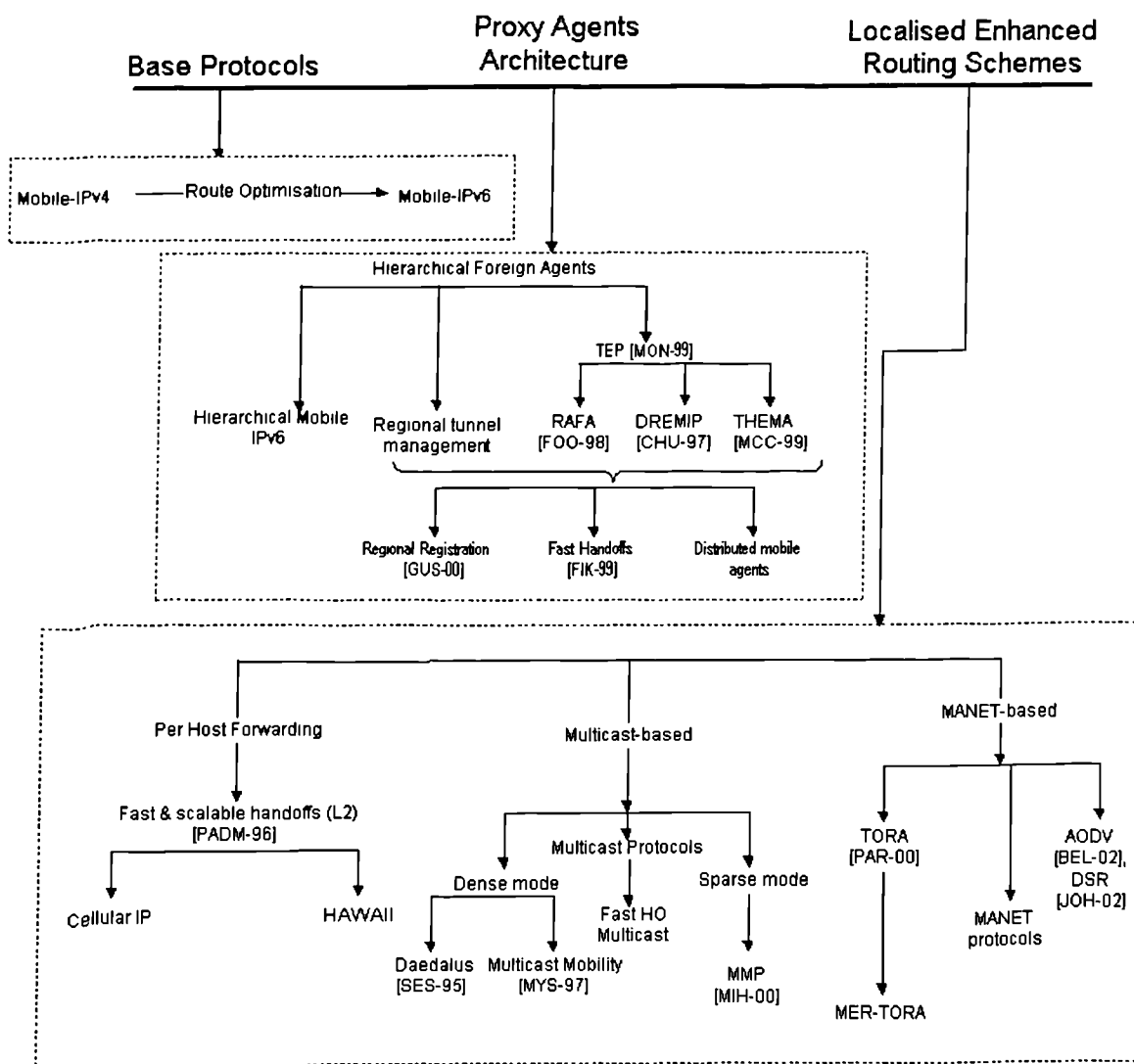


Figure 5-4: Classification of IP mobility protocols.

## **5.3 Multilayer Mobility Management**

### **5.3.1 Differentiation between global and local mobility**

Mobile IP is optimised for macro-level mobility, i.e. mobile user roams from one network to another. It is not suitable for an environment where highly MTs often move from one cell to another during active data transfer. Hence, Mobile IP is not adequate for mobile users in a cellular, LAN network, where the changes of serving AP could be of high frequency. Mobile IP must be complemented by another mechanism to ensure fast handoff. Therefore, Mobile IP could be used to make sure that packets for mobile terminals are routable to the network (or subnet), which it attaches to. While, micro-mobility protocols are needed to handle the handoff. The demand for a unified solution for global and local mobility is a misleading requirement. Such an approach would not provide the best solution to either of the problems.

### **5.3.2 Mobility management in All-IP networks: Protocols and architectures**

The starting point for the design of efficient mobility management architecture for next-generation all-IP networks is with Mobile IP, an IETF proposed standard. Indeed, this work is being looked at within the IETF Mobile IP and Seamoby WGs. Mobile IP provides a network layer solution to node mobility across IP networks. While roaming, a MT maintains two IP addresses, a permanent home address used in all transport layer connections, and a topologically correct care-of address that reflects the current point of attachment. The CoA is obtained either through a Foreign Agent or auto-configuration process (i.e. is co-located in the MT). While at home, the MT uses its permanent home address. A location register on the home subnet, referred to as Home Agent, maintains a mobility binding that maps the MT's home address to a CoA. The HA acts as a proxy on the home subnet, attracting packets addressed to the MT, and employs tunnelling to re-direct packets to the MT's CoA. MTs send registration requests to inform the HA of any change in CoA or to renew a mobility binding. Mobile IP provides an elegant solution for node mobility when the MT moves infrequently. However, Mobile IP introduces significant latency when used in a cellular environment simply because handoffs occur frequently and registration messages may travel large distances before packet redirection occurs. Additionally Mobile IP suffers from indirect communications, which increases the delay and causes an overhead due to tunnelling, which decreases the bandwidth utilization. This is mainly critical for real-time communications.

The Session Initiation Protocol [ROS-01] is an emerging protocol, designed to provide basic call control and application-layer signalling for voice and multimedia sessions in a packet-switched

network. Several wireless technical forums (e.g., 3GPP, 3GPP2, SIP forum) have agreed upon SIP utilization to provide session management as a means to provide personal and service mobility. Furthermore, [WED-99] has proposed to extend SIP with functions that would support terminal mobility, alleviating some of the shortcomings associated with Mobile IP and its route optimisation variants [MIP-RO]. The author argues that for real-time traffic over IP, which is mostly RTP, there is a need for fast handover, low latency and high bandwidth utilisation and therefore SIP should be considered as an alternative mobility solution to Mobile IP and its variants.

The *Extended SIP mobility* approach (Figure 5-5) introduces mobility awareness at a higher layer than the network layer. SIP already supports user mobility and the approach is to extend SIP as an application layer signalling protocol, in order to support *end system mobility*. The main assumption behind the *Extended SIP mobility* approach is, that a mobile user is identified by a unique address (e.g. callee/caller@example.com). This unique address is mapped to the current IP address of the mobile user's end system. No explicit home IP address is required. SIP introduces a SIP agent at the user's side and a SIP server (SIP redirect server or SIP proxy server) and location server to the network infrastructure. The *user* mobility is supported as follows: Suppose a user wishes to initiate a session, an invitation is directed to the SIP server, which in turn queries the location server for the current IP address of the mobile user's end system. The SIP server sends the invitation to the calling user. The invitation contains the IP address of the callee (called). If the mobile user moves, the location server is updated and new sessions will be set up to that new IP address. *End system* mobility to that scheme is mainly understood as an increased roaming frequency and as a change of an IP address during an ongoing session. Assuming that a session is already established, then the mobile registers with the location server the new temporary address and the mobile re-invites the correspondent host with the same session identifier and the new temporary address (in the *Contact field of the SIP message*). The session can be continued, although the IP address has changed.

It is important to note that SIP does not support TCP. Therefore, *Extended SIP mobility* applies only for real-time communications over UDP, as it breaks TCP connections. Supporting terminal mobility for TCP with SIP [VAK-01], requires a tracking agent on the MT to maintain a record of its ongoing TCP connections, as well as, IP encapsulation capabilities on each correspondent host.

Although Mobile IP and SIP could work complementary to support inter-domain mobility (mobility across different administrative domains), they are both unsuitable for handling intra-domain mobility (micro-mobility). Intra-domain mobility must either be supported by lower layers, or by some other more efficient solutions. In this research, we propose such a domain-based approach to support mobility for multimedia applications, by using an IP-based micro-mobility solution (e.g., HMIP, Cellular IP) to handle the local mobility and support fast handoffs

and paging. While SIP or Mobile IP will support inter-domain mobility for all types of traffic from/towards the MT.

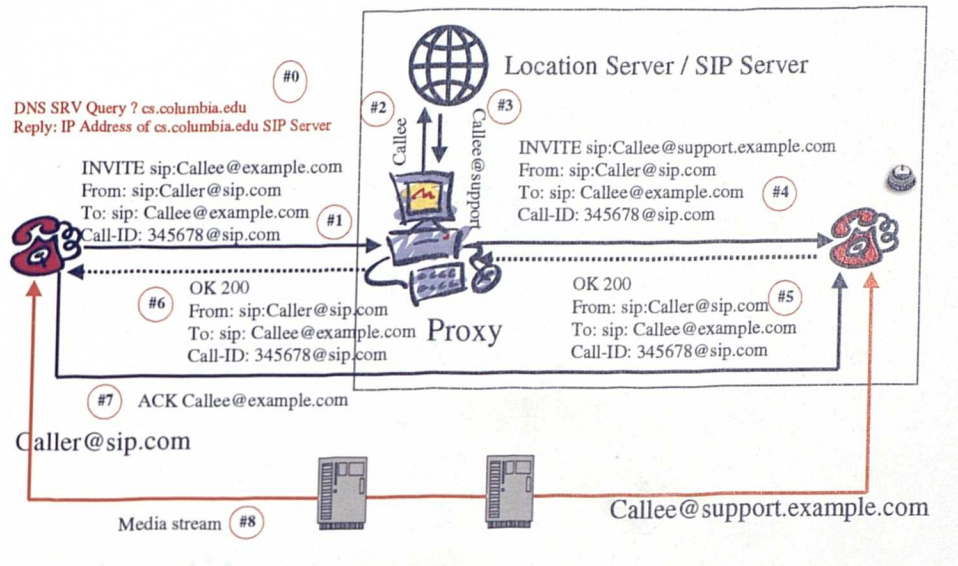


Figure 5-5: Session establishment with the extended-SIP approach.

### 5.3.3 SIP-based approach for mobility management

This approach uses SIP signalling to support macro-mobility, while fast handoffs and paging are supported by an IP-based micro-mobility solution.

Figure 5-6 illustrates this hierarchical scheme, which *handles mobility of hosts in two layers*. The MTs are identified with their private home addresses within the micro-mobility areas. This provides location privacy and application transparency, while the mobile terminal roams within a visited domain.

A micro-mobility protocol (CIP or HMIP) could be extended with functions that enable its smooth interworking with SIP. More specifically, we have defined a new type of control information in the micro-mobility protocol route/binding/paging-update packets, being the SIP user identifier. This is an email-like address of the form “user@host”, where “user” is a user name and “host” is a domain name or numerical address. This information is inserted in the payload of the first route/binding-update packet after handoff and may be repeated in a few subsequent route/binding/paging-update packets for reliability. Upon receiving the first route/binding-update packet, the micro-mobility gateway (“green” gateway in

Figure 5-6) performs admission control to register the CH (Correspondent Host) in its caches (Local Registration) [EVO-D2.1].

If the MT will move during a session, the SIP User Agent sends a SIP re-INVITE request message to each one of its CHs. In this message, the MT includes its original SIP user identifier into the `from` field of the SIP header. It also includes the GW's (Gateway) address into [GAT-02]:

- the `Contact` field of the SIP header, in order to inform the corresponding host where does it want to receive future SIP messages, and
- the `c` (connection) field of the SDP (Session Description Protocol) header [HAN-01] that contains a description of the session, in order to redirect the data traffic flow towards its new location.

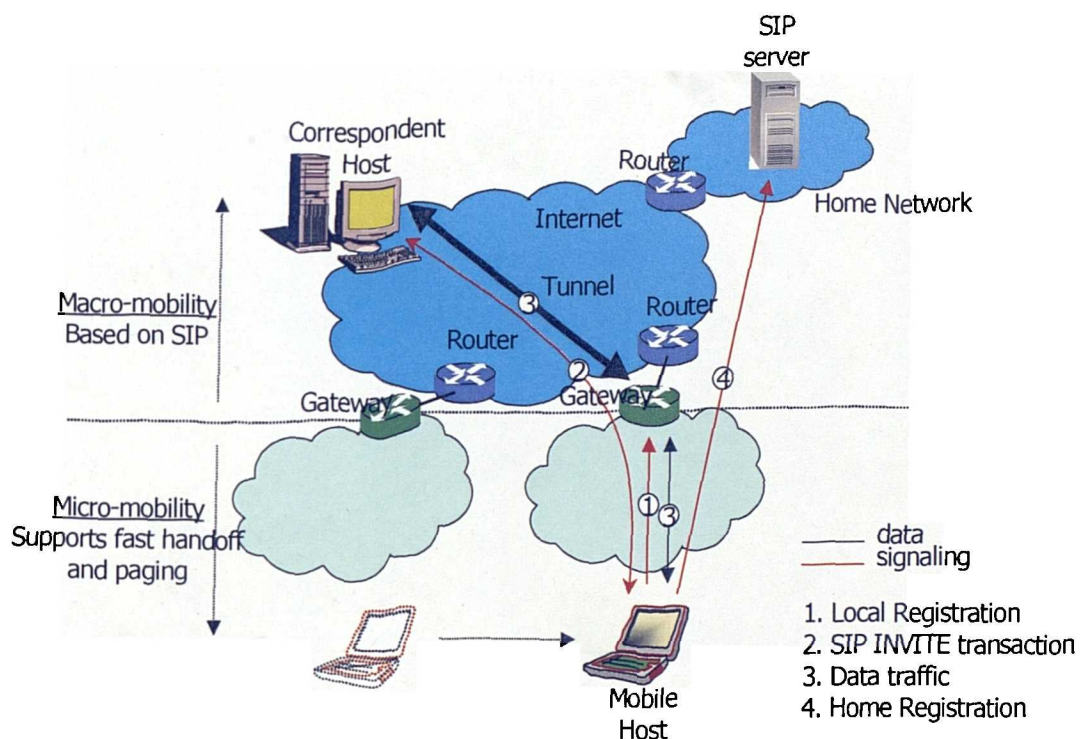


Figure 5-6: SIP-based approach for mobility management.

The micro-mobility GW will thus act as the point where SIP responses will reach. To forward the SIP responses to their original destination, the GW is equipped with a SIP message-tracking agent. This agent checks the GW's binding caches to determine whether a SIP response message must be forwarded towards a registered MT and then forwards the message after modifying the destination IP address. This assures that the aforementioned SIP re-INVITE transaction is correctly completed upon handoff (see Figure 5-6 for SIP message exchanges {1-4} after handoff, prior to handoff the micro-mobility protocol is in operation). A corresponding host after

receiving the aforementioned SIP INVITE message uses IP encapsulation to forward data information to the MT. The encapsulated data packets are captured by the GW, which in turn decapsulates and forwards them to the recipient MT following micro-mobility routing. Data traffic from the MT is regularly routed without the use of tunnelling. The MT completes the handoff by sending a SIP REGISTER message towards a SIP server on its home network. The main drawback of this approach is that it requires IP encapsulation capabilities on each host, while Mobile IP triangulation is avoided, since MH registers with SIP server after handing over to the new domain.

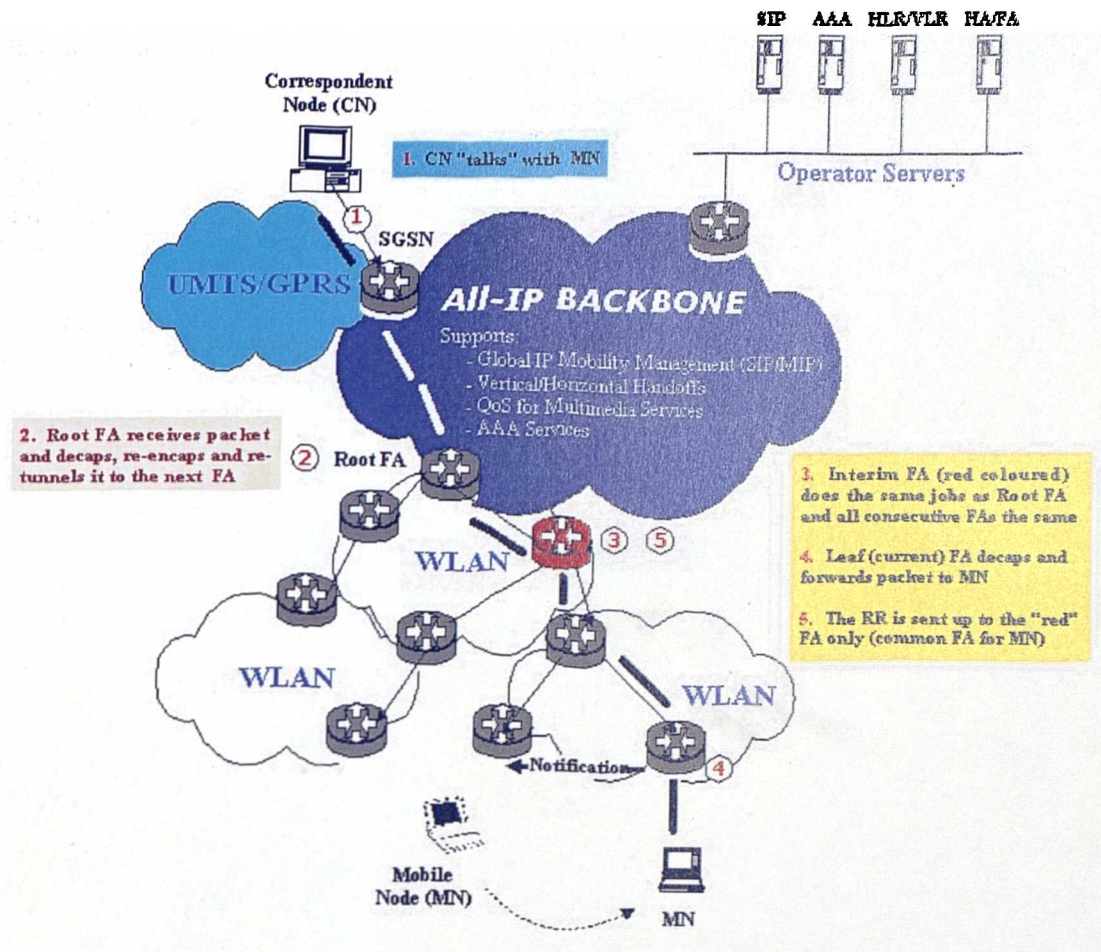
## **5.4 Simulation results and discussion for micro-mobility [CF -PO'02]**

In the following sections, we evaluate the performance of Cellular IP and Hierarchical Mobile IP protocols, with respect to reliability (handoff quality) and scalability (signalling overhead) metrics. The best performing protocol will be used to provide micro-mobility management support in All-IP environments (see Figure 5-7) and specifically in the WLAN domain. Then, a complete mobility suite will be proposed, where SIP will provide application layer mobility and location-based services, while MIP will define the global mobility domain.

The simulator was implemented using OPNET. Both the simulation models are based on the description on the protocols [ZAC-00], [SARI-00]. Paging functions for HMIP is incorporated based on [SOL-00]. Figure 5-7 shows the network level view of the simulator. It consists of a home agent, Internet cloud, 1 gateway and 16 access routers, and a number of MTs. The traffic source nodes represent the correspondent nodes of MTs. Access routers are grouped into 4 paging areas, i.e. PA1-PA4.

The analysis focuses on data packet transmission and mobility management functionalities. Two versions of traffic model are defined. The first is the constant bit rate source, which generates 1024 bytes packets at 10 ms interval. The second is an on/off traffic source, with exponentially distributed on time with mean value of 30 seconds, during which data packets are generated, and exponentially distributed off time with mean value of 90 seconds. The on/off model is used so that a MT would have chances to be in both active and idle mode. All traffic is directed from correspondent node to MTs (for more details on the traffic characteristics refer to Chapter 3, sections 3.6.1 and 3.6.4).

The mobility model in our simulations is the random waypoint model, unless otherwise stated. In the beginning, nodes are placed randomly in the simulation area. Afterwards, each node selects a



**Figure 5-7: The simulated network topology.**

destination location in a random fashion and moves towards it with a velocity selected from a predefined range. Once the destination point is reached, the node stops there for a pause time of exponentially distributed value with mean of 60 seconds. After that the procedure is repeated throughout the simulation (note: all handoffs are hard handoffs).

An ideal wireless interface was used. Hence, packets transmitted over the wireless interface encounter no bit error or loss. All routers are assumed to have an unlimited buffer size. As such, the only reason for packet loss is merely due to interruption during handoff. All links are of 10 Mb/s, with delay of 5 ms, whereas the effective data rate of wireless link is 1.5 Mb/s.

Our simulation studies were carried out on several performance measures, some of which are defined as following:

- **Host Specific Routing Entry Size** – Number of MT-specific routing, or forwarding, entries established in the network for data delivery to the MT. This is referring to the size of router cache for the case of Cellular IP; and the size of active binding cache plus the number of collocated addresses for the case of HMIP.

- Packet Loss Ratio – Ratio of total number of packets lost to the total number of packets transmitted.

### **5.4.1 Handoff Performance**

We first present the simulation results for the basic handoff performance for each micro-mobility protocol. During the simulation, one MT is constantly moving among neighbouring access routers at a constant speed, while receiving data packets constantly sent by its CN. Handoffs take place as MT crosses different coverage areas. Route update (for the case of CIP) and binding update (for the case of HMIP) procedures are executed by MT. The average number of packets lost during handoff for each protocol is recorded. The simulations were performed for different values of speed. Figure 5-8 shows the packet lost ratios for both protocols. Data is recorded from hundreds of independent handoff events.

Our first observation is that CIP gives better handoff performance than HMIP. The total number of packet lost, and hence the packet loss ratio, are higher for the case of HMIP. This is because CIP handoff delay is related to the packet delay between the access router and the closest cross over router. Whereas in HMIP, as binding updates packets are always destined for gateway, the handoff delay is higher. HMIP therefore does not benefit from the fact that a crossover router is topologically closer to the MT.

Next, we set a more realistic scenario for our simulations. Instead of continuous movement, the random waypoint mobility model is adopted, whereby MT pauses after moving for certain distance. In addition, instead of constant packet source, an on/off traffic source is used. In this scenario, a concrete mobility management operation comes into play. This includes paging update and paging for MTs that are in idle state. Various performance metrics were recorded. Figure 5-9 shows the packet lost ratio for both protocols. In contrast to Figure 5-8, the handoff performance of HMIP is better than that of CIP. Closer studies reveal that the higher amount of packet lost in Cellular IP is due to the higher queuing delay that exists between mobile terminal and the access router. In Cellular IP, the control overhead is higher because the frequency of sending route update packets (as well as paging update packets) is higher. Furthermore, when an access router that has not kept a Paging Cache receives a data packet, and if the destination is not found in its Route Cache, the packet is broadcasted on all its interfaces. As multiple data packets, destined to the same MT, arrive at the same router before the required forwarding entry is established (in other words, before the destined MT sends route update packets to the router), the data packets are broadcasted too. This results in multiple broadcasts, which in turn create higher traffic load in the lower layers, and hence higher queuing delay. Since the delay created by this phenomenon in

Cellular IP is higher than the packet delay for sending binding update(s) to the gateway in HMIP, the handoff performance has become worse.

**Set of results A – Handoff performance, one user (CBR traffic source):**

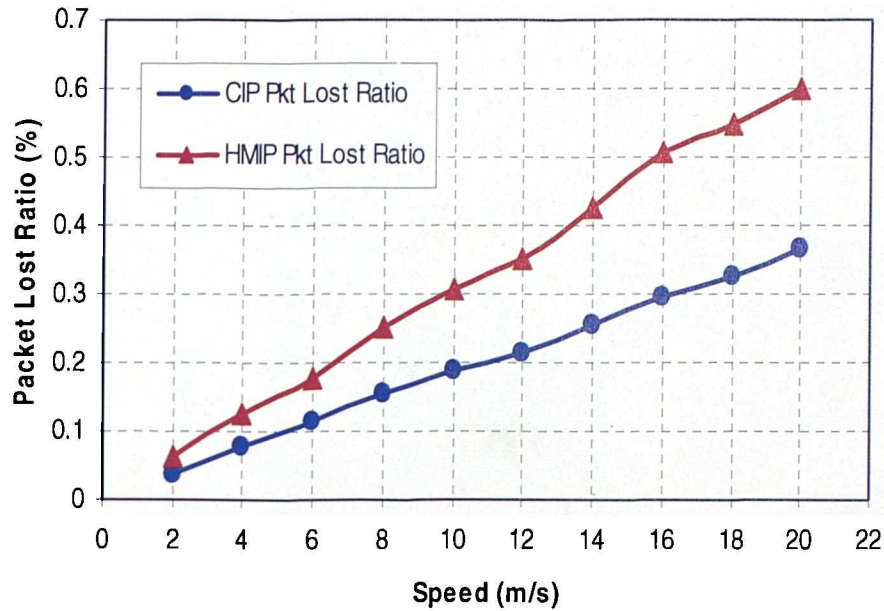


Figure 5-8: Packet lost ratio due to handoff (CBR traffic source).

**Set of results B – Handoff performance, many users (on/off traffic source):**

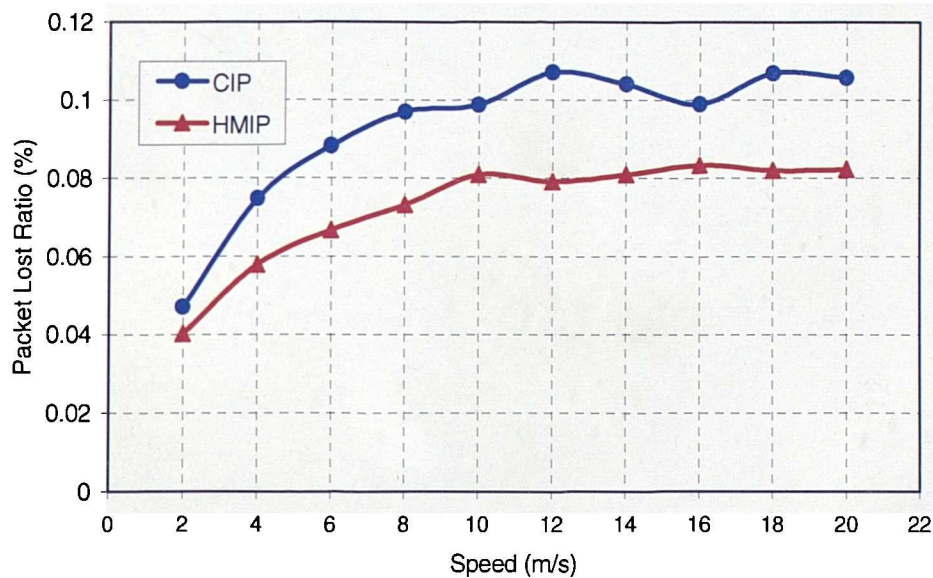


Figure 5-9: Packet lost ratio due to handoff (on/off traffic source).

### 5.4.2 Overhead

In the next instance, we investigate the overhead aspect of the protocols (in terms of scalability). The total number of control packets (i.e. route update, page update and paging packets) generated by both protocols is recorded. In addition, the number of host-specific forwarding entries installed in GW and different access routers is also recorded. On/off traffic sources are used in this set of study. Simulations were run for different node density ranging from 16 to 80 MTs.

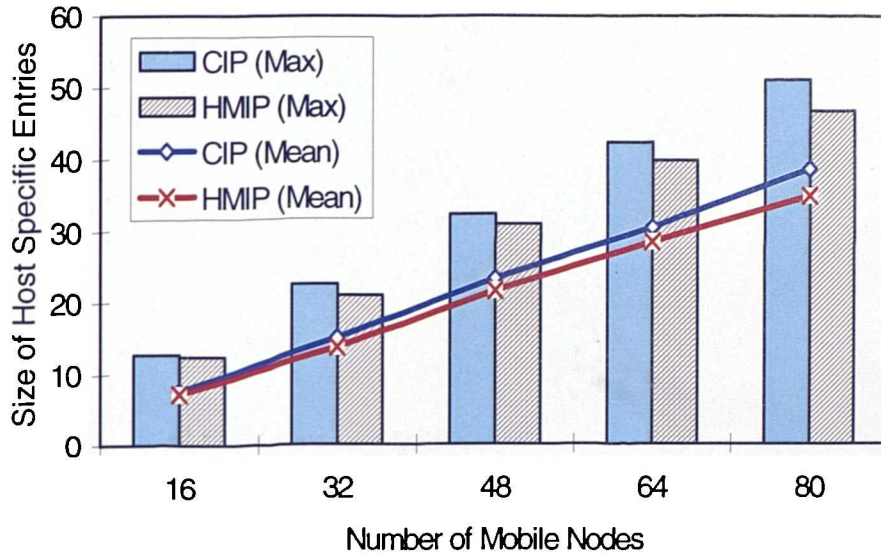


Figure 5-10: Size of host-specific entry created in gateway.

Figure 5-10 and Figure 5-11 show the average and maximum size of host-specific forwarding entries created in the gateway and access routers respectively. We can observe that for both Cellular IP and Hierarchical MIP, the entries size in GW is similar. This is because for both protocols, some sort of host-specific forwarding information for active MTs is kept in the gateway. This is in the form of Mobile IP binding cache for the case of HMIP, and routing cache for CIP. The number of entries for HMIP is consistently lower than that of Cellular IP, due to the fact that routing entries in CIP are removed only when the associated timer expires. Whereas in HMIP, binding update for a MT is removed immediately when a paging update packet (sent by a MT which is moving into idle mode) is received by the gateway. Nevertheless, the difference in both cases is insignificant. However, the entry size of AR is significantly higher in CIP than in HMIP, as shown in Figure 5-11. This is because, when a MT is in active mode, a forwarding entry is maintained in all routers connected along the path towards the gateway. In contrast, for HMIP, only two nodes, i.e. the gateway and the serving AR maintain routing-related information for an active MT. An entry is created in the Mobile IP binding cache in the GW; another one is created in the serving access router in the form of co-located IP address. This demonstrates that Cellular IP is far less scalable than HMIP.

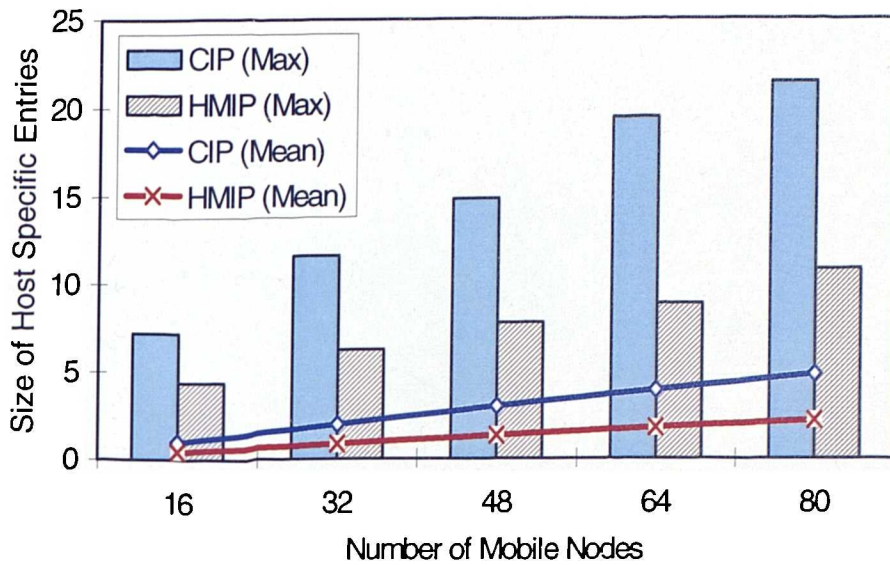
**Set of results C – Overhead:**

Figure 5-11: Size of host-specific entry created in access routers.

**5.4.3 Proposed Micro-mobility Protocol for IP-based WLAN Systems**

The study and further evaluation of these two micro-mobility protocols, as a part of an IP-based WLAN architecture, has demonstrated that Hierarchical Mobile IP performs better than Cellular IP under certain circumstances and scenarios (non-real time traffic, load conditions, signalling/overhead effect, etc). However, Figure 5-8 indicates that CIP could perform better, when the traffic conditions and the seamless fashion of services and applications is the system design target (multimedia applications). Therefore, in a mobility framework for IP-based WLAN infrastructures, both solutions (i.e. Cellular IP and Hierarchical Mobile IP) should be considered. The ultimate selection, it is more a philosophical and polymorphous antagonism rather than a trade off between scalability and performance. Consecutively, the rest of this chapter sheds some light on several other mobility concepts, as we try to propose a complete mobility suite for the next generation All-IP Wireless LANs.

**5.5 SIP/Cellular IP-based Mobility Solution**

The proposed mobility solution is based on the interworking of SIP with another protocol to support local mobility. For the latter, we have selected Cellular IP (despite both protocols behave in a similar manner this is purely a by-default selection). Furthermore, we briefly describe both

the SIP/DHCP and the SIP/CIP based solutions for mobility support, while in the following section; we discuss the overall architecture and the associated signalling in more detail.

As mentioned earlier, mobility support can be ensured by using SIP for macro-mobility together with lower layer mechanisms for handling local mobility. In the first scenario, SIP is used in conjunction with DHCP. The role of DHCP in this setup is to enable the MT to acquire a topologically correct IP address, when it moves within a micro-mobility domain. When the MT first attaches to an access point, it acquires an IP address via a DHCP server associated with the network. To initiate a session with a correspondent host, the MT sends an INVITE message and the normal SIP signalling procedure is performed to establish the session. Let's assume that the MT moves, while the session is ongoing and leaves the coverage area of the AP and moves into a neighbouring cell. The host then disassociates from the old AP and associates with the nearest AP acquiring a new IP address via DHCP. During this process, the ongoing session is disrupted. To re-establish the session, the MT sends another INVITE message maintaining the same Call-ID of the existing session but replacing the *contact* and *c* fields in the SIP and SDP headers with the new IP address. Finally, the UA sends a REGISTER message to the home SIP server to update the location information stored there.

Other IP-related solutions like DHCP can support local mobility, but they are not very efficient, since they can add significant latency during handoffs. On the other hand, Cellular IP supports fast handoff and paging. Hence, we can optimise the handover performance with the interworking of SIP and Cellular IP. In order to support that, we consider the same scenario as before. In this case, when a handover takes place, the MT does not need to acquire a new IP address and it can keep using its home address. If the new point of attachment is within the same CIP domain, the host only has to send a *route update message to the CIP GW*. On the other hand, when the handoff is between two CIP domains, first a route update is sent to the new gateway. Then, the SIP UA on the MT will send a re-INVITE message to each of its CHs. This message will contain the address of Cellular IP GW in the *contact* field and *c* field of the SDP header. Once again, the same Call-ID is maintained. Thus we assume that each domain GW will also have a SIP proxy server, which will act as the outbound proxy for the MTs attached to the domain. When the INVITE is received by the CH(s), it will reply back with a 200 OK message and the session(s) can be then resumed. As in the SIP/DHCP case, the handover is completed by sending a REGISTER message to the home SIP server to inform it about the current location of the UA. One implication of using the Cellular IP GW address is that all packets originated from the CH will be encapsulated. The gateway receives and decapsulates these packets and performs a local binding table look-up to route the packets to the MT. On the uplink, the MT sends packets without any encapsulation. The handoff is completed once the MT informs the home register of its current location.

### 5.5.1 SIP/DHCP Test-bed Environment

In this real-time experiment (Test-bed), all the routers have a built-in DHCP server, each of them acting as the access router of an IEEE 802.11b WLAN. In each WLAN subnet, L2 switches are used to connect a number of IEEE 802.11b APs to the ARs (Access Router). Several laptops act as mobile terminals, having been equipped with IEEE 802.11b wireless client, DHCP client and SIP user agent. In each WLAN, the access points use static IP addresses, while the MTs use a dynamic IP address obtained from the DHCP server.

A number of desktop computers are connected to the network via L2 switches, which are in turn connected to the IP core network via a L3 switch. The desktop computers are equipped with SIP user agents and act as wired corresponding hosts. With this configuration, once the MT enters the coverage of a new AP, DHCP allocates a new IP address to the MT, and SIP signalling is used to re-establish the communication between the MT and its CH. Figure 5-12 shows the overview of the system configuration for the case when DHCP is used with SIP to provide terminal and network mobility.

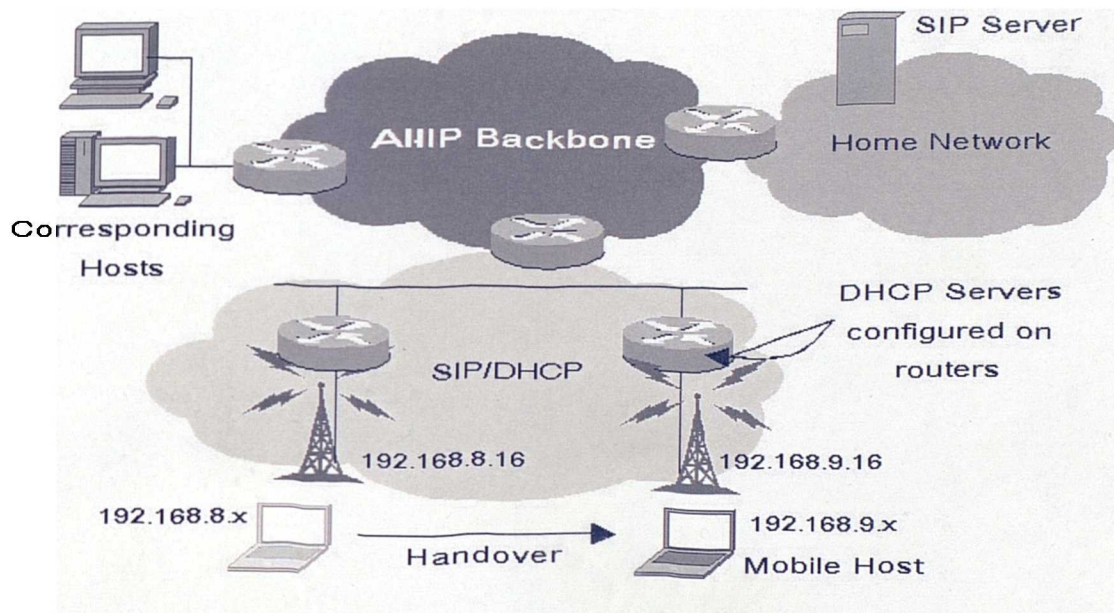


Figure 5-12: Subnet handover using SIP/DHCP.

### 5.5.2 SIP/Cellular IP Test-bed Environment

For the SIP/Cellular-IP approach the Test-bed is reconfigured as shown in Figure 5-13. For this set-up, we have replaced the two access points with two laptops, configured to behave as CIP

GWs. Similarly, we configured the MT to behave as a CIP mobile client. For these clients the implementation of Columbia University [COL-CIP] was used.

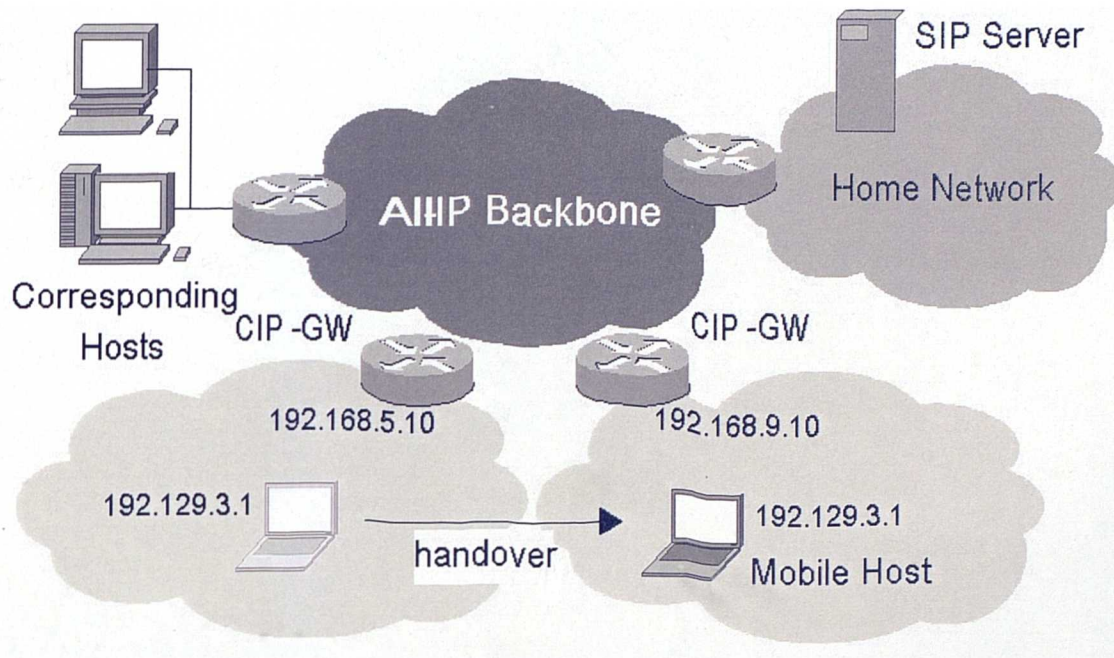


Figure 5-13: Subnet handover using SIP/CIP.

### 5.5.3 Handover Procedure using SIP/DHCP Scheme

The handover achieved in the Test-bed involves the following procedure and is shown in Figure 5-14:

- A session is established using SIP signalling between MT and CH.
- During the ongoing session the MT moves into a new subnet.
- Once the MT is out of coverage of its current access point and enters a new subnet the old path is lost.
- At the new subnet the MT acquires a new IP address from a DHCP server.
- Once the new IP address is obtained, the SIP user agent sends a re-INVITE message with the new IP address and sends it to CH maintaining the same call-ID. The re-INVITE message may be sent directly or via SIP servers to CH.

After the 200 OK and ACK message exchange, a new media path is established with the new IP address.

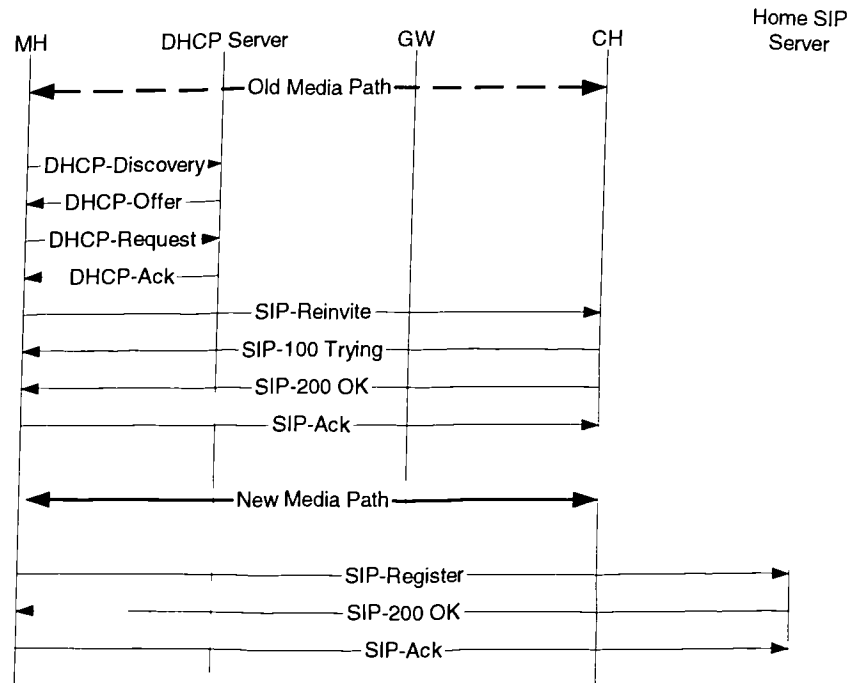


Figure 5-14: SIP/DHCP handover signalling.

#### 5.5.4 Handover Procedure using SIP/Cellular IP Scheme

For this case, the handover achieved in the Test-bed includes the following procedure and is shown in Figure 5-15:

- A session is established using SIP signalling between MT and CH.
- During the ongoing session the MT moves into a new subnet.
- Once the MT is out of coverage of its current AP and enters a new subnet, the old path is lost.
- At the new subnet the MT sends a *CIP-RU* packet to the *CIP-GW*.
- When entering a new subnet (i.e. new *CIP-GW*) the initiation of a *CIP-RU* is also used as a trigger for a SIP re-INVITE message.
- Once the CH receives the SIP-ACK message, the two hosts have all the information to continue their multimedia session over the new media path.

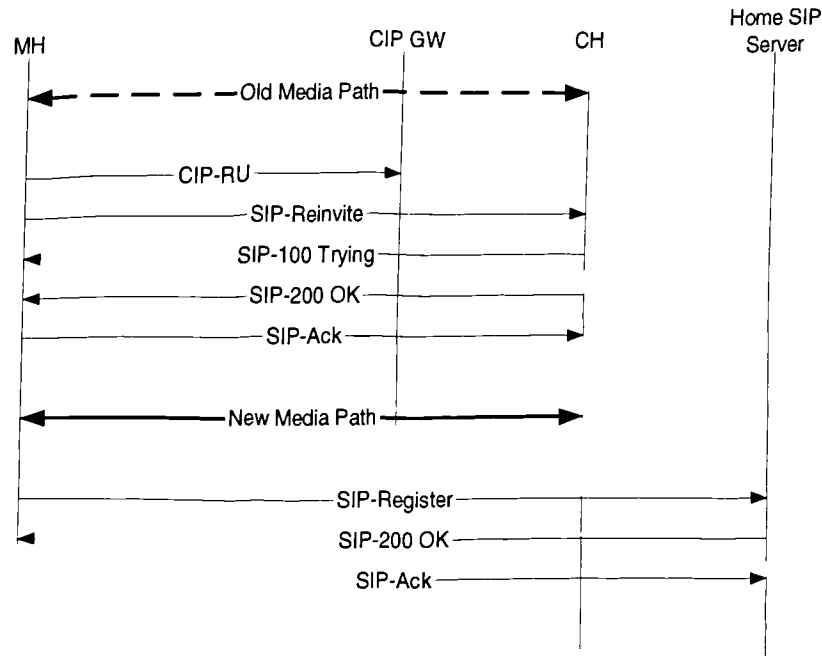


Figure 5-15: SIP/Cellular IP handover signalling.

### 5.5.5 Performance Evaluation and Discussion

The SIP terminal as well as the network mobility performance was determined by measuring the handover delay. We consider, the total handover delay to be the duration between loss of the old media path and establishment of the new media path. The old media path is lost once a MT executes a handover. In an inter-subnet handover scenario, the total handover delay may consist of the following components:

- Physical link and L2 handover duration (this duration is short enough to be ignored).
- SIP re-INVITE procedure delay. SIP re-INVITE procedure delay is defined as the duration between the time the new IP address is obtained and the time when the ACK is received by the CH.
- DHCP delay is defined as the duration between the time when a new physical link is established and the time when a new IP address is acquired.
- Cellular IP procedure delay is the time required by the route update packet to inform the CIP GW of the local handover. This may be ignored, if it overlaps with the SIP signalling.

### 5.5.5.1 Performance evaluation results

For both the SIP/DHCP and SIP/Cellular IP approaches, the delay of a number of handover attempts (HA) was measured. Table 5-3 shows the handover delay in seconds experienced by the mobile terminal in the SIP/DHCP and SIP/Cellular IP schemes respectively.

Protocols Used	Handover Delay (sec)					
	HA1	HA2	HA3	HA4	HA5	HA6
SIP/DHCP (DAA)	28.2	27.3	31.2	28.6	30.3	29.2
SIP/DHCP (SAA)	10.4	8.9	12.3	9.5	11.2	10.5
SIP/Cellular IP	0.55	0.54	0.53	0.55	0.54	0.55

Table 5-3: Handover delay for SIP/DHCP and SIP/CIP [AKH-POL].

For comparison purposes, in the SIP/DHCP scheme, measurements of the delay were also taken with the DHCP Server configured to provide a static address (i.e. provide the same IP address rather than selecting it from a pool of IP addresses as in dynamic allocation). The results show that the MT experiences an overall handover delay of about 30 sec using SIP/DHCP (DAA) and about 10 sec when using SIP/DHCP (SAA). In the SIP/CIP case, the handover delay was on average only 0.54 sec. The SIP re-INVITE procedure was also measured, showing variation between 0.51 and 0.53 sec. Table 5-4 shows a breakdown of the signalling involved during the SIP/Cellular IP handover for 5 different handover attempts.

The results have shown that on average the difference between 0.52 sec of the total 0.54 sec, is due to the SIP procedure delay.

Messages	HA1	HA2	HA3	HA4	HA5
CIP-RU	0.0000	0.0000	0.0000	0.0000	0.0000
SIP INVITE	0.0243	0.0243	0.0241	0.0244	0.0242
100 Trying	0.0323	0.0289	0.0319	0.0296	0.0314
SIP 200 OK	0.5482	0.5418	0.5347	0.5525	0.5352
SIP ACK	0.5498	0.5434	0.5362	0.5541	0.5367
SIP Delay	0.5255	0.5191	0.5249	0.5298	0.5254

Table 5-4: SIP delay in SIP/CIP [AKH-POL].

### **5.5.5.2 Discussion**

According to the test results, the handover delay was approximately 30 sec for SIP/DHCP (DAA), 10 sec for SIP/DHCP (SAA), and 0.54 sec for SIP/CIP. As shown in Table 5-4 the SIP procedure delay was also measured independently and was found to be on average 0.52 sec. This indicates that in the SIP/DHCP scheme, the DHCP procedure was the dominant part of the total handover delay. Hence, DHCP is unsuitable for supporting network mobility in WLANs. It was indicated in [DUT-01] that the standard DHCP procedure delay is about 15-20 sec with ARP checking. The difference in the delay might be due to: (a) the difference in the operating system of the end systems or (b) the difference in hardware configuration of the laptops, which could result in different processing capability. By configuring the DHCP server to provide static address allocation the overall delay was greatly reduced from 30 to about 10 sec. This difference is due to the fact that with the dynamic setting, the DHCP server needs to dynamically select a new IP address for the host whereas in the static case, a fixed IP address is allocated. In general the SIP/DHCP approach results indicated that the use of a L2 protocol like DHCP, is not a very efficient solution for supporting local mobility. Nevertheless, the replacement of the DHCP procedure(s) with a micro-mobility protocol, like Cellular IP is evangelised, since it optimises the handover performance. In the SIP/CIP scheme the MT maintains the same IP address avoiding the need for IP re-configuration and thus minimising the handover delay. In this scheme, the two major delay components are due to the Cellular IP signalling and the SIP signalling. Since, we have configured the MT to send a re-INVITE, when a Cellular IP route update packet is triggered; the SIP signalling procedure highly overlaps with the CIP procedure, thus causing insignificant additional delay to the overall handover delay.

## **5.6 Mobility management Framework for IP-based WLAN Systems**

### **5.6.1 Comparison of SIP-based and MIP-based mobility approaches [PO-CH'03]**

In this section, we present two mobility management configurations based on SIP (session, service and personal mobility) and Mobile IP (network mobility) protocols, which compliment the mobility framework proposed for local domains. These architectures mainly provide mobility in

two different layers application and network layers and due to that fact our proposal is so called multilayered.

The motivation for using the *extended SIP mobility* approach [WED-99] derives from its ability to support terminal mobility, alleviating some of the shortcomings associated with Mobile IP and its route optimisation variants. Figure 5-16 introduces mobility awareness at a higher layer than the network layer. It is important to note that SIP does not support TCP. Therefore, extended SIP mobility applies only for real-time communications over UDP, as it breaks TCP connections. Supporting terminal mobility for TCP with SIP requires a tracking agent on the MT to maintain a record of its ongoing TCP connections, as well as, IP encapsulation capabilities on each CH.

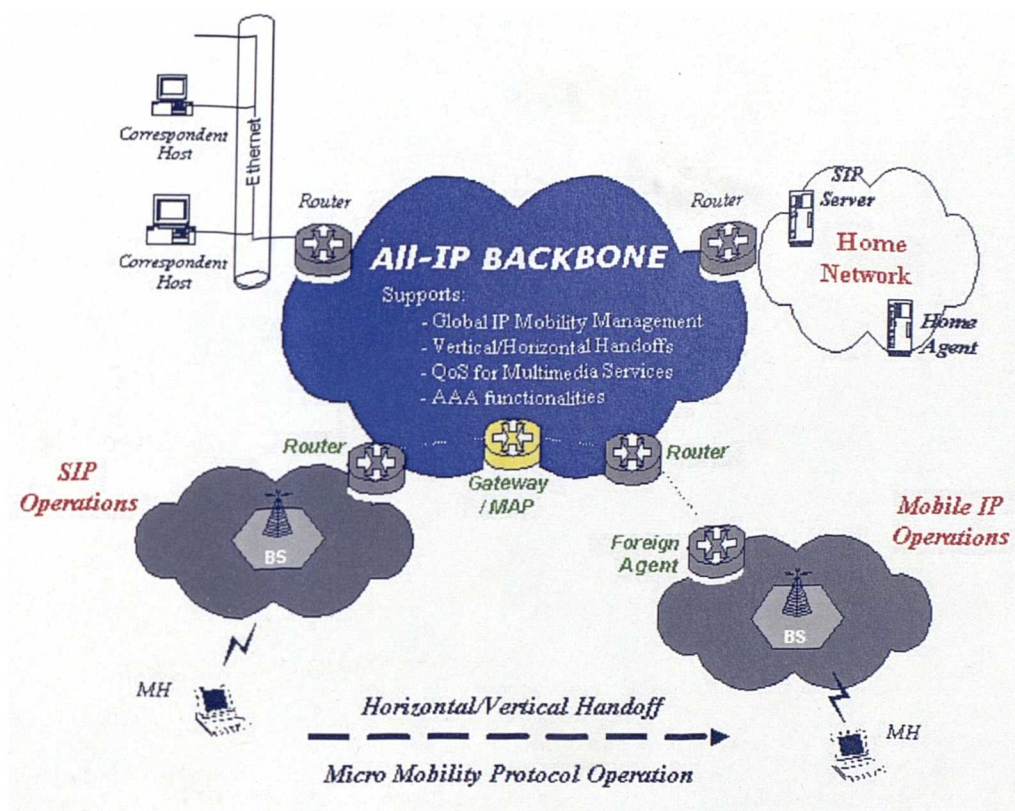


Figure 5-16: Network architecture for SIP and MIP approaches.

On the other hand, Mobile IP [MIP-02] provides a network layer solution to node mobility across IP networks. While roaming, a MT maintains two IP addresses, a permanent home address used in all transport layer connections, and a topologically correct CoA that reflects the current point of attachment.

Although MIP and SIP could work complementary to each other, as mentioned above, to support inter-domain mobility (mobility across different administrative domains), they are both unsuitable for handling intra-domain mobility. Therefore, the micro-mobility protocols of section 5.4.3 are considered for this job, while two approaches are nominated to complete the mobility “puzzle”.

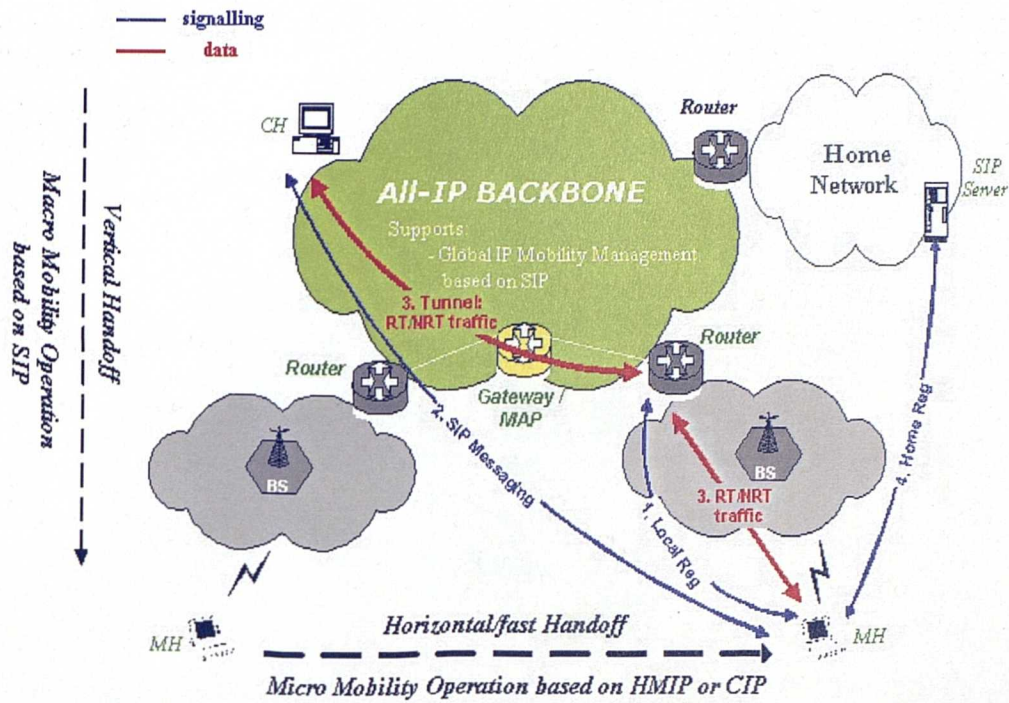


Figure 5-17: SIP-based approach for mobility management.

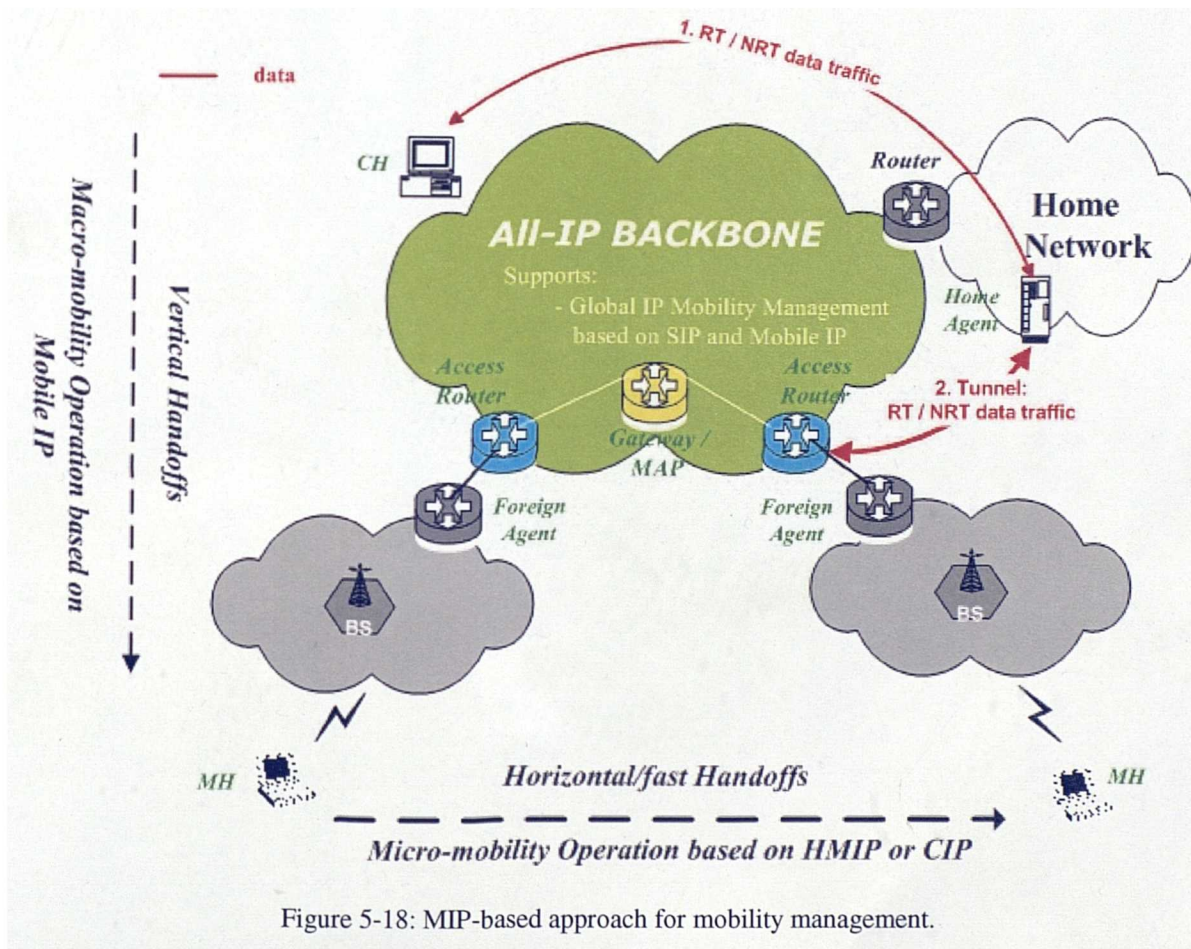


Figure 5-18: MIP-based approach for mobility management.

The first uses SIP in combination with IP encapsulation mechanisms on CHs to support mobility for all types of traffic from/towards the MT. On the other hand, the second approach employs MIP to support mobility for non-real-time traffic (mainly TCP-based applications) as well as real-time traffic. These two approaches are described in detail in [PO-CH'03] and schematically shown in Figure 5-17 (SIP-based approach) and Figure 5-18 (Mobile IP-based approach). Furthermore, in the Figure 5-19 (flow chart), the message sequence involves the DHCP operation, then the SIP messages exchange for supporting mobility and finally the tunnelling of data packets. In the Mobile IP approach, similar message exchanges are involved, as in the SIP-based approach, however the Mobile IP protocol is in operation now, as is shown in Figure 5-20 (flow chart).

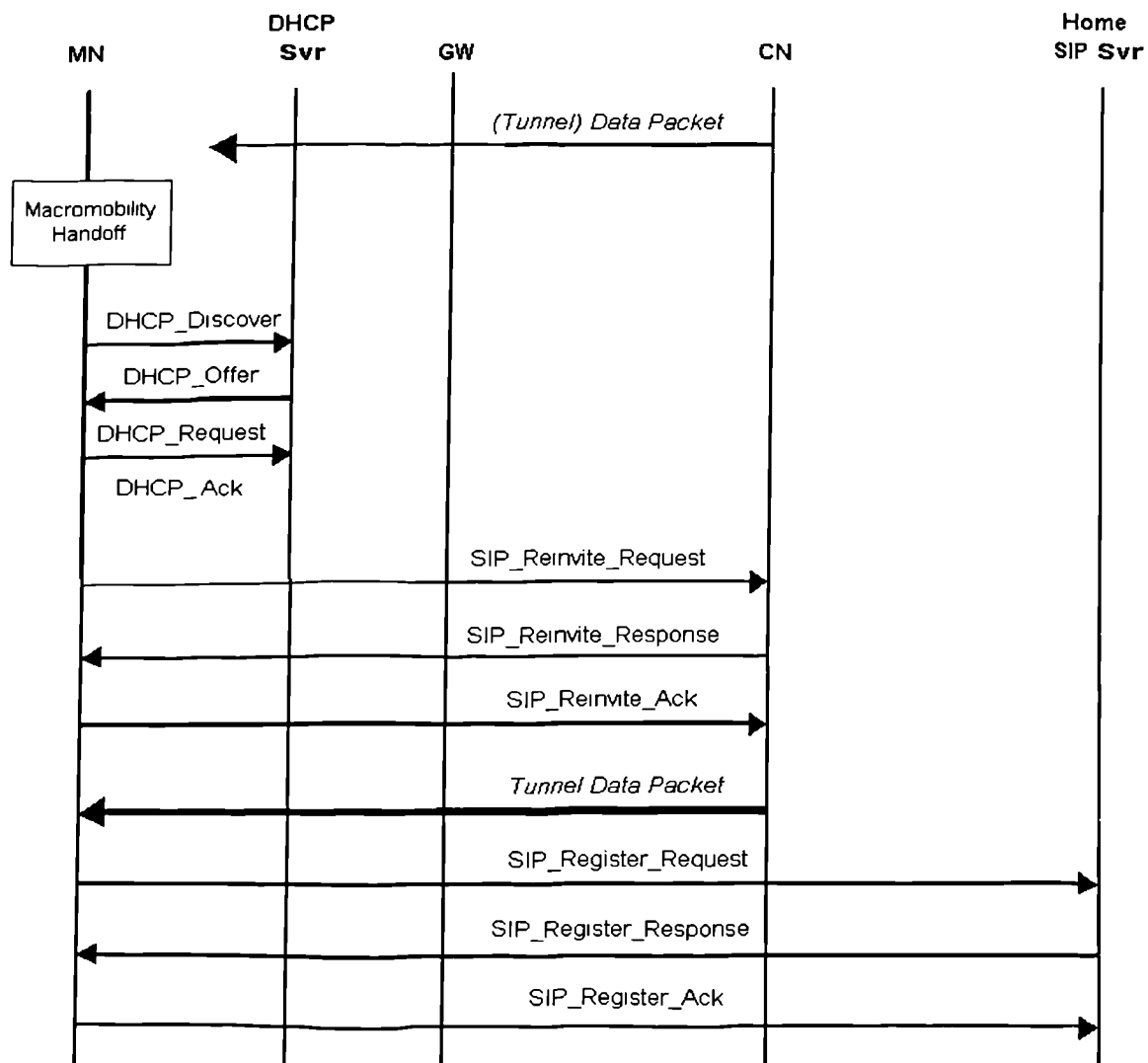


Figure 5-19: Message Sequence for the SIP-based approach.

Here it must be stressed that, locating Mobile IP Foreign Agents at micro-mobility gateways would require the seamless interworking of Mobile IP with the candidate micro-mobility solution. It is known that non-real-time traffic (e.g., TCP), will bypass the NAT (Network Address Translation). This traffic would normally be routed towards the MH through its home network using tunnelling. The overhead introduced by IP-in-IP encapsulation is not so critical for this type of applications. However, macro-mobility for real-time traffic is supported with SIP signalling. Therefore, NAT functionality is required on domain edge routers, since MHs are identified by their private home addresses within micro-mobility domains. The use of NAT for real-time traffic introduces problems that involve the blocking of IP communications, because the IP-based voice and video devices behind the NAT, have private IP addresses that are not routable outside their local domain or on the public Internet. This is due to IP address of the calling end-host embodiment in the exchanged VoIP signalling messages (e.g. H.323/H.225, MGCP, SIP). This problem is currently being investigated by the IETF MidCom (Middlebox Communication) WG, which focuses its attention on communication with firewalls and NATs. In the framework of midcom, STUN protocol [ROS-02] has been proposed as a candidate solution for allowing entities behind a NAT to first discover the presence and type of NAT, and then to learn the addresses binding allocated by the NAT. The integration of SIP user agents with STUN client functionality, will allow a wide variety of applications, such as VoIP, VoD, etc to work through the existing NAT infrastructure.

Despite the NATs/STUN concept should be considered in a real-time communication system, in the framework of this research has been omitted, for simplicity reasons. The modelling (simulation scenario) of the NATs/STUN effect has been established by allowing the private domains (behind the gateways) to have some knowledge of the type of traffic to be received.

In the following few lines, we evaluate and discuss the two proposed mobility architectures (SIP- and Mobile IP-based) and provide some useful results regarding the overall mobility management framework for IP-based WLANs and its interworking/interrelation with the All-IP 3G infrastructure (global mobility management support).

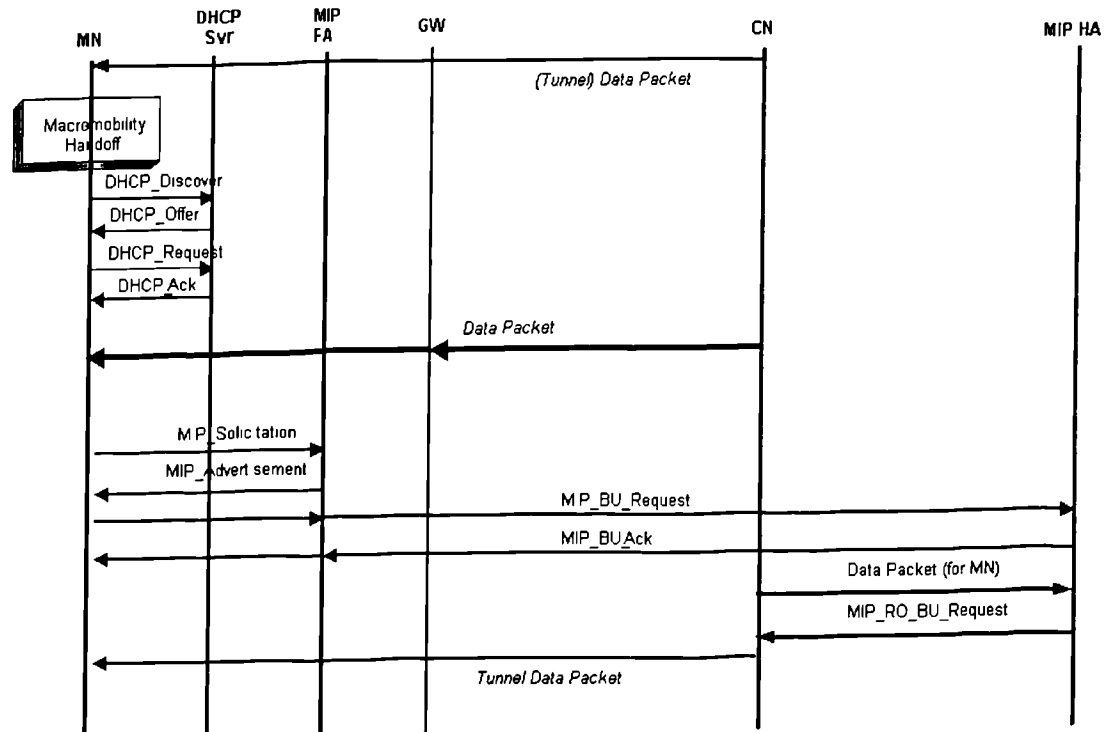


Figure 5-20: Message Sequence for the MIP-based approach.

### 5.6.2 RTP/UDP Real Time Applications

In order to evaluate the two proposed mobility architectures, we have selected a set of simulation scenarios that consider real time and non-real time applications. Therefore, we split the mobility domains in two: the first is SIP-based where real time applications are mainly treated; while the second is MIP-based where non-real time applications are primarily targeted.

In this simulation scenario, the MT performs handoff regularly. It can be seen from Figure 5-21, that the PLR (packet loss ratio) for MIP marginally increases only as handoff frequency increases (as the main source of packet loss for MIP is packet discard due to playout delay<sup>3</sup> or  $t_{po}$ ). In contrast, SIP creates more packet loss as the number of handoffs increases (as the main source of packet loss is due to RTP pause time). Generally speaking, SIP gives better performance if MT is relatively static, while MIP is better if MT handoffs more frequently. In this setup, when the mean handoff interval is greater than about 15 sec (for wireless link delay of 50 ms) and about 10 sec (for wireless link delay of 20 ms), SIP is better. Since the handoff interval in practice is very

<sup>3</sup> The playout delay is set 1000ms for VoD traffic and 500ms for VoIP traffic. If the packet delay is higher than the playout delay, then the packet is discarded.

likely to be higher, we can deduce that SIP gives a better handoff performance, as far as real time applications are concerned (performance can also be affected by other factors such as distance between CH and MT, wireless link delay & delay of DHCP, AAA and security operation if applied).

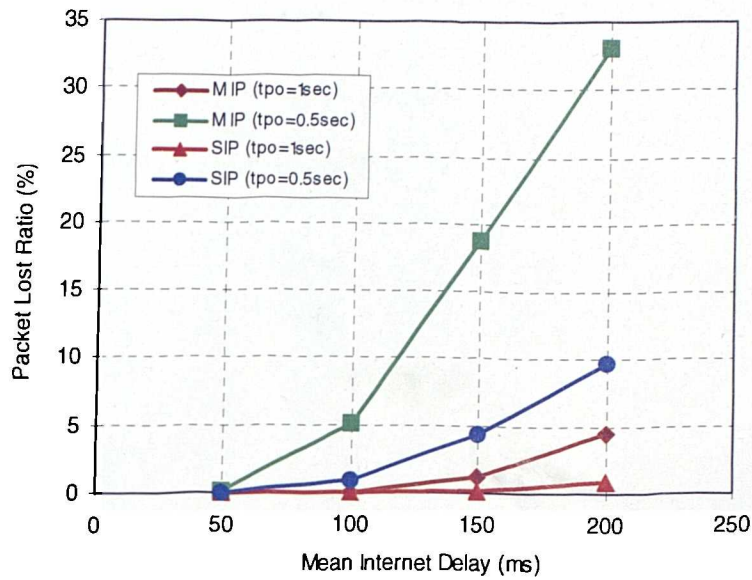


Figure 5-21: PLR for RTP session running over SIP and MIP (MT is static).

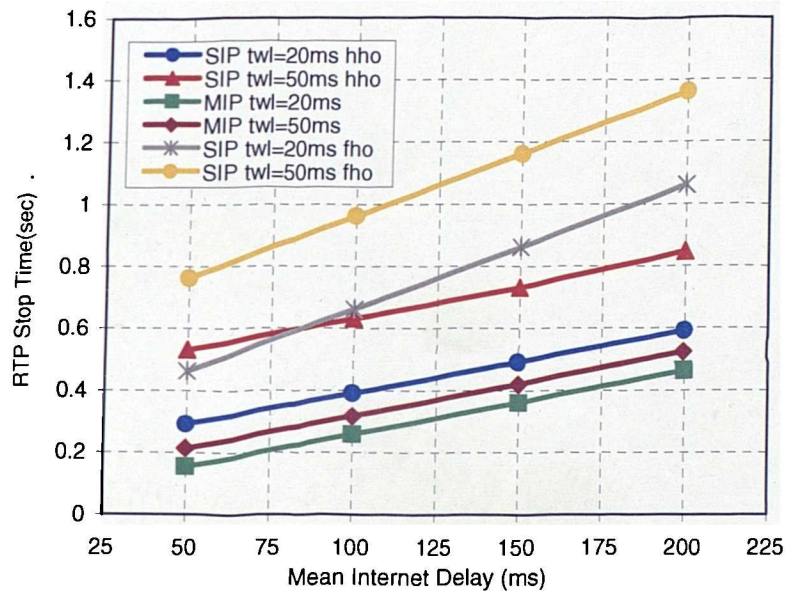


Figure 5-22: RTP stop time for SIP and MIP during handoff.

As it is shown in Figure 5-22, RTP session pauses when handoff starts, while the SIP MT takes longer time to resume RTP. Concurrently, the higher the wireless link delay ( $= twl$ ), the worse the handoff performance is. The SIP-based MT performs the following actions before resuming RTP session:

- DRCP/DHCP (4 messages involved), no DAD (Duplicated Address Detection), no ARP, involved, therefore it is similar to DRCP.
- Send SIP-Reinvite message to CN
- Receive SIP-OK from CN

On the other hand, the MIP-based MT performs the following process before resuming RTP session:

- Configure CoA based on advertisement information
- Send MIP-BU to HA
- Receive BU-ACK from HA

This is due to separate address acquisition procedure (DHCP) used, thus the total delay for SIP is higher.

Additionally, we have half handoff (hho), when SIP MT resumes RTP communication, once OK is received from the CN, while its handoff is completed once signalling is done with CN (without informing the HR). In the full handoff (fho) case, SIP MT resumes RTP communications only after OKs are received from the CN and home registrar, while it is completed after all signalling is exchanged between the Home SIP server and CN.

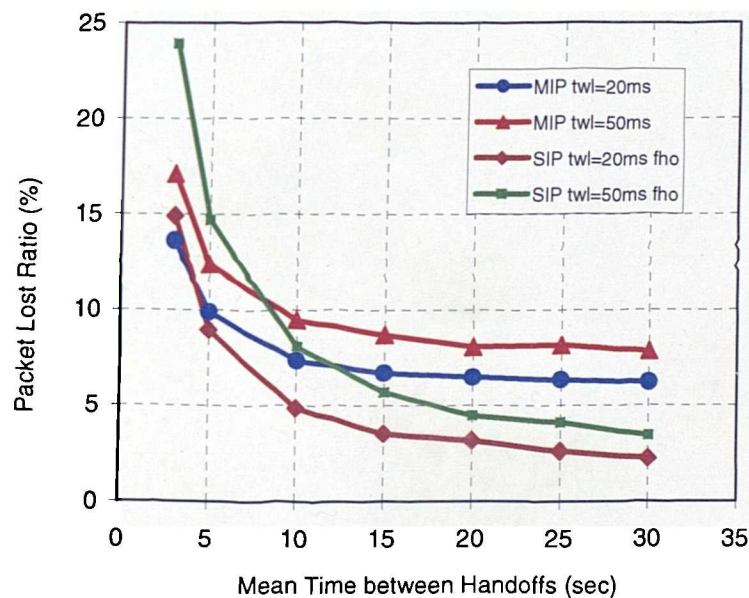


Figure 5-23: PLR for SIP (fho) and MIP.

Note that in this implementation SIP mobile terminal resumes RTP after receiving response from CN, it does not wait for SIP-Reregister-OK from its Home Registrar. However, in a commercial implementation, MT needs to wait for the OK from the HR, which imposes even higher delay.

Furthermore, DHCP here is the simplest implementation. Neither DAD nor ARP packets are sent. Thus, a commercial DHCP implementation will have imposed a much greater delay.

The results shown in Figure 5-23 and Figure 5-24 are based on MIP handoff without the old MIP FA forwarding packets to new FA. It can be clearly seen that the MIP behave worse in all cases than SIP.

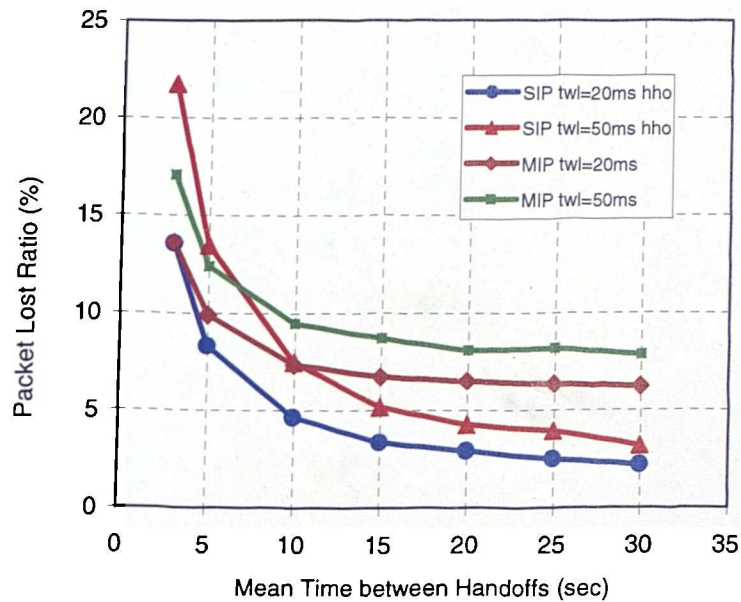


Figure 5-24: PLR for SIP (hho) and MIP.

Note that, hho emerges when SIP MT resumes RTP once OK is received from CN. While, fho is accomplished when SIP MT resumes RTP only after OKs received from CN and home registrar.

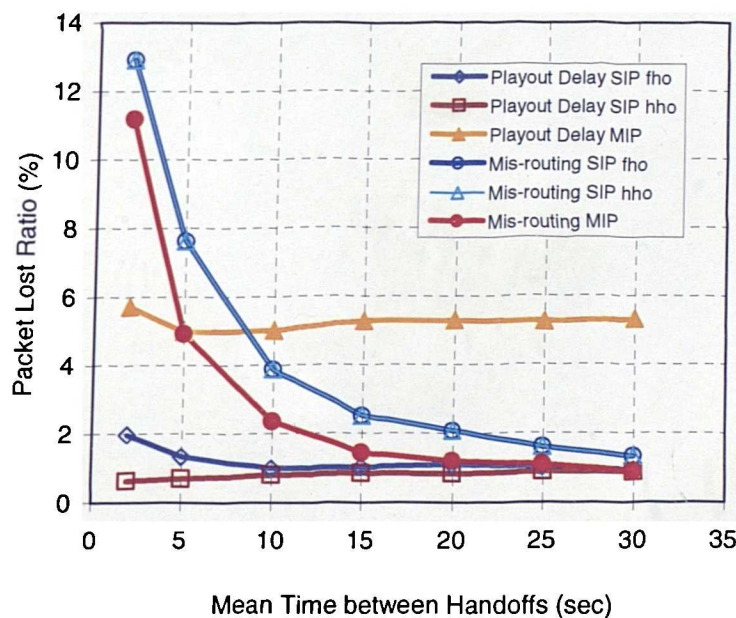


Figure 5-25: PLR for SIP and MIP, for  $twl = 20$  ms (without old MIP FA forwarding packets to new FA).

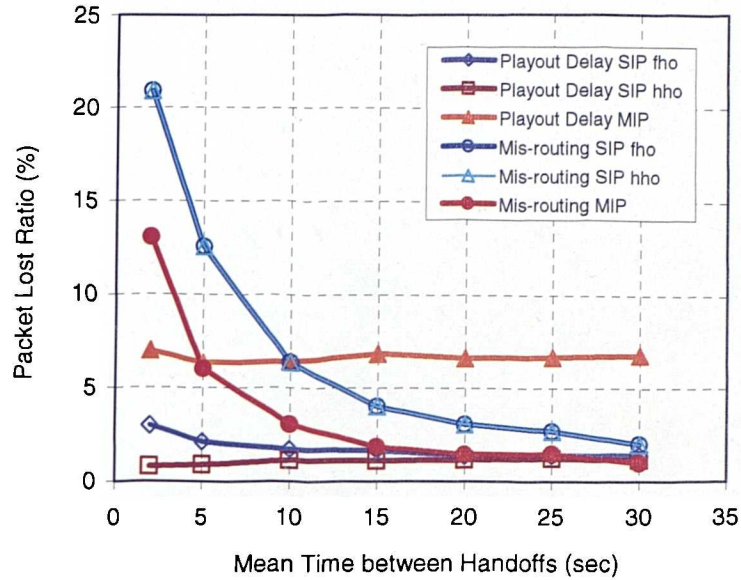


Figure 5-26: PLR for SIP and MIP, for  $t_{wl} = 50\text{ms}$  (without old MIP FA forwarding to new FA).

Referring to Figure 5-25 and Figure 5-26, these illustrate the PLR for RTP packets for both MIP and SIP mobility management schemes versus the delay as wireless link was set to 20 ms and 50 ms. In this scenario, the old FA is not forwarding any packets at the new FA (mis-routing case, where packets are routed to the MH's old visited network). As it is shown, the more frequent the handoffs in SIP, the RTP pause time increases and thus more packets are lost. On the other hand in the MIP case, frequent handoffs cause less packet loss, since handoff in MIP is relatively faster. However, MIP still has higher packet loss because of 2-step packet forwarding (triangular routing, tunnelling operation). Therefore, despite MIP is a “faster” mobility protocol, it is still not a complete solution for RTP/UDP traffic, due to its well-known drawbacks.

### 5.6.3 Non-Real Time Applications (TCP / FTP)

In the case of TCP-based traffic, we consider 2 MB files transfer. For the Mobile IP scenario, each IP packet, from MT and FTP server, is encapsulated. If it is tunnelled by HA, it will be further encapsulated. For the SIP scenario, each IP packet from MT and FTP server is also encapsulated.

As it can be seen from Figure 5-27, the FTP download response time is almost the same for both mobility approaches. In Figure 5-28 and Figure 5-29, it appears that higher throughput can be achieved for the SIP-based approach rather than in the MIP-based one.

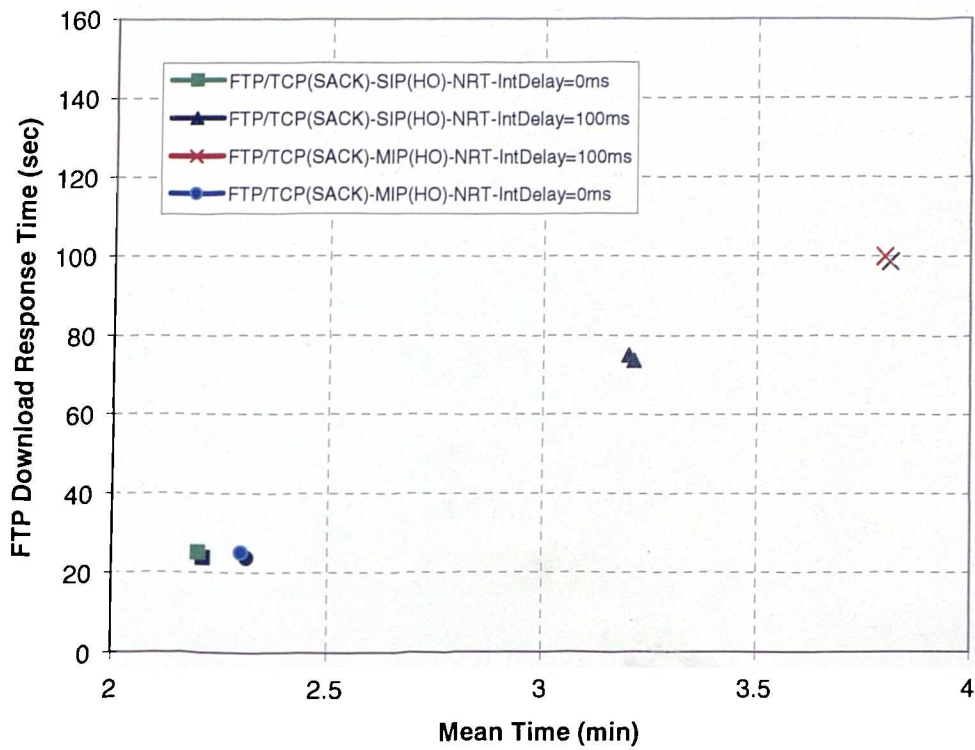


Figure 5-27: The impact of MIP and SIP mobility management schemes on FTP download response time.

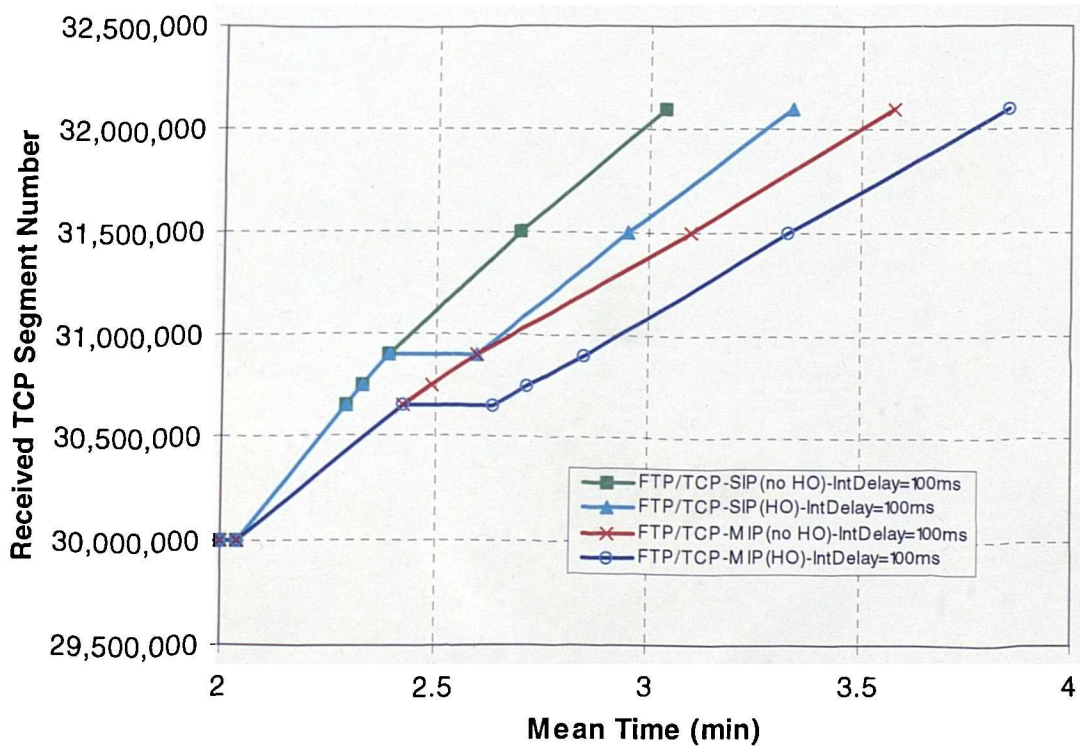


Figure 5-28: The impact of MIP and SIP mobility management schemes on the received TCP segment number.

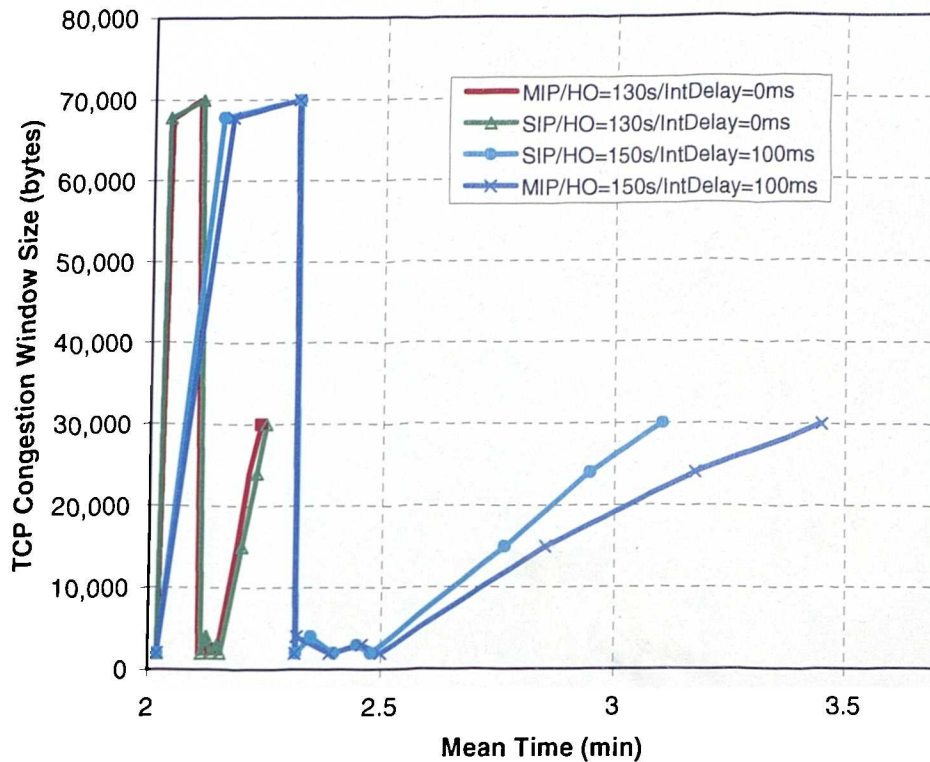


Figure 5-29: The impact of MIP and SIP mobility management schemes on TCP Congestion window.

Because of triangular routing in MIP, SIP performs better than MIP in all aspects, including file download time, congestion window recovery time, and sequence number sent.

Both SIP and MIP employ tunnelling. In MIP, tunnelling is from HA to MT. In SIP, tunnelling is from CN (FTP server in this case) to MT. There is not much different in both cases, in terms of extra IP headers.

The main difference, in terms of data delivery, is still the extra path in triangular routing in MIP. However, the better performance in SIP for TCP has been based on certain assumptions, which are fundamentally unsound [POL-DAG] for real-time implementations and hardware/software systems (while the simulation domain is a virtual dream-world, where these assumptions are considered and can be valid):

1. All IP hosts (PC, servers etc.) run SIP for both real time and non-real time applications.
2. All MTs keep a list of ongoing TCP sessions. Keeping a list of TCP may be non-trivial.
3. SIP-based messages are sent based on existing TCP session(s). This requires cross-layer operation among SIP, TCP, IP and perhaps the OS of the MT.
4. All IP hosts perform IP encapsulation (this is the reason why Route Optimisation MIP is unacceptable and hence is dropped. In fact, SIP further requires MT to perform IP encapsulation when sending TCP-packets to CN.)

Case	Settings	MIP (sec)	SIP (sec)
<b>A</b>	File size/transfer = 2 MB MTU = 1500 Internet Delay = 50 ms Wireless link delay = 20 ms TCP Setting: SACK Buffer size = 32768 MSS = Ethernet <sup>4</sup> Nagle's SWS Avoidance = Enabled	18.1678	14.4211
<b>B</b>	Same as Case A, except: HO at 129 sec	25.4107	18.7808
<b>C</b>	Same as Case A, except: Internet Delay = 100 ms	28.6107	21.3432
<b>D</b>	Same as Case C, except: HO at 135 sec	38.3842	26.6087
<b>E</b>	Same as Case A, except: Buffer size = 16384	52.0107	38.9429
<b>F</b>	Same as Case E, except: HO at 144 sec	56.6107	43.6087

Table 5-5: FTP download time for different TCP settings.

Let us assume that the MH reassumes its connection to the home network, after a handoff and if there are still files to be downloaded, intelligent FTP gives to the user three options: Stop, skip, overwrite/restart or resume. "Stop" means do not perform anything and forget about continuing the download. "Skip" means go to the next file download. "Overwrite" starts the download from the beginning. While, "Resume" means continue the download from where it stopped.

Figure 5-30 and Table 5-5 show the different file download times for various TCP configurations for both mobility schemes. It can be seen that with SIP the file download time is slightly lower than with Mobile IP.

---

<sup>4</sup> **Nagle's Algorithm:** A technique, when TCP is used then small packets are not sent. TCP uses the arrival of an acknowledgement to trigger the transmission of additional packets. Nagle's Algorithm = Send-side Silly Window Avoidance: Wait for ACK before sending another data segment, or wait until data segment is large enough.  
The problem with excess ACKs originated from the MH, which creates retransmission timeout in FTP server, then goes away.

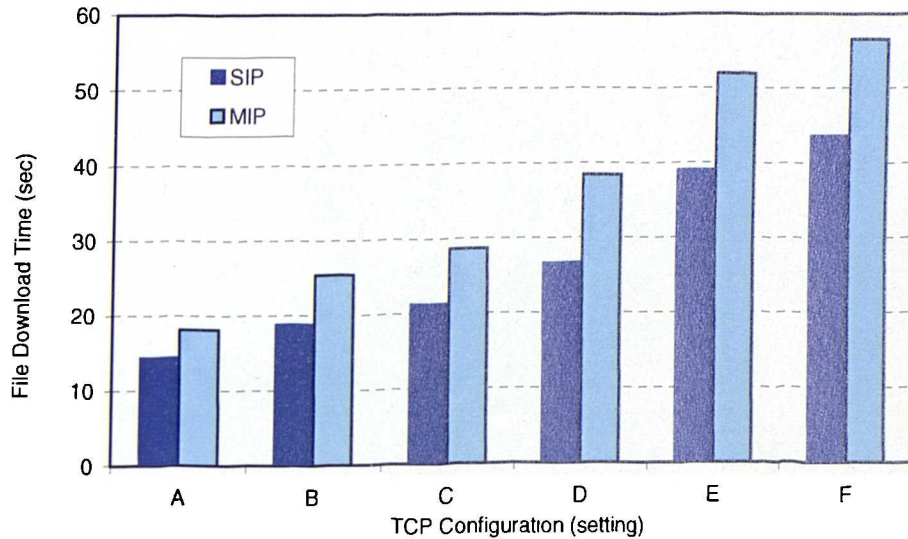


Figure 5-30: FTP download time for different TCP settings.

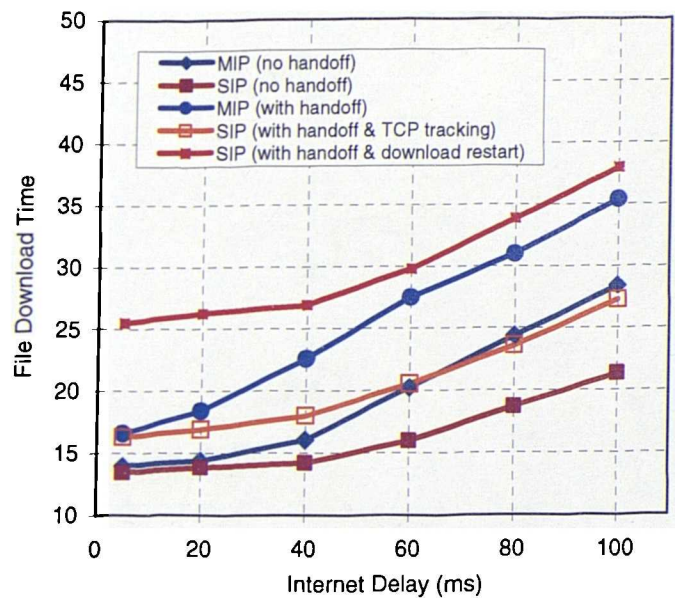


Figure 5-31: File download time for MIP and SIP (tracking agent –enabled).

Figure 5-31 shows that FTP restarts and overwrites the file download that was interrupted. This assumes that FTP client/application/program/software is intelligent enough to detect an incomplete and interrupted download. Furthermore, it is clear that when a CH (SIP UA) is not equipped with a TCP tracking agent, the performance of the SIP-based mobility scheme is further degraded on average, around 10 times.

## 5.7 Discussion and Summary

In this chapter a complete mobility framework for IP-based WLAN networks was proposed. Initially, two micro-mobility protocols were evaluated namely Hierarchical Mobile IP and Cellular IP. It was demonstrated that HMIP performs better than CIP under certain circumstances and scenarios. As a test case and for the remainder of the chapter, Cellular IP was selected to support mobility in IP-based WLAN domains.

Consecutively, multilayer mobility architecture, using a domain-based approach was presented in a real-time scenario (Test-bed). Terminal (and application layer) mobility is supported by SIP, while network mobility is provided by either DHCP or CIP. Both schemes are capable of realizing seamless mobility but the results have shown that the SIP/DHCP scheme is not a suitable mobility solution. On the other hand, the performance of the SIP/Cellular IP scheme was impressive, offering relatively small handover delay between subnets in an IP-based wireless network.

This work also considers the impact of two mobility management schemes in the performance of multimedia services in All-IP network infrastructures. The study has shown that SIP is not worse than MIP for handling mobility. For real time applications (RTP/UDP based) SIP performs better than MIP, while for non-real time applications (TCP-based), SIP performs better under the following conditions (which make it unsuitable for non-real time traffic, since all increase the signalling overheads):

- All CHs interpret SIP messages
- All CHs perform IP encapsulation
- Support from TCP tracking (to allow TCP session to continue)
- Support from application layer
- Cross-layer cooperation is needed between SIP, TCP, IP and perhaps the OS of the MH
- While MIP performs better, when the triangular problem is treated.

Therefore, the standing out approach is to have a multi-layer, multi-protocol and hierarchical mobility management framework, i.e.:

- *CIP or HMIP for local mobility and fast/seamless handoffs;*
- *MIP for non-real time applications and global mobility/vertical handoffs;*
- *SIP for real time applications and terminal/service/personal mobility.*

# Chapter 6

## 6 AAA Schemes for IP-Based WLAN

### 6.1 Introduction

In an IP-based complex architecture that provides multimedia services for WLAN users, the management of the authentication and authorisation of users as well as the agreements between the ISPs in terms of policies, charging and billing is of paramount value. An important step in proposing a secure and seamless WLAN architecture is simulation studies, since these are the basis for making decisions. Moreover, the AAA concepts are fairly complicated and a great need for simulations arises. These activities allow us to evaluate the AAA architecture in a virtual environment, which imitates real-life scenarios. We investigate whether the AAA model fits in the overall WLAN architecture (QoS provisioning and mobility management) proposed in this work.

Therefore, the studies and simulation in this context give insights into the following concepts:

- **Network access studies:** Proposed AAA architecture for WLAN systems, the IEEE802.1x protocol and various AAA protocols (Radius, Diameter etc) analysis.
- **Network access simulation:** Two WLAN technologies are proposed for accessing an all-IP network infrastructure, IEEE 802.11 and ETSI HiperLAN/2. However, only H2 authentication is simulated during the handoffs, this is due to the existence of similar studies<sup>1</sup> in the 802.11 authentication area but not in the H2 one.

### 6.2 AAA Architecture in All-IP networks

Considering the increasing number of Internet users, the Internet Service Providers are facing various challenges. An ISP should be able to provide a service required by a user whose network access is of a different nature and may belong to another ISP. This service as well as the network access should be offered in a secure way. On the other hand, the ISPs should also be able to protect their interests from malicious users by authenticating them and charging them afterwards if such services are offered.

---

<sup>1</sup> They are described in [PAC-02].

The Authentication, Authorisation and Accounting infrastructure offers a solution to the above challenges. The AAA architecture has been specified by the IETF AAA WG in [RFC-2903], [RFC-2904] and [RFC-2905]. The AAA framework takes into account various access technologies, like Mobile IP and roaming, aiming to provide a simple, scalable and secure infrastructure.

This work aims to provide an all IP-based network infrastructure that will offer seamless multimedia services to users who access the network via a WLAN technology. Thus, we need to propose a scalable and robust AAA infrastructure. AAA includes the following concepts [GAT-POL]:

**Authentication:** Authentication is the process of determining who the user is. This process assumes that the user possesses a unique piece of information, namely a username/password pair, a secret key or perhaps biometric data (fingerprints for example) that serves as identification credentials. However, the access policy should be flexible, i.e. it could be applied on per user, group, global or location basis. In this research, two kinds of authentication are required for a user whose network access is a WLAN. The first is to allow access to the network, this is via the 802.1x protocol [802.1x-01] and the second one is to allow access to a multimedia service. The SIP protocol can be used for the latter.

**Authorisation:** defines what rights and services the user is allowed to get once the network access is granted. This depends on the users dial-in properties and eventually on local policies. For example, an ISP might allow a subscriber to start and receive calls over the network but he is not allowed to use some other services like a voicemail system. Usually authentication and authorisation are treated together.

**Accounting:** Accounting is the collection of data on resource usage for the purpose of analysing, auditing and billing. Accounting depends on the type of data, the frequency of the message exchange and the charging model. The Accounting includes the control of data gathering, the transport and the storage of the accounting data. The Accounting process can be configured through accounting policies. Accounting data may be application/service-dependent based on the user's profile and the required service. In this work's architecture, data can be collected at the AAA servers or at the SIP servers.

### **6.2.1 AAA in WLAN**

The WLAN corporate environments are presumed to provide unrestricted access to the network. This open nature requires the authentication of the wireless stations to the access points. The only way to restrict access to the WLAN network is through WEP (Wireless Equivalent Privacy) or 802.1x (Port Based Network Access Control).

### **6.2.1.1 WEP**

WEP is a part of the 802.11 specifications [802.11-WG]. Some static keys are used for each station (workstations/mobile terminals and access points) to provide both authentication and encryption. When WEP is enabled, each station (workstation/MT or access point) has up to four keys and the decryption is successful, if all the keys are decrypted. However, this set of keys is the same for everyone using the same AP, so the network in this case is not secure enough. On the other hand, WEP suffers from another problem, which is the weakness of its encryption process. It should also be noted that WEP is only available when an AP is in use. For an Ad hoc network, the stations cannot enforce authentication and must pass all data in clear.

### **6.2.1.2 802.1x protocol**

The 802.1x protocol applies to the association of wireless hosts (supplicants) to Access Points (Authenticators) with a supporting authentication server (Radius or Diameter). The 802.1x authentication process uses the Extensible Authentication Protocol (EAP) [802.11-WG] between the MT and the authentication server. EAP is a general protocol and allows different authentication protocols to be selected, such as smart cards, Radius, Diameter, Kerberos, Public Key Encryption, One Time Password, TLS, and others.

The Authenticator is configured with two logical access points to the LAN through a single physical LAN port, namely the controlled port and the uncontrolled port. The uncontrolled port allows an uncontrolled exchange between the authenticator and other systems of the LAN regardless of the system authorisation state. Normally, the uncontrolled port will only accept packets to establish authentication. In 802.1x, these are EAP over LAN or EAPOL packets. The second logical AP allows an exchange between a system on a LAN and the authenticator services only if the system is authorised. Here, the so-called Network Access Identifiers are used to identify the users.

The standard for access authorisation between a MT and the AP is 802.1x and the standard between the AP and the AAA server is any AAA protocol mentioned above. When the Access Point receives EAPOL frames and relays them to the authentication server, the Ethernet header is stripped and the remaining EAP frame is re-encapsulated in the Radius format for example. The EAP frames are not modified or examined during encapsulation. When the AP receives frames from the authentication server, the server frame header is removed, leaving the EAP frame, which is then encapsulated for Ethernet transmission and sent to the MT.

A typical authentication process in 802.1x is shown in Figure 6-1.

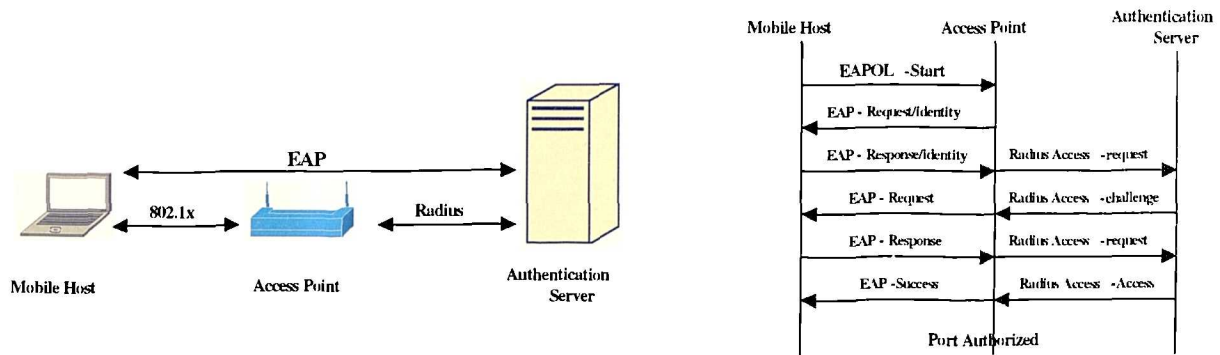


Figure 6-1: Authentication Process in 802.1x.

- The AP requests an identity from the MT;
- The MT sends its identity to the AP;
- The AP forwards the information to the AAA server via EAP (Extensible Authentication Protocol);
- The AAA server and the MT have an EAP authentication dialogue. The authentication dialogue between the MT and the AAA server must be negotiated between them as part of EAP dialogue. The authentication dialogue between the MT and the Authentication server is carried in EAP frames. The EAP frames are carried as EAPOL in 802.1x, and as EAP attributes in Radius, if this latter is used. It should be noted that mutual authentication eliminates rogue Access Points. EAP protocol supports a variety of authentication methods as explained earlier;
- If the dialogue is successful, the MT and the AAA server share a session key;
- The AAA server sends the session key to the AP in a Radius attribute as a part of Radius acceptance message;
- The AP enables its controlled port for the MT.

## 6.2.2 AAA in the Internet

We attempt in this section to describe a general architecture for an AAA infrastructure that covers a large variety of Internet services. We will also provide a brief description of the AAA protocols used for communicating between the AAA servers.

### 6.2.2.1 General AAA architecture

Here, an AAA activity related to a user requiring a multimedia service is explained. The user network access could be a network where he already has a registration and a profile or one that

he/she is visiting. On the other hand, the service that he/she wants access to may require a single end user like video streaming or Fax over IP, or multiple end users like VoIP or multimedia conferencing. Whatever the service on offer, we are interested only in the AAA activity related to users issuing the request. More complex scenarios could be found in [RFC-2905]. In general, the Provider who provides network resources is different from the one providing multimedia services. In our discussion, the network access is assumed “granted”. We mainly focus on the authentication, authorisation and accounting of the offered services. A general AAA scenario may involve the components depicted in Figure 6-2 and described further on.

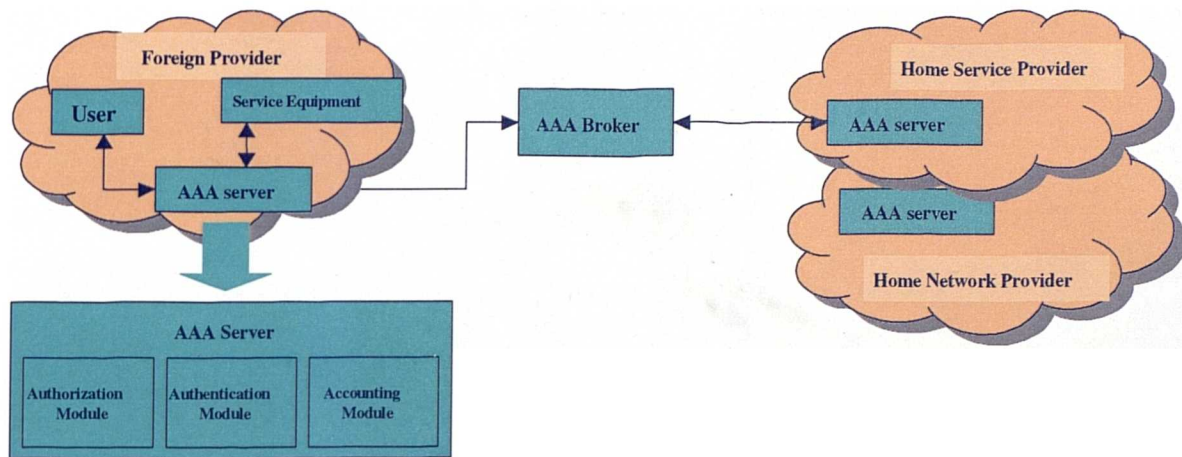


Figure 6-2: General AAA architecture [REB-02].

- **The User:** An entity requiring authentication to get access to the network resources and multimedia services.
- **The Foreign Provider:** It is a provider that a user (having subscription with another Network Provider), wants to get access to the network through it.
- **The Home Network Provider:** The Provider that has an agreement with the user to provide him access to the network. This agreement may be a subscription. If the user is roaming, the Foreign Provider will eventually contact the Home Network Provider to check whether the user is allowed to get an IP address. This is based on an agreement between the two Providers and the user profile.
- **The Home Network Provider AAA server:** This server authorises access to the network based on the profile of the user.
- **The Home Service Provider:** This entity provides services (VoIP, video, printing, Fax over Internet, etc) to the user, if the request is local. If the user got access to the network through a Foreign Provider, the last one will check with the Home Service provider whether the user has rights to the service he/she is requiring.

- **The Foreign Provider AAA server:** Authorises a service based on an agreement with the Home Provider. This agreement may involve elements that are not relevant to an individual user, for example the total agreed bandwidth between the Foreign Provider and the Home Service Provider. The AAA server should be generic, since it deals with a variety of applications requiring AAA. Since, each application has its own requirements, an AAA server should support some application specific protocols to communicate with application specific service components. To handle that, an application interface should exist between the generic AAA server and a set of Application Specific Modules, which can carry out the functionality required by each application.
- **The Foreign Provider Service Equipment:** This entity provides the service itself. It might be a NAS in a dial service, or a router in the QoS service, or print service in the Internet Printing service. If the request could be satisfied locally, the Foreign Home Service Provider Service Equipment will provide the service. In the roaming case, the Service Equipment also provides the service but after a user's inter-authorisation between the Foreign Provider and the Home Provider.

In fact, the ASM (Application Specific Module) is a component that allows Policies to influence the Service Equipment. The Service Equipment can be anything, and therefore there is no standard API to communicate with. An ASM has a standard API, and it can be used to translate requests into something the Service Equipment understands.

- **The AAA Broker:** A broker provides scalability for the AAA protocol. It can be either a proxy or redirect server, and may be operated by roaming consortiums. A broker may also provide Certificate Authority services, by issuing certificates to all AAA servers within the consortium (or alternatively signing existing certificates).

An AAA scenario effectively consists of a user who wants access to a particular service. This could be from his/her home domain or a domain that he/she is visiting. In the first case, the local AAA server will be in charge of the AAA activities. In the second case, the local AAA server contacts the user's Home domain AAA server to which it forwards the request. This might be direct or through a broker.

### ***6.2.2.1.1 Authentication phase***

The authentication phase may be handled in three different ways. We should also distinguish between an AAA scenario within a single domain and another one where the home domain is different from the foreign domain (i.e. roaming scenario).

**Single domain model**

- *Agent Sequence:* In this scenario, the user sends a request to the AAA server, which applies a policy associated to the user and the requested service. It also sends a request to the Service Equipment. The last one sets up the service and notifies the AAA server about that. In turn, the AAA server notifies the user.
- *Pull Sequence:* Here the user sends a request to the Service Equipment, which forwards it to the AAA server. This last evaluates the request and returns a response to the Service Equipment, which sets up the service (if granted) and notifies the user.
- *Push Sequence:* In this scenario, the user gets from the AAA server a certificate giving him the right to access the service. The user includes this certificate in the request to the Service Equipment. The last one verifies that the request is approved by the AAA server and sends OK to the user.

**Roaming model**

In this case, the user who is visiting another domain is assigned an IP address, because he/she wants to be allowed to use a multimedia service. The above sequences can be adapted to this new situation.

- *Agent Sequence:* The user sends a request to the Foreign Provider AAA server, which contacts the AAA server of the home domain after checking the local policies. The AAA server of the home domain authenticates the user and sends a response to the AAA server of the foreign domain. This last applies a policy associated to the user and the requested service. It also sends a request to the Service Equipment. The Service Equipment sets up the service and notifies the AAA server about that. In turn, the AAA server notifies the user.
- *Pull Sequence:* Here the user sends a request to the local Service Equipment, which forwards it to the AAA server of the same domain. The last one evaluates the request locally or through the AAA server of the home domain and returns a response to the Service Equipment, which sets up the service (if granted) and notifies the user.
- *Push Sequence:* In this scenario, the user gets from the local AAA server or through the AAA server of the home domain, a certificate giving him right to access the service. The user includes this certificate in the request to the local Service Equipment. The last one verifies that the request is approved by the local AAA server and sends OK to the user.

### **6.2.2.1.2 Authorisation phase**

Once the user is authenticated, an authorisation phase is required. This means that the Service Provider checks the services, which the user is allowed to access. As mentioned earlier, the authorisation will be performed by the local AAA server or by the Home Service Provider AAA server. The authorisation is based on the user profile (and some local policies).

The user profile is just a record that includes information about the services the user is allowed to use and the restrictions (if there are any) related to these services.

Policies are a set of rules to manage and control access to the network resources. For example, the user is not allowed to log on after 23h or he/she is not allowed to require a QoS support of 2 Mb/s. A policy rule consists of a policy condition and a policy action. The latter is evaluated and depending on the evaluation of the condition, a policy action is taken or not. Policy definitions are maintained and stored in a policy repository by the organisation that requires them. A policy is evaluated at a Policy Decision Point and enforced at a Policy Enforcement Point, which is the Service Equipment in our case. Policies are of two kinds: Intra-domain policies which are evaluated entirely within a single administrative domain and Inter-domain policies which may originate in a foreign domain and be evaluated in the local domain, or they may originate in the local domain and be evaluated in a foreign domain.

### **6.2.2.1.3 Accounting phase**

The user got the access to the requested service and starts using it. At this level, the AAA server needs to keep track of the multimedia session, in order to charge the user for the network resources he/she is consuming. One major issue in accounting is *the specification of the parameters on which the accounting will be based*. Accounting may be based on metering and gathering network usage parameters. Metering is the process of measuring the parameters *within* the network related to the customer service usage, like call duration, time of the day and destination of the call. It might be also the measurement of the traffic flowing inside a network, from or to other networks. In other words, the user should be charged for the transport and the content of the consumed data. The transport accounting means charging the user with regards to the transport of IP packets, for instance the number of packets transmitted. Mechanisms like DiffServ base their operation on the transport accounting notion. The charging may then be based on traffic volume, QoS class, time of day and the destination. It may also differentiate between local, national and international traffic. The content accounting means that the user will be charged based on the nature of packet: video or audio stream, simple text or image and so on. Another issue for accounting is who will get paid; the Network Provider or the Content Provider.

Accounting can occur in a single domain or between domains administrated by different providers. Intra-domain accounting involves the collection of information on resource usage within an administrative domain, for use within that domain. However, in an inter-domain accounting, the collected information will be used within another administrative domain. In this case, accounting packets and session records will typically cross administrative boundaries.

The accounting management architecture involves interactions between network devices, accounting servers and billing servers. The network device collects raw resource consumption data, which will be sent, using an accounting protocol, to the accounting server that transforms this information into accounting metrics forms. Afterwards, this information is sent to a billing server or to another server for auditing or analysing purposes. This process is depicted in Figure 6-3.

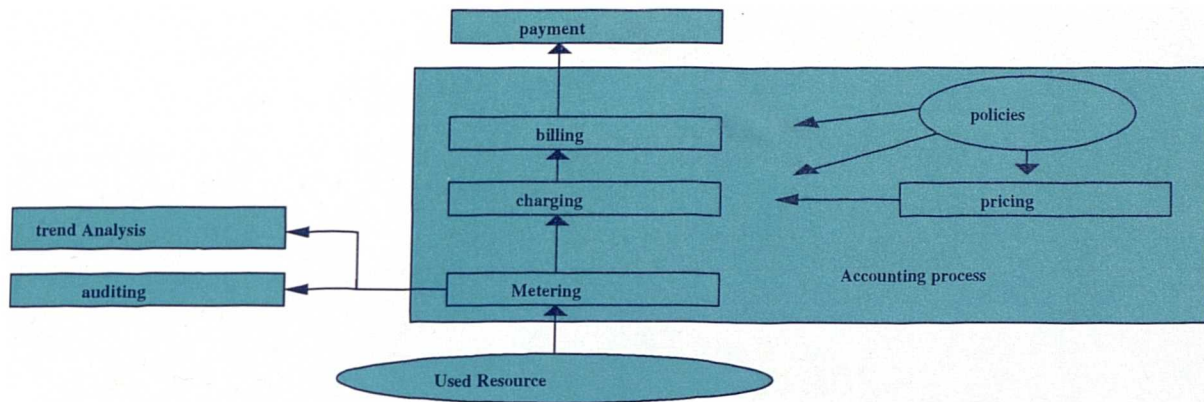


Figure 6-3: Accounting process [REB-02].

### 6.2.2.2 AAA protocols

The AAA protocols are needed to carry out the requests for the AAA activities. Many AAA protocols are actually used, namely, Radius [RADIUS], TACACS [TACACS], Diameter [DMTR] and COPS [COPS]. These protocols are different with respect to the functionalities that they can provide. For instance, COPS is more a Policy-based protocol in a RSVP environment. Diameter is considered as an extension of Radius and it provides functionalities that Radius does not provide, in particular AAA activities for Mobile IP. These protocols are also different with respect to their features: flexibility, security mechanisms and ability to support other protocols.

In the following subsections, we provide a brief overview for Radius and Diameter, since they will be used further on in this research. For detailed descriptions, the given references might be useful.

#### **6.2.2.2.1 Radius**

Radius is an AAA client-server protocol between a NAS (Network Access Server) and a centralised Radius server. In the beginning, Radius was designed for intra-domain use, where the NAS, proxy, and home server exist within a single administrative domain, and proxies may be considered as trusted components. Radius is an open protocol and is distributed as source code. This protocol is based on the UDP transport protocol. It runs over IPv4/IPv6 and uses the AVPs format for data delivery. Radius provides hop-by-hop security and a variety of authentication methods. A shared secret key is used between the different servers or the end users and the servers. This means that transactions between the client and Radius server are authenticated through the use of a shared secret, which is never sent over the network. In addition, any user passwords are sent encrypted between the client and Radius server, to eliminate the possibility that someone snooping on an unsecured network could determine a user password. However, Radius does not provide end-to-end security. Also, it does not provide mechanism for congestion control.

#### **6.2.2.2.2 Diameter**

With the new emerged access technologies and the increasing number of the Internet users, the concept of Internet roaming services is introduced. Roaming is just the fact that remote end-users connect to the network but they are authenticated by a local (visited) ISP that is different from their home ISP. This is achieved by allowing service providers to cross-authenticate their users. As explained earlier, this was one problem that Radius was suffering from. Contrary to Radius, the new protocol Diameter was designed from the beginning to be used in a multi-domain environment.

Diameter is an end-to-end AAA protocol. It consists of a base protocol and a set of extensions. The base protocol should not be used by itself and must be used with a service-specific extension (for instance, Mobile IP, Nasreq, Strong Security, etc). The term “extension” denotes an extension to the base protocol functionality to manage the requirements, in terms of AAA, of a particular service. This protocol is based on the TCP or SCTP transport protocols. Diameter also uses the AVPs for delivering data. It provides hop-by-hop, end-to-end security and a mechanism for congestion control. It runs over IPv4/IPv6. Diameter is using IPsec as the underlying security mechanism, so it does not require the use of shared secrets for message authentication. Diameter also allows AAA activities when roaming.

### 6.3 AAA Simulations for WLAN Systems

This section provides an insight into the AAA simulation activity in the context of this work. As mentioned in previous chapters, a robust AAA infrastructure for IP-based WLANs that provides fast access granting to multimedia services is required. Therefore, a user wishing a service from All-IP infrastructure will be only authenticated once, i.e. to get access to the network. Thus, these AAA activities are evaluated here.

#### 6.3.1 Network Access Simulation

In the context of this study, the network access can be a WLAN network. For a WLAN-based technology (802.11 or HiperLAN/2), the IEEE standard 802.1x is chosen to provide authentication, authorisation and accounting. The H2 standard is also considered. Even though H2 claims to be technically superior (as it was discussed in Chapters 3 and 4), the standard 802.11 came first into the market and is considered as technologically mature and the predominant WLAN standard across the world. Table 6-1 summarizes the characteristics of 802.11a (Higher Speed Physical layer in the 5 GHz band) and HiperLAN/2.

	IEEE 802.11a	HiperLAN/2
<b>Throughput (Max PHY rate)</b>	54 Mb/s (for max size packets)	54 Mb/s
<b>Data Transport</b>	Ethernet	Support ATM, Ethernet, IP
<b>Spectrum</b>	5 GHz	5 GHz
<b>Real-time Data</b>	Delivers good streaming even without QoS features	H2 was designed with multimedia in mind. For time-critical applications requiring QoS, H2 performs better
<b>Network Compatibility</b>	Not compatible with 802.11b and 802.11	Not compatible with 802.11b
<b>QoS</b>	Best effort transmission	QoS inherited from ATM/IP

Table 6-1: Comparison between IEEE 802.11a and HiperLAN/2 [GRE-02].

Therefore, the simulation of the AAA activities for WLANs includes a comparison between the manners of authenticating a user in H2 and 802.11. This can be based on the authentication process, fast access and security.

Let us quote that if the 802.11-based WLAN is using WEP (see subsection 6.2.1.1) for authentication and encryption, then HiperLAN/2 is a much better choice. However if the 802.1x is used, the situation is quite different and the difference between 802.1x and H2 resides in:

- *Mutual authentication:* Mutual authentication allows the MT and the AP to identify each other. Both H2 and 802.1x provide this concept. Moreover, the authentication protocol for both is a challenge-response protocol [3GPP-33.9].
- *Key management:* In HiperLAN/2 [3GPP-33.9], the authentication is based on pre-shared key or RSA signatures. Once the keys are generated, each MT is assigned an authentication key identifier. The AKI (Authentication Key Identifier) will be sent to the access point with which the MT has a security association. Among the AKI that can be used, the NAI, the IEEE address and the X.509. As explained in [EVO-D2.1], the standard 802.1x is EAP-based and a shared key can be negotiated within the EAP dialogue.
- *Encryption:* In H2 [3GPP-33.9], the algorithms DES and 3DES could be used to provide confidentiality protection. In 802.1x/EAP, TKIP or AES can be used.

Note: For a user who wishes to access an All-IP infrastructure through a 3G network, UMTS AKA protocol is chosen to perform the authentication and the authorisation.

### 6.3.1.1 IEEE 802.1x Model for HiperLAN/2

In this section, we provide a brief description of the IEEE 802.1x model, while we study and simulate its applicability for the H2 case, by considering a pre-authentication scheme for fast handoffs [PAC-02].

The 802.1x protocol applies to the association of wireless hosts/MT (supplicants) to Access Points (Authenticators) with a supporting authentication server (e.g. Radius). The 802.1x authentication process uses EAP between the MT and the authentication server.

The standard for access authorisation between a MT and the AP is 802.1x and the standard between the AP and the AAA server is any AAA protocol mentioned above. When the AP receives EAPOL frames and relays them to the authentication server, the Ethernet header is stripped and the remaining EAP frame is re-encapsulated in the Radius format for example (Radius server is used from now on). The EAP frames are not modified or examined during encapsulation. When the AP receives frames from the authentication server, the server frame header is removed, leaving the EAP frame, which is then encapsulated for Ethernet and sent to the MT.

The basic 802.1x architecture is shown in Figure 6-4.

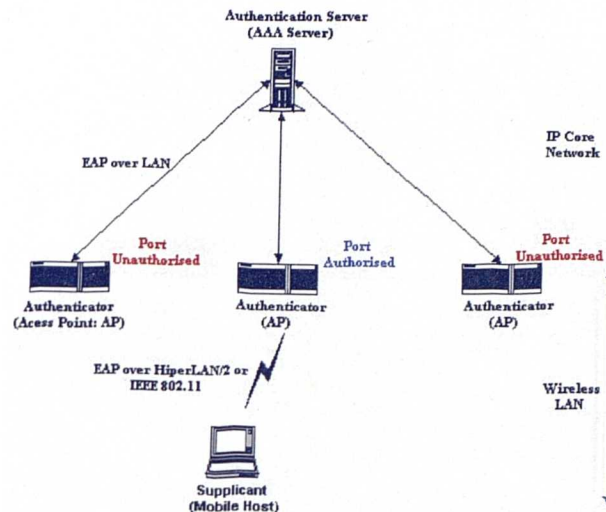


Figure 6-4: The basic IEEE 802.1x architecture.

Consecutively, we discuss and analyse a fast inter-AP handoff scheme. In the scheme, a MT performs authentication procedures not only for the current AP but also for neighbouring APs, when it handoffs. Figure 6-5 shows the basic components involved in the pre-authentication. Since IEEE 802.1x provides a network port access control scheme, this is more scalable than others. The supplicant system (MT and AP) is an entity at one end of the point-to-point LAN segment that is being authenticated by the authenticator attached to the other end of the link. The authenticator system (adjacent AP and AAA server) facilitates authentication of the entity attached to the other end of the link via the authentication server. The authenticator controlled port is in the unauthorised state; therefore the authenticator makes use of the uncontrolled port to communicate with the supplicant, using EAPOL and EAP to communicate with the AAA server. The innovation here, is that MT authenticates with several APs during the scanning process, so that when association is required, the MT is already authenticated.

Figure 6-6 and Figure 6-7 show the key distribution in IEEE 801.x for the H2 case. Despite a MT sends an authentication request to the AAA server, the latter sends multiple authentication responses to all APs within that frequent handoff region. The FHR<sup>2</sup> (Frequent Handoff Region) is a set of adjacent APs. It is determined by factors such as location of APs in a WLAN service area and users' movement pattern. By definition, the FHR is comprised of APs, which MTs are likely to move shortly. Although, there are a lot of APs in a public WLAN, the movement ratio between different APs is not the same (depending on geographical profile/location, velocity of users, QoS contracts i.e. applications type and services class etc). In order to take these factors into account

<sup>2</sup> This FHR is based on two fundamental conditions: 1) the user service class (here we use the service classes defined in Chapter 4) and 2) the user moving pattern [PAC-02].

and evaluate the pre-authentication process for HiperLAN/2 systems, we use the FHR selection algorithm described in [PAC-02].

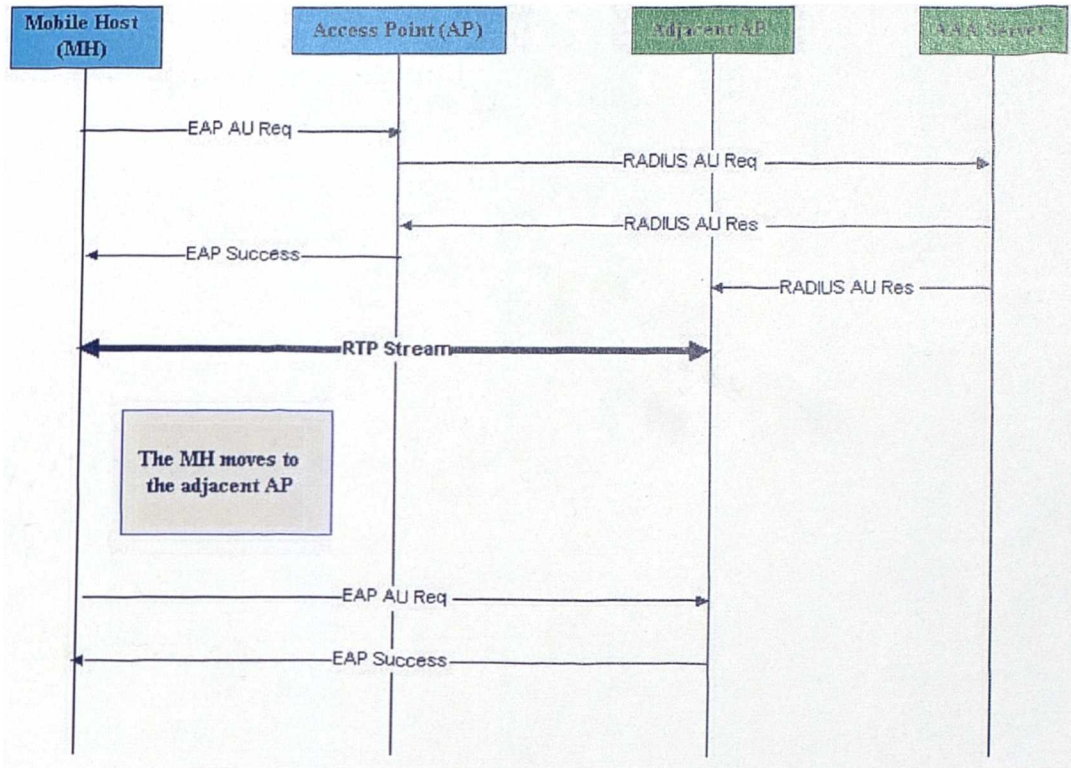


Figure 6-5: Pre-authentication scheme for fast handoff in H2.

After having received responses, all APs, except the current one, keep the authenticated information until the specific timer has expired. If there is no handoff during that period, the information expires and the MT should perform re-authentication when a handoff occurs.

Figure 6-7 shows the re-authentication message flow after handoff. We assume that the adjacent AP belongs to the FHR. If a MT hands off to the adjacent AP, no further message exchanges are needed (due to pre-authentication). In the 802.1x model, the controlled port changes into an authorised port after authentication. However, in the pre-authentication scheme, the port in “ready” state can change imminently to “authorised” state just by checking the identifier of the MT, without further interaction with the AAA server.

Finally, we compare the handoff delay for the fast handoff scheme as proposed in [PAC-02] and for different service classes as described in Chapter 3, section 5. The total delay during handoff can be measured by the summation of all the delays across the wired and wireless links [AKH-POL].

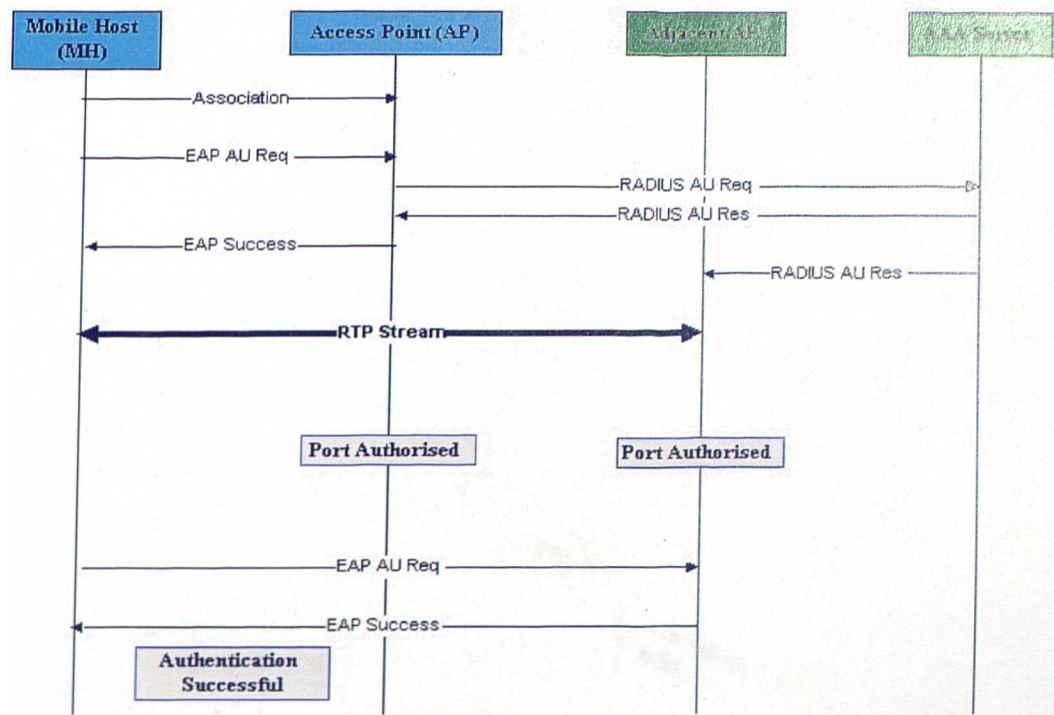


Figure 6-6: Message exchanges before handoff.

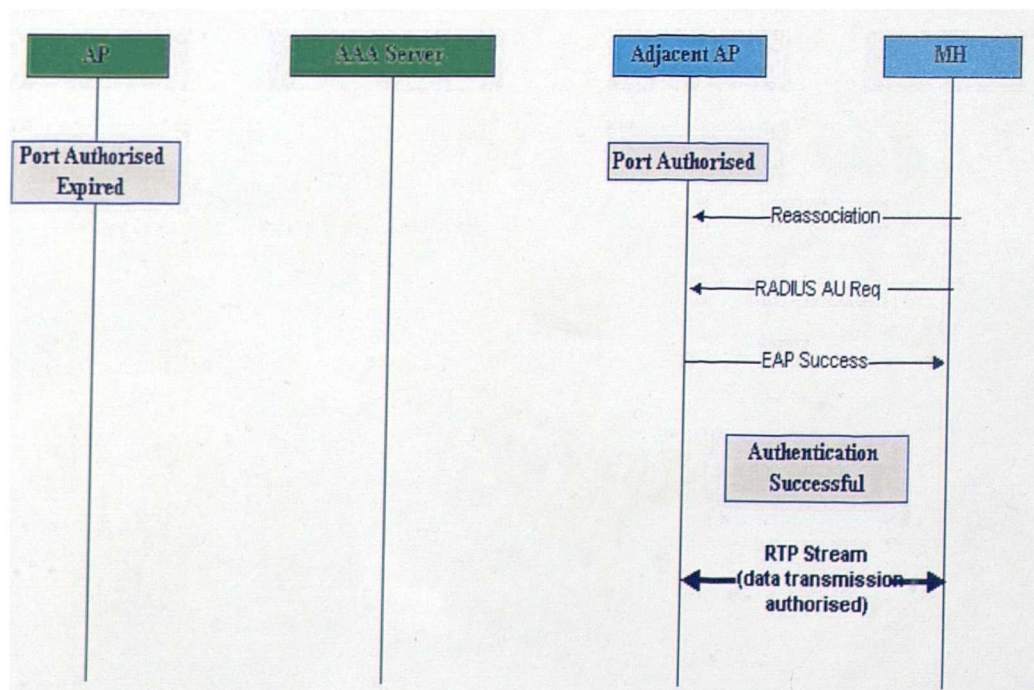


Figure 6-7: Message exchanges after handoff.

Figure 6-8 shows the average handoff delay when the AAA server is located in the local domain. It can be seen that the average delay is lower when the pre-authentication scheme is used rather than the BE (Best Effort) class/user (normal handoff process) [DA-PO'03]. Additionally, users belonging to CO (Conversational) class have lower delays compared to users in other classes, e.g. ST (Streaming) or IN (Interactive). Figure 6-9 shows the simulation results in the remote AAA server case. Here the pattern is similar to the one in Figure 6-8. However, the average handoff delay is approximately 20, 25 and 30 msec higher than the delays for IN, ST and CO class users respectively. Therefore, the proposed pre-authentication scheme based on the IEEE802.1x protocol for H2 (WLAN) systems is more appropriate for multimedia services and other delay sensitive applications.

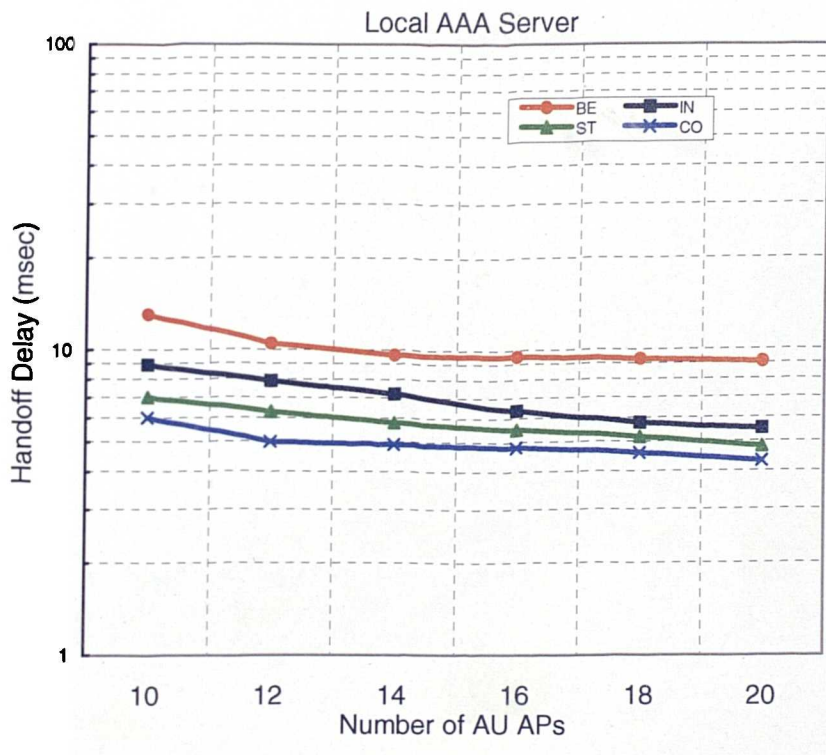


Figure 6-8: Handoff delay for pre-authenticated H2 APs (Local AAA server).

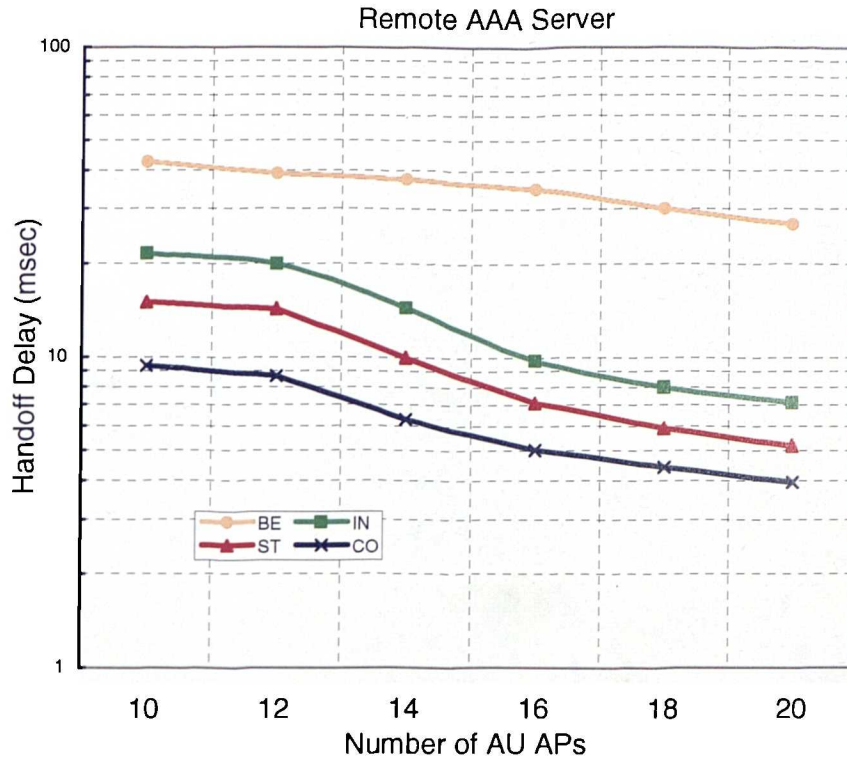


Figure 6-9: Handoff delay for pre-authenticated H2 APs (Remote AAA server).

### 6.3.2 Service Access Simulation

The multimedia services deployed on the Internet and offered through a WLAN infrastructure are SIP-based. This means in particular that SIP is coupled with the AAA procedures in order to grant or deny the service(s). Several scenarios for this coupling can be studied and compared with each other based on the following criteria:

- *Performance*: this involves an estimation about the number of messages needed to be transported to authorize the usage of a service in addition to the establishment of the SIP session
- *Complexity*: this criteria describes the changes needed to existing components and the additional overhead required managing new components
- *Flexibility*: this part deals with how difficult is the integration of new services or the replacement of the current ones into the communication scenarios

However, such work is out of the scope of this research, forming nonetheless a strong item for future research and test-bed/simulation works.

## **6.4 Discussion and Summary**

The access to multimedia services through an all-IP based network infrastructure can be achieved in two steps. First, the access to the network is required. Secondly, the user needs to get authenticated and then authorised, in order to use the services provided by the SIP Provider. The AAA simulation work handled here, concentrated on the performance evaluation of the proposed AAA architecture for network access. We have thoroughly investigated various AAA protocols (Radius, Diameter) for their suitability to fit into the proposed WLAN architecture. In the authentication front, an extended study was performed regarding the use of the IEEE801.x protocol for supporting fast handoffs in HiperLAN/2 systems. The performance evaluation shows that the proposed scheme saves significant amounts of time during handoffs compared with the traditional one. Various service classes were used based on our findings from the previous chapters (QoS provisioning); the mobility models from Chapter 5 were also used.

# Chapter 7

## 7 Conclusions and Future Work

This chapter concludes the work described in the previous chapters of this Ph.D. thesis. It mainly focuses on the objectives and the achievements of this work. Possible extensions of this research, along with the publications list, are also presented.

### 7.1 Major Achievements

The following bullet points outline the achievements and constitute the major stations of this work, as discussed in the previous chapters:

- An analytical/mathematical approach was followed to identify the interworking requirements between various H2 layers (PHY-DLC/MAC-CL) and the cross-layer and cross-protocol QoS issues both on data as well as control planes. Various entities of the three layers were specified and a design approach for a catholic QoS framework was presented.
- A link adaptation technique has been studied and an algorithm based on Chow loading technique proposed to enhance and potentially improve the QoS support/guarantees (in terms of gain and throughput). The enhanced H2 system was compared to the existing design and the overall system performance showed significant improvement, in the order of approximately 1 to 3 magnitude units.
- On the MAC level, two well-known traffic controllers (algorithms) can be potential candidates for providing QoS in the IP-based HiperLAN/2, namely WFQ and CBQ. The WFQ for both domains and for all traffic types performs better than CBQ. While, the IntServ domain proves its superiority (for real-time services) against DiffServ, with the drawback being that of scalability. Furthermore, both (WFQ and CBQ) algorithms were shown to be able to support the specified QoS contracts, in terms of throughput and delay requirements. The selection of such algorithms has the advantage of their previous suitability for IP networks, and the minor modifications to be imposed in order to work in WLANs environments. Similarly, RSVP proved to perform well with HiperLAN/2, by improving significantly the performance of the whole system. Finally, the QoS framework was further enhanced by introducing congestion (reuse of WFQ, CBQ) and connection admission (a simple and efficient resource-polling algorithm, specified by H2) control algorithms.

- A complete mobility framework for IP-based WLAN networks was proposed. Initially, two existing micro-mobility protocols were evaluated namely Hierarchical Mobile IP and Cellular IP. It was demonstrated that HMIP performs better than CIP, in terms of scalability but rather worse in terms of reliability. Furthermore, the overall performance of the two protocols was very similar and the small fluctuations were recorded due to the type of traffic or the unpredicted load balancing of the network. As a by-default test case and for the remainder of the chapter, Cellular IP was selected to support mobility in IP-based WLAN domains.
- Consecutively, a domain-based approach was presented in a real-time scenario (Test-bed). Application layer and terminal mobility was supported by SIP, while network mobility was provided by either a layer-2 solution (e.g. DHCP) or a micro-mobility protocol (in this case Cellular IP). Both schemes were capable of realising seamless mobility but the performance of the SIP/Cellular IP scheme was impressive, offering relatively small handover delay between subnets in an IP-based wireless network (based on IEEE 802.11 WLAN).
- The design, specification and evaluation of two mobility management architectures (namely SIP-based and Mobile IP-based) has shown that SIP is not worse than MIP for handling mobility. For real time applications SIP performs better than MIP, while for non-real time applications, SIP performs better but the imposed signalling overhead is overwhelming (two-way encapsulation, TCP tracking agents on all CHs, application layer support, etc). On the other hand, Mobile IP can facilitate a better performance, when the triangulation is treated. Therefore, we proposed an approach, where a multi-layer and multi-protocol mobility management framework is activated: CIP or HMIP for micro-mobility (local) mobility and fast/seamless handoffs; MIP for non-real time applications and macro-mobility (global) mobility/vertical handoffs and SIP for real time applications and terminal/service/personal mobility.
- We have thoroughly studied and investigated various AAA protocols (e.g. Radius, Diameter) for their suitability to fit in the proposed WLAN architecture. Furthermore, an AAA architecture for WLAN systems was proposed. In the authentication front, an extended study was performed regarding the use of the IEEE801.x protocol for supporting fast handoffs in HiperLAN/2 systems. The performance evaluation showed that the proposed pre-authentication scheme (FHR approach [PAC-02]) saves significant amount of time for multimedia services (VoD, VoIP, streaming applications etc) during handoffs compared with the traditional one (the service classes of Chapter 4 were also used).

## 7.2 Future Work

Significant research is still required in the B3G systems and especially the hotspot WLAN arena. The physical layer that has been touched on in this work definitely needs more attention. Apart from the PHY issues, there are other research topics and unidentified problems that demand solutions. Some of them are discussed below.

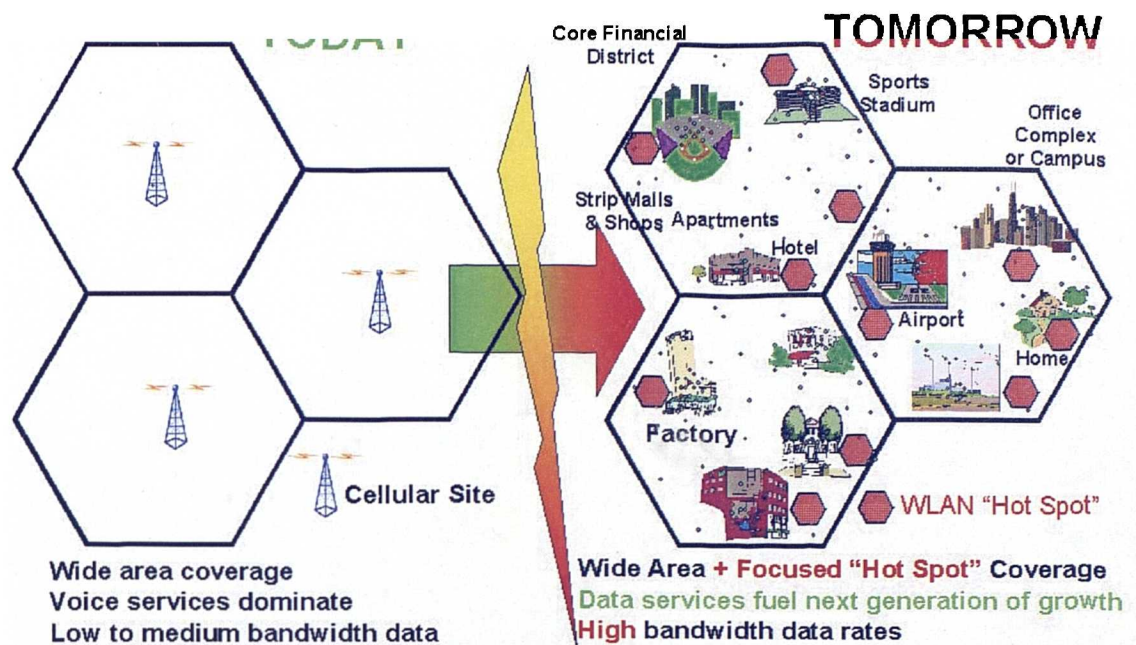
In the PHY area, an innovation may be the design of a PHY-layer enhanced with selected Multiple Transmit Multiple Receive techniques in order to increase the capacity and link reliability. These techniques increase the number of simultaneous users in the same cell without any deep modification of the transmission standards for HiperLAN/2. The higher layers (DLC) should be modified to take into account the capacity increase brought by the MTMR [MIN-POL].

The recent advances in the telecommunication market in the form of high speed access networks and open service models based on SIP or OSA (Open Service Access) for example as well as the availability of low-cost wireless access technologies, is leading to the emergence of a new breed of network and service providers. The idea is to provide a security framework for establishing security and trust relations among, network operators, service providers and users, allowing thereby smooth roaming among heterogeneous administrative domains/networks. To achieve that, there is a need to realise a programmable inter-domain AAA infrastructure for supporting the dynamic establishment of trust relations between independent providers, in a secure and distributed manner over a heterogeneous IPv4 (including Firewalls and NATs) and IPv6 environment.

Furthermore, the support of security context transfer mechanisms to enable fast and smooth intra- and inter-domain handover performance in mobile environments is another hot topic within the same research area. The context transfer operation should be further studied, analysed, and evaluated [GEO-POL]. The dependency and interoperability between context transfer protocol [LOU-03], [NAK-01] and handoff protocols should be investigated. Various mobility management-related protocols might be considered, including but not limiting to Mobile IP, Hierarchical Mobile IP and Cellular-IP, [GE-PO'03]. Initially existing packet formats defined by these mobility protocols should be investigated to check whether they could be utilised as carriers for context transfer information or as triggers to context transfer [KOO-03].

Additionally, the multimedia services deployed on the Internet and offered through a WLAN infrastructure are SIP-based. This means in particular that SIP is coupled with the AAA procedures in order to grant or deny the service(s). Several scenarios for this coupling can be studied and compared to each other based on the following criteria: *performance, complexity and flexibility* (extensions to Chapter 6).

Not too long ago, communications meant voice and mobility meant cellular. But today we see that subscribers are increasingly relying upon diverse communications solutions for a complex array of voice, data, and multi-media needs, many of which are being addressed by Internet/Intranet connected networks, e.g. at offices, homes, shopping areas, transport facilities, and the like.



*The Coming Discontinuity*

Figure 7-1: Today and Tomorrow in the Wireless World [DAG-VEL].

The missing part of the puzzle is any overlaying strategy for integration of these disparate solutions into what, from the customers perspective, appears as a single fabric. The core components for this integration strategy include cross-network and cross-service solutions for mobility, authentication, subscriber administration and consolidated accounting and billing. These are all elements, which the current cellular world provides better than anyone else, but only for itself. The opportunity for the cellular/wireless community is to broaden its focus and associations by extending these core services to enterprise networks, ISPs, public access hot spots, such as airports and shopping malls, and to private hot spots, such as home networks (see Figure 7-1) [DAG-VEL].

Concluding, we can say that, the communication environment is changing. The cellular community has a unique opportunity to leverage its current strengths to turn threat into opportunity by using the seamless mobility solution paradigm:

- A family of seamless mobility handsets that will operate simultaneously in both cellular and WLAN environments.

- Intelligent gateways that will interface between the cellular and the WLAN systems to hand off connections and enable the consolidated mobility tracking, authentication, subscriber administration and accounting services.
- An approach to consolidated authentication that will be compatible with 2G, 2.5G, 3G Systems.

### **7.3 Publications**

Accepted papers:

- i. C. Politis, R. Tafazolli, "MAC Protocol for All-IP Hiperlan2", IEEE VTC Spring Conference, Rhodes island, Greece, p.p. 657, 6-9 May 2001.
- ii. C. Politis, R. Tafazolli, "MAC Protocol for Supporting QoS in All-IP HiperLAN2", 4th European Wireless Conference (EW'02), Florence, Italy, 25-28th February 2002.
- iii. C. Politis, V. Almenar, R. Tafazolli, "Physical Layer Techniques for Optimum Performance of HiperLAN2 Systems", IEEE 3Gwireless'02/WAS'02 Conference, San Francisco, USA, 29-31st May 2002.
- iv. Gatzounas D, Sisalem D, Politis C, Grilli S, Rupp S, Velentzas R, Tafazolli R, et al, "Seamless Multimedia Services Over All-IP Based Infrastructures: The EVOLUTE Approach", IST Mobile and Wireless Telecommunications Summit, Thessalonica, Greece, 16-19th June 2002.
- v. C. Politis, R. Tafazolli, "Connection Admission Control for Supporting QoS in IP-Based HiperLAN2 Networks" IEEE International Symposium on Advances in Wireless Communications (ISWC'02), Victoria, BC, Canada, 23-24 September 2002.
- vi. K. Chew, C. Politis, R. Tafazolli, "Performance Evaluation of Micro-mobility Protocols for All-IP Based Infrastructures" Wireless World Research Forum (WWRF), 7th meeting, Eindhoven, The Netherlands, December 3-4, 2002.
- vii. C. Politis, K. Chew, R. Tafazolli, "Multilayer Mobility Management for All-IP Networks: Pure SIP vs. Hybrid SIP/Mobile IP", IEEE VTC Spring Conference, Jeju, Korea, 21-24th April 2003.
- viii. N. Akhtar, M. Georgiades, C. Politis, "SIP-based End System Mobility Solution for All-IP Infrastructures", IST Mobile & Wireless Communications Summit 2003, 15-18th June 2003, Aveiro, Portugal.
- ix. E. Mino, M. Engels, E. Cianca, M. Chenu-Tournier, A. Hutter, I. Forkel, C. Politis, P. Rosson, P. Wachsmann, D. Nogu  t, "STRIKE: spectrally efficient fixed wireless network

- based on dual standards”, IEEE SympoTIC Symposium, 26 – 28th October 2003, Bratislava, Slovakia.
- x. D. Noguet, E. Mino, M. Engels, M. Chenu-Tournier, C. Politis, P. Rosson, P. Wachsmann, “STRIKE: spectrally efficient fixed wireless network based on dual standards” poster presentation on IST Mobile & Wireless Communications Summit 2003, 15-18th June 2003, Aveiro, Portugal.
  - xi. T. Oda, D. Sisalem, C. Politis, J. Nielsen, J. Urban, M. Palola, M. Sawada, O. A. Gonzalez, T. Dagiuklas, "White Paper on Cooperative Networks (CoNet) Research Challenges and Priorities" Wireless World Research Forum (WWRF), 9th meeting, Zürich, Switzerland, 1-2 July 2003.
  - xii. M. Georgiades, K. Chew, C. Politis, R. Tafazolli, "Context Transfer Extension to Mobility Protocols for All-IP Based Infrastructures" Wireless World Research Forum (WWRF), 9th meeting, Zürich, Switzerland, 1-2 July 2003.
  - xiii. C. Politis, T. Dagiuklas, "On the performance of Mobility Management Schemes in all-IP Network Architectures", International HET-NET'03 Conference, Ilkley, U.K, 21-23 July, 2003.
  - xiv. L. Fan, C. Politis, R. Tafazolli, "New DLC Layer Design and Evaluation for MTMR HiperMAN System", International HET-NET'03 Conference, Ilkley, U.K, July 21-23, 2003.
  - xv. M. Georgiades, C. Politis, N. Akhtar, R. Tafazolli, "Context Transfer Extension to Cellular-IP" IETF draft, draft-georgiades-seamoby-ctecip-01.txt, expires September 2004.
  - xvi. M. Georgiades, N. Akhtar, C. Politis, R. Tafazolli, “AAA Context Transfer for seamless and secure multimedia services over All-IP networks”, 5<sup>th</sup> European Wireless Conference (EW'04), February 24-27, 2004, Barcelona, Spain.
  - xvii. Tasos Dagiuklas, Christos Politis, Glaucio Bigini, Yacine Rebani, Dorgham Sisalem and Steven Xarhoulacos, "Hierarchical AAA Architecture for User and Multimedia Service Authentication in Hybrid 3G/WLAN Networking Environments" to appear IEEE International Conference on Network Protocols, Berlin, Germany, October 5th-8th 2004.
  - xviii. C.Politis, K.Chew, N.Akhtar, R.Tafazolli, "Multilayer Mobility Management for Supporting Multimedia Services over All-IP Networks: Performance Evaluation", IST EVOLUTE International Workshop, University of Surrey, Guildford, UK, November 10, 2003.
  - xix. N.Akhtar, M.Georgiades, C.Politis, R.Tafazolli, "Real-time Evaluation of Mobility Management Schemes for all IP-based Infrastructures", IST EVOLUTE International Workshop, University of Surrey, Guildford, UK, November 10, 2003.

- xx. L. Fan, C. Politis, R. Hoshyar, R. Tafazolli, “Design of New MAC-PHY Interface for MTMR-based HIPERMAN System” Wireless World Research Forum (WWRF), 10th meeting, New York, USA, October 27-28, 2003.
- xxi. Z. Li, C. Politis, R. Tafazolli, V. Sdralia, T. Dodgson, “SIP and MIP Inter-working for Mobility Management in All-IP Networks” Wireless World Research Forum (WWRF), 10th meeting, New York, USA, October 27-28, 2003.
- xxii. C. Politis, T. Oda, S. Dixit, M. Smirnov, et al, “Cooperative Networks for the Future Wireless World” to appear IEEE Communications Magazine, publication date 3<sup>rd</sup> quarter 2004.

Submitted papers:

- xxiii. T. Dagiuklas, C. Politis, S. Grilli, et al, “Seamless Multimedia Sessions and Real-Time Measurements across Hybrid 3G and WLAN Networks” submitted to International Journal of Wireless and Mobile Computing (IJWMC).
- xxiv. C. Politis, T. Dagiuklas, R. Tafazolli, “Mobility Management and AAA Support Characterisation for Providing Seamless Multimedia Services over IP-Based Infrastructures” submitted to IEEE Wireless Communications Magazine.

# References

- [30701F] BRAN WG3 PHY subgroup, "Criteria for Comparison", ETSI-BRAN document No 30701F, 1998.
- [3GPP'03] 3GPP Technical Specification Group Services and Systems; 3<sup>rd</sup> Generation mobile system Release 5 specifications. 3GPP TS 21.103 V5.3.0 (2003-03).
- [3GPP-33.9] 3GPP TR 33.9de v0.1.0, "Wireless Local Area Network (WLAN) Interworking Security", release 6, July 2002.
- [802.11-94] IEEE Project 802-Local & Metropolitan area networks, "Wireless LAN Medium Access Control (MAC) Physical Layer (PHY) Specifications, draft standard 01-12-1994.
- [802.11-WG] IEEE 802.11 Working Group, <http://grouper.ieee.org/groups/802/11/index.html>
- [802.1x-01] IEEE Standards for Local and Metropolitan Area Networks: Port based Network Access Control, IEEE Std 802.1x-2001, June 2001.
- [AKH-POL] N. Akhtar, M. Georgiades, C. Politis, R. Tafazolli "SIP-based End System Mobility Solution for All-IP Infrastructures", IST Mobile & Wireless Communications Summit 2003, 15-18th June 2003, Aveiro, Portugal.
- [BEL-02] C. Perkins, E. Belding-Royer, S. Das, "Ad hoc On-Demand Distance Vector (AODV)", draft-ietf-manet-aodv-11.txt, June 2002.
- [BER-00] Bernet, P. Ford, et al., "A framework for integrated services operation over DiffServ networks", RFC 2998, November 2000.
- [BING-00] T. Bing, M. Dell'Anna, et al, "Physical Layer Techniques for the BRAIN Air Interface" in proceedings of London International Workshop on BRAIN, November 2000.
- [BOL-00] E. Bolin, G. Lombardi, et al, "BRAIN Enhancements for the HIPERLAN/2 Air Interface to support QoS in Wireless Communications", IST Mobile Communications Summit, Galway, 2000.
- [BON-00] Servane Bonjour et al, "IP convergence layer for HIPERLAN/2", BRAIN London workshop, November 2000.
- [BRA-69] Paul T. Brady, "A Model for Generating On-Off Speech Patterns in Two-Way Conversation", The Bell System Technical Journal, No. 9, Vol. 48, pp. 2445—2472, September 1969.
- [CAC-98] Ramon Caceres, Venkata N. Padmanabhan, "Fast and scalable wireless handoffs in support of Mobile Internet Audio" ACM/Baltzer Mobile Networks and Applications, vol.3, n.4, pp. 351-363, 1998.
- [CALH-00] P. Calhoun, T. Hiller, et. al. "Foreign Agent Assisted Hand-off ". draft-calhoun-mobileip-proactive-fa-03.txt. Nov 2000.
- [CAMP-00] A. T. Campbell, J. Gomez, C-Y. Wan, S. Kim, Z. Turanyi, and A. Valko, *Cellular IP*, work in progress, draft-ietf-mobileip-cellularip-00, January 2000.
- [CAMP-02] A. Campbell, et al., "Comparison of IP Micromobility Protocols," IEEE Wireless Communications, February 2002, vol. 1, pp. 72-82.
- [CHO-99] Choi, H. K. and Limb, J.O., "A Behavioural Model of Web Traffic", Proc. of the 7th International Conference on Network Protocols (ICNP 99), Ontario, Canada,

- October 1999.
- [CHOW-95] P.S. Chow, J.M. Cioffi, J.A.C. Bingham, "A practical Discrete Multitone Transceiver Loading Algorithm for Data Transmission over Spectrally Shaped Channels", IEEE Transactions on Communications, 43, Feb/Mar/Apr 1995, pp 773-775.
- [CH-PO'02] Kar Ann Chew, Christos Politis, Rahim Tafazolli, "Performance Evaluation of Micro-mobility Protocols for All-IP based Infrastructures", in Proc. of the Wireless World Research Forum, 7th Meeting, December 2002.
- [CHU-97] M.C. Chuah, Y. Li, "Distributed Registration Extension to Mobile IP," Internet Draft, draft-chuahli-mobileip-dremip-00.txt, October 1997, Work in Progress.
- [COL-CIP] [http://www.comet.columbia.edu/cellularip/linux\\_src\\_code.htm](http://www.comet.columbia.edu/cellularip/linux_src_code.htm)
- [COPS] The COPS (Common Open Policy Service) Protocol, IETF RFC 2748, <http://www.ietf.org/rfc/rfc2748.txt>
- [DAG-VEL] T. Dagiuklas, S. Velentzas, "3G and WLANs Interworking Scenarios: Qualitative Analysis and Business Models", International HET-NET'03 Conference, Ilkley, U.K, July 21-23, 2003.
- [DA-PO'04] T. Dagiuklas, C. Politis, S. Grilli, et al, "Seamless Multimedia Sessions and Real-Time Measurements across Hybrid 3G and WLAN Networks" submitted to International Journal of Wireless and Mobile Computing (IJWMC).
- [DAS-00] S. Das, A. Misra et. al. "TeleMIP: Telecommunications-Enhanced Mobile IP Architecture for Fast Intradomain Mobility", IEEE Personal Communications System Magazine, vol.8, August 2000 pp.50-58.
- [DMTR] DIAMETER Base Protocol, IETF DRAFT, <http://search.ietf.org/internet-drafts/draft-ietf-aaa-diameter-11.txt>
- [DUT-01] A. Dutta, J. C. Chen, S. Das, S. Madhani, A. McAuley, S. Baba, N. Nakajima, Y. Ohba and H. Sculzrinne, "Implementing a Testbed for Mobile Multimedia", IEEE Globecom International Conference, 2001.
- [ETSI-475] ETSI TS 101 475 V1.2.1, BRAN; HIPERLAN Type 2; Physical (PHY) layer, TS/ 11-2000.
- [ETSI-493.1] ETSI TS 101 493-1 V1.1.1, BRAN; HIPERLAN Type 2; Packet based Convergence Layer; Part 1: Common Part, TS/ 04-2000.
- [ETSI-761.1] ETSI TS 101 761-1 V1.2.1, BRAN; HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 1: Basic Data Transport Functions, TS/ 11-2000.
- [ETSI-957] ETSI TR 101 957 (V1.1.1): "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Requirements and Architectures for Interworking between HIPERLAN/2 and 3<sup>rd</sup> Generation Cellular Systems".
- [EURE-00] EURESCOM Project, P810-PF, "Requirements and solutions on the wireless link", Deliverable 2, October 2000.
- [EVO-D2.1] IST-2001-32449 EVOLUTE (<http://evolute.intranet.gr/>), "EVOLUTE Architecture Specification", Deliverable 2.1, August 29th, 2002.
- [EVO-D2.2] IST-2001-32449 EVOLUTE (<http://evolute.intranet.gr/>), "EVOLUTE Architecture Simulation", Deliverable 2.2 (updated), January 31st, 2004.
- [FIK-99] K.E. Malki, N.A. Fikouras, S.R. Cvetkovic, "Fast Handoff Method for Real-Time Traffic over Scaleable Mobile IP Networks", Internet Draft, draft-elmalki-mobileip-

fast-handoffs-01.txt, June '99.

- [FISC-96] R.F.H. Fischer, B.H. Hiber, "A New Loading Algorithm for Discrete Multitone Transmission", in Proceedings of IEEE Globecom 1996.
- [FOO-98] Foo, S., F., Chua, K., C., "Regional Aware Foreign Agent (RAFA) for Fast Local Handoffs", draft-chuafoo-mobileip-rafa-00.txt, Work in progress, November 1998.
- [GAT-02] D. Gatzounas, D. Theofilatos, T. Dagiuklas, "Transparent Internet Mobility using SIP/Cellular IP Integration", IPCN2002, Paris, France, April 23-26, 2002.
- [GAT-POL] Gatzounas D, Sisalem D, Politis C, Grilli S, Rupp S, Velentzas R, Tafazolli R, et al, "Seamless Multimedia Services Over All-IP Based Infrastructures: The EVOLUTE Approach", IST Mobile and Wireless Telecommunications Summit, Thessalonica, Greece, 16-19th June 2002.
- [GEO-POL] M. Georgiades, C. Politis, N. Akhtar, R. Tafazolli, "Context Transfer Extension to Cellular-IP" draft-georgiades-seamoby-ctecip-01.txt.
- [GE-PO'03] M. Georgiades, K. Chew, C. Politis, R. Tafazolli, "Context Transfer Extension to Mobility Protocols for All-IP Based Infrastructures", in Proc. of the Wireless World Research Forum (WWRF), 9th meeting, Zürich, Switzerland, 1-2 July 2003.
- [GILB-60] E.N. Gilbert, "Capacity of a Burst-Noise Channel", Bell Systems Technical Journal, Vol.39, pp. 1253-1265, 1960.
- [GRE-02] J. W. Green, S. Henrichon, M. A. Said, S. Roberts, "802.11a or Hiperlan2: Which technology will emerge as the 5 GHz WLAN standard", IEEE Project 802-Local & Metropolitan area networks, draft standard 2002-05-09, University of Colorado, Boulder, May 2002.
- [GUS-00] E. Gustafsson, A. Jonsson, C. E. Perkins, "Mobile IP Regional Registration" IETF draft, draft-ietf-mobileip-reg-tunnel-03.txt, July 2000.
- [H/2-00] ETSI TR 101 683 V1.1.1, BRAN HIPERLAN Type 2: System Overview, TR/ 02-2000.
- [HAN-01] M. Handley, V. Jacobson and Colin Perkins, "SDP: Session Description Protocol", draft-ietf-mmusic-sdp-new-03.txt, July 2001, work in progress.
- [HAV-00] H. Haverinen and J. Malinen, "Mobile IP Regional Paging" Internet draft, draft-haverinen-mobileip-reg-paging-oo.txt, June 2000, work in progress.
- [IST-STRIKE] IST-2001-38354 STRIKE project, <http://ist.strike.org/>
- [ITU-323] ITU-T, H.323, "Packet-based Multimedia Communications System", Nov-2000.
- [ITU-723] ITU-T, G.723, "Speech coders: Dual Rate Speech Coder for Multimedia Communications transmitting at 5.3 and 6.3 kbit/s", Mar-1996.
- [ITU-96] ITU-T Recommendation G.114, "General Characteristics of International Telephone Connections and International Telephone circuits: One Way Transmission Time," Feb 1996.
- [JOH-02] D. Johnson, D. Maltz, Y. Hu, J. Jetcheva, "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)" draft-ietf-manet-dsr-07.txt, February 2002.
- [KAD-00] A. Kadelka, A. Masella, "Servicing IP Quality of Service with HiperLAN/2", in Proceedings of European Wireless Conference 2000, Dresden, Germany, Sept. 2000.
- [KAD-99] A. Kadelka, A. Hettich, S. Dick, "Performance Evaluation of the MAC protocol of the ETSI BRAN HIPERLAN/2 standard" in Proc. of European Wireless Conference

- 1999, Munich, Germany, October 1999.
- [KAR-00] M. Karam, F. Tobagi, "On the Traffic and Service Classes in the Internet", in Proc. of IEEE Globecom 2000, San Francisco, CA, US.
- [KES-91] S. Keshav, "On the Efficient Implementation of Fair Queuing", Journal of Internetworking Research and Experience, Vol. 2, No. 3, pp. 3-26, September 1991.
- [KOO-03] R.Koodli, C.E. Perkins, "Context Transfer Protocol for Seamless Mobility", Internet Draft, Internet Engineering Task Force, Work in Progress, draft-koodli-seamoby-ct-04.txt.
- [KRU-95] Krunz, Sass and Hughes "Statistical Characteristics and Multiplexing of MPEG Streams" in proc. of IEEE INFOCOM '95, The Conference on Computer Communications, 1995.
- [KRU-97] Krunz, Tripathi, "On the Characterization of VBR MPEG Streams" in ACM SIGMETRICS, Conference on Measurement and Modelling of Computer Systems Performance Evaluation Review 25, 1997.
- [LOM-98] Lombardo, Morabito, Palazzo and Schembra "A Fast Simulation Model of MPEG Video Traffic" in IEEE GLOBECOM 1998.
- [LOU-03] J.Loughney, "Context Transfer Protocol", Internet Draft, Internet Engineering Task Force, Work in Progress, draft-ietf-seamoby-ctp-03.txt.
- [M-1079] ITU-R M.1079 recommendation "Performance and Quality of service requirements for international mobile telecommunications-200 (IMT-2000)" June 1999.
- [MAL-00] K. El-Malki, H. Soliman "Fast Handoffs in Mobile IPv4", IETF draft, draft-elmalki-mobileip-fast-handoffs-03.txt September 2000.
- [MAN-03] Manner, X. Fu, "Analysis of Existing QoS Signalling Protocols," Internet Engineering Task Force, Next Steps in Signalling (NSIS) Working Group, February 2003 (draft-ietf-nsis-signalling-analysis-01.txt).
- [MAR-98] G. Maral, M. Bousquet, "Satellite Communications Systems", third edition, Wiley, 1998.
- [MCC-99] P. McCann, T. Hiller et al, "Transparent Hierarchical Mobility Agents (THEMA)", Internet Draft, draft-mccann-thema-00.txt, March '99.
- [MED-98] J. Medbo, P. Schramm, "Channel models for HIPERLAN/2 in Different Indoor Scenarios", ETSI EP BRAN, 3ERI085B, March 1998.
- [MENA-00] A.Mena, J.Heidemann, "An Empirical Study of Real Audio Traffic", Proceedings of IEEE INFOCOM, Tel Aviv, Israel, March 2000.
- [MIH-00] A. Mihailovic, M. Shabeer, A.H. Aghvami, "Multicast for Mobility Protocol (MMP) for emerging internet networks", in Proceedings of PIMRC2000, London, UK, September 2000.
- [MIN-POL] E. Mino, M. Engels, E. Cianca, M. Chenu-Tournier, A. Hutter, I. Forkel, C. Politis, P. Rosson, P. Wachsmann, D. Noguét, "STRIKE: spectrally efficient fixed wireless network based on dual standards", IEEE SympoTIC Symposium, 26 – 28th October 2003, Bratislava, Slovakia.
- [MIP-02] a) Perkins (Editor), IP mobility support for IPv4, RFC 3220, January 2002.  
b) C. Perkins, "IP Mobility Support," Internet RFC 2002, Oct. 1996.
- [MIP-RO] C. Perkins and D. Johnson. Route Optimization in Mobile IP. Internet draft, draft-ietf-mobileip-optim-11.txt, September 2001, work in progress.

- [MOB-IP] Mobile IP WG web page: <http://www.ietf.org/html.charters/mobileip-charter.html>
- [MON-99] C. Perkins, G. Montenegro, P. Calhoun, "Private Addresses in Mobile IP" draft-ietf-mobileip-privaddr-00.txt, June 1999.
- [MOO-01] Moon, H. Aghvami, "RSVP Extensions for Real-Time Services in Wireless Mobile Networks," IEEE Communications Magazine, pp. 52-59, December 2001.
- [MYS-97] J. Mysore and V. Bharghavan, "A New Multicasting-based Architecture for Internet Host Mobility", Proceedings of ACM Mobicom Conference, September 1997.
- [NAK-01] M.Nakhjiri, "Time Efficient conteXt Transfer (TEXT), Internet Draft, Internet Engineering Task Force, draft-nakhjiri-seamoby-text-ct-00.txt, Nov 2001.
- [OIK-01] K. Oikonomou, I. Tenidis, I. Stavrakakis, "A Mechanism to Enable Differentiated Services QoS in HIPERLAN/2", 8<sup>th</sup> IEEE International Conference on Telecommunications (ICT'01), Bucharest, Romania, June 2001.
- [PAC-02] Sangheon Pack and Yanghee Choi, "Pre-Authenticated Fast Handoff in a Public Wireless LAN based on IEEE 802.1x Model, " IFIP TC6 Personal Wireless Communications 2002, Singapore, October 2002.
- [PADM-96] R. Caceres and V. Padmanabhan, Fast and Scalable Handoffs for Wireless Internetworks", Proceedings of ACM Mobicom, November 1996.
- [PAR-00] V. Park, M. S. Corson. "Temporally-Ordered Routing Algorithm (TORA) Version 1 Functional Specification", Internet-Draft, draft-ietf-manet-tora-spec-03.txt, Nov 2000.
- [PERK-96] C. Perkins. "Mobile-IP Local Registration with Hierarchical Foreign Agents". draft-perkins-mobileip-hierfa-00.txt Feb 1996.
- [PO-CH'03] C. Politis, K. Chew, R. Tafazolli, "Multilayer Mobility Management for All-IP Networks: Pure SIP vs. Hybrid SIP/Mobile IP", IEEE VTC Spring Conference, Jeju, Korea, 21-24th April 2003.
- [POL-01] C. Politis, R. Tafazolli, "MAC Protocol for All-IP Hiperlan2", IEEE VTC Spring Conference, Rhodes island, Greece, 6-9 May 2001.
- [POL-02] C. Politis, R. Tafazolli, "MAC Protocol for Supporting QoS in All-IP HiperLAN2", 4th European Wireless Conference (EW'02), Florence, Italy, 25-28th February 2002.
- [POL-ALM] C. Politis, V. Almenar, R. Tafazolli, "Physical Layer Techniques for Optimum Performance of HiperLAN2 Systems", IEEE 3Gwireless'02/WAS'02 Conference, San Francisco, USA, 29-31st May 2002.
- [POL-DAG] C. Politis, T. Dagiuklas, "On the performance of Mobility Management Schemes in all-IP Network Architectures", International HET-NET'03 Conference, Ilkley, U.K, 21-23 July, 2003.
- [POL-TAF] C. Politis, R. Tafazolli, "Connection Admission Control for Supporting QoS in IP-Based HiperLAN2 Networks" IEEE International Symposium on Advances in Wireless Communications (ISWC'02), Victoria, BC, Canada, 23-24 September 2002.
- [RAA-01] J. Manner, K. Raatikainen, "Extended Quality-of-Service for Mobile Networks," IEEE/IFIP 9<sup>th</sup> International Workshop on Quality of Service (IWQoS 2001) Karlsruhe, Germany, June 6 - 8, 2001. Published in the Springer LNCS Series, Vol. 2092, pp. 275-280.
- [RADIUS] Remote Authentication Dial In User Service (RADIUS), IETF RFC 2865, <http://www.ietf.org/rfc/rfc2865.txt>

- 
- [RAM-00] R. Ramjee, T. La Porta. "IP micro-mobility support using HAWAII". draft-ietf-mobileip-hawaii-01.txt, July 2000.
  - [RAPP-00] J. Rapp, "New approach for system performance study in HIPERLAN/2", in Proc. of European Wireless Conference 2000, Dresden, Germany.
  - [REB-02] Y. Rebahi, D. Sisalem, "AAA Management in the Internet for Wireless and 3G Users", Wireless World Research Forum (WWRF), 7th meeting, Eindhoven, Netherlands, December 2002.
  - [RFC-2210] J. Wroclawski, "The Use of RSVP with IETF Integrated Services." RFC2210, Sept. 1997.
  - [RFC-2211] IETF, RFC 2211, "Specification of the Controlled-Load Network Element Service", September 1997.
  - [RFC-2212] IETF, RFC 2212, "Specification of Guaranteed Quality of Service", September 1997.
  - [RFC-2474] IETF, RFC 2474, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", December 1998.
  - [RFC-2903] Generic AAA Architecture, IETF RFC 2903, <http://www.ietf.org/rfc/rfc2903.txt>
  - [RFC-2904] AAA Authorization Framework, IETF RFC 2904, <http://www.ietf.org/rfc/rfc2904.txt>
  - [RFC-2905] AAA Authorization Application Examples, IETF RFC 2905, <http://www.ietf.org/rfc/rfc2905.txt>
  - [ROS-01] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Jonston, J. Peterson, R. Sparks, M. Handley, E. Schooler, "SIP: Session Initiation Protocol", draft-ietf-sip-rfc2543bis-05.txt, work in progress, October 2001.
  - [ROS-02] J. Rosenberg, J. Weinberger, C. Huitema, R. Mahy, "STUN-Simple Traversal of UDP Through NATs", IETF draft, draft-ietf-midcom-stun-00.txt, April 2002.
  - [SARI-00] B. Sarikaya, H. Haverinen, J.T. Malinen, V. Magret, "Mobile IPv6 Regional Paging", draft-sarikaya-mobileip\_hmipv6rp-00.txt, November 2000.
  - [SES-95] S. Seshan, H. Balakrishnan and R.H. Katz, "Handoffs in Cellular Wireless Networks: The Daedalus Implementation and Experience", ACM/Baltzer Journal on Wireless Networks, Vol. 1, No. 4, December 1995.
  - [SMBY] SeaMoby WG web page: <http://www.ietf.org/html.charters/seamoby-charter.html>
  - [SOL-00] Hesham Soliman, Claude Castellucia, Karim El-Malki, Ludovic Bellier, "Hierarchical MIPv6 mobility management", IETF draft, September 2000.
  - [STR-2.1] "4G scenarios and system requirements", IST-2001-38354 STRIKE, deliverable D2.1, April 2003.  
<http://ist.strike.org/>
  - [TACACS] Access Control Protocol, or TACACS, IETF RFC 1492, <http://www.ietf.org/rfc/rfc1492.txt>
  - [TS-23.107] 3GPP TS 23.107 V3.5.0 (2000-12) 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; QoS Concept and Architecture (Release 1999).

- [VAK-01] F. Vakil, A. Dutta, J-C. Chen, M. Tauil, S. Baba, N. Nakajima, H. Schulzrinne, "Supporting Mobility for TCP with SIP", draf-itsumo-sipping-mobility-tcp-00.txt, work in progress, June 2001.
- [WAK-95] Wakeman, I., Ghosh, A., Crowcroft, J., Jacobson, V., and Floyd, S., "Implementing Real Time Packet Forwarding Policies using Streams", Usenix 1995 Technical Conference, January 1995, New Orleans, Louisiana.
- [WED-99] C. E. Wedlund, H. Schulzrinne, "Mobility Support using SIP", Second ACM/IEEE International Conference on Wireless and Mobile Multimedia (WoWMoM'99), Seattle, Washington, August 1999.
- [WIND-99] "Modem requirements for baseband, RF/IF subsystem and DLC/MAC layer", IST-Windflex deliverable D2.1, 1999.  
<http://www.vtt.fi/ele/research/els/projects/windflex.htm>
- [WIS-01] Dave Wisely et al, "BRAIN-an architecture for a broadband radio access network of the next generation", Wireless Communications & Mobile Computing Journal, Jan-Mar 2001, vol.1 No. 1, pp. 55-75.
- [YAT-82] Y. Yatsuzuka, "Highly Sensitive Speech Detector and High-Speed Voice band Discrimination in SDI-ADPCM system", IEEE Transactions on Communications, No 4, Vol. 30, pp. 739—750, April 1982.
- [ZAC-00] Zach D. Shelby, Dionisios Gatzounas, Andrew Campbell, Chieh-Yih Wan, "Cellular IPv6 " IETF draft, November 2000.