Quality-Oriented Mobility Management for Multimedia Content Delivery to Mobile Users

Bogdan Ciubotaru-Petrescu

M.Sc. in Computer Science

A Dissertation submitted in fulfilment of the

requirements for the award of

Doctor of Philosophy (Ph.D.)

to the

Dublin City University

Faculty of Engineering and Computing, School of Electronic Engineering

Supervisor: Dr. Gabriel-Miro Muntean

January 14, 2011

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Student ID: 57112266

Signed: ______________________

Date: ________________________

To my lovely wife and to my dear parents

Acknowledgements

First of all I want to express my gratitude to my supervisor **Dr. Gabriel-Miro Muntean** for supporting and guiding me during the last three years spent in Dublin City University as a PhD research student. I have learned much from his vast technical and academic expertise and knowledge which he kindly shared with me during the time when I have worked close to him. I also want to thank him for the great support and advise he offered me in developing my professional track record towards reaching my career goals.

I also want to thank the staff members of PEL: **Professor Liam Murphy**, **Professor John Murphy**, **Dr. Jennifer McManis**, **Dr. Olga Ormond** and **Dr. Hrishikesh Venkataraman** and my colleagues **Ramona**, **Sabine**, **Kevin**, **Seung-Bum**, **Martin**, **Zenhui**, **Lejila** and **Arthy** for their help, support and advice.

I cannot forget the important contribution of the lecturers and professors from the **Computer Science Department of ”Politehnica” University** in **Timisoara, Romania** to my professional development and my education. I especially want to thank **Professor Dr. Mircea Stratulat**, **Dr. Mihai V. Micea** and **Dr. Dan Pescaru** for their support and valuable advice during the period I spent with their research groups.

At last but not the least I very much thank to the technical staff from School of Electronic Engineering, Dublin City University for their valuable support.

Dublin, January 2011

Bogdan Ciubotaru

List of publications

Bogdan Ciubotaru and Gabriel-Miro Muntean. "SASHA - A Quality-Oriented Handover Algorithm for Multimedia Content Delivery to Mobile Users" *IEEE Transactions on Broadcasting*, Special Issue on IPTV, vol. 55, no. 2, June 2009.

Bogdan Ciubotaru, Gabriel-Miro Muntean and Ghiorghita Ghinea. "Objective Assessment of Region of Interest-Aware Adaptive Multimedia Streaming Quality" *IEEE Transactions on Broadcasting*, vol. 55, no. 1, June 2009.

Bogdan Ciubotaru, Ghiorghita Ghinea and Gabriel-Miro Muntean. "Subjective Assessment of Region of Interest-aware Adaptive Multimedia Streaming Quality" submitted to *ACM Transactions on Multimedia Computing Communications and Applications*, IGI Global, Ed. Dr. Susmit Bagchi.

Bogdan Ciubotaru and Gabriel-Miro Muntean. "Quality-Oriented Mobility Management for Ubiquitous Multimedia Content Delivery" accepted for publication in *Ubiquitous Multimedia and Mobile Agents: Models and Implementations*.

Bogdan Ciubotaru and Gabriel-Miro Muntean. "Smooth Adaptive Soft Handover Algorithm for Multimedia Streaming over Wireless Networks" *IEEE Wireless Communications and Networking Conference (WCNC 2009)*, Budapest, Hungary, April 2009.

Bogdan Ciubotaru and Gabriel-Miro Muntean. "QMS-Quality of Multimedia Streaming Metric for Soft-Handover in Heterogeneous Wireless Environments" *IEEE Conference on Local Computer Networks (LCN 2010)*, Denver, Colorado, U.S.A., October 2010.

Bogdan Ciubotaru and Gabriel-Miro Muntean. “Quality of Multimedia Streaming-oriented Handover Management Solution for Mobile Applications” *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB 2009)*, Bilbao, Spain, May 2009.

Bogdan Ciubotaru and Gabriel-Miro Muntean. “M3S - Multimedia Mobility Management and Load Balancing in Wireless Broadcast Networks” *IET China-Ireland International Conference on Information and Communications Technology (CIICT 2008)*, Beijing, China, September 2008.

Awards

2010 Scott Helt memorial award for the best paper published in IEEE Transactions on Broadcasting.

Bogdan Ciubotaru, Gabriel-Miro Muntean and Ghiorghita Ghinea. "Objective Assessment of Region of Interest-Aware Adaptive Multimedia Streaming Quality" *IEEE Transactions on Broadcasting*, vol. 55, no. 1, June 2009.

Abstract

The heterogeneous wireless networking environment determined by the latest developments in wireless access technologies promises a high level of communication resources for mobile computational devices. Although the communication resources provided, especially referring to bandwidth, enable multimedia streaming to mobile users, maintaining a high user perceived quality is still a challenging task. The main factors which affect quality in multimedia streaming over wireless networks are mainly the error-prone nature of the wireless channels and the user mobility. These factors determine a high level of dynamics of wireless communication resources, namely variations in throughput and packet loss as well as network availability and delays in delivering the data packets. Under these conditions maintaining a high level of quality, as perceived by the user, requires a quality oriented mobility management scheme. Consequently we propose the Smooth Adaptive Soft-Handover Algorithm, a novel quality oriented handover management scheme which unlike other similar solutions, smoothly transfer the data traffic from one network to another using multiple simultaneous connections. To estimate the capacity of each connection the novel Quality of Multimedia Streaming metric is proposed. The QMS metric aims at offering maximum flexibility and efficiency allowing the applications to fine tune the behavior of the handover algorithm. The current simulation-based performance evaluation clearly shows the better performance of the proposed Smooth Adaptive Soft-Handover Algorithm as compared with other handover solutions. The evaluation was performed in various scenarios including multiple mobile hosts performing handover simultaneously, wireless networks with variable overlapping areas, and various network congestion levels.

Contents

List of Tables

List of Figures

List of Abbreviations

3GPP: Third Generation Partnership Project

ABC: Always Best Connected

AHP: Analytic Hierarchy Process

AMR: Adaptive Multirate

AP: Access Point

AVC: Advanced Video Coding

BCMCS: Broadcast and Multicast Services

BER: Bit Error Rate

BSA: Basic Service Area

BSS: Basic Service Set

CBR: Constant Bit Rate

CDMA: Code Division Multiple Access

CIR: Carrier-to-Interferences Ratio

CN: Corresponding Nodes

DCCP: Datagram Congestion Control Protocol

DCF: Distributed Coordination Function

DCT: Discrete Cosine Transform

DS: Distribution System

DVB: Digital Video Broadcasting

EDCF: Enhanced Distributed Coordination Function

EDGE: Enhanced Data Rates for GSM Evolution

ESS: Extended Service Set

FA: Foreign-Agent

FIS: Fuzzy Inference System

FLUTE: File Delivery over Unidirectional Transport

GOP: Group of Pictures

GPRS: General Packet Radio Service

GPS: Global Positioning System

GRA: Grey Relational Analysis

GSM: Global System for Mobile Communications

HA: Home Agent

HAP: High Altitude Platforms

HAWAII: Handoff-Aware Wireless Access Internet Infrastructure

HCF: Coordination Function

HD: High Definition

HIP: Host Identity Protocol

HiperACCESS: High Performance Radio Access

HiperMAN: High Performance Radio Metropolitan Access Network

HMIPv6: Hierarchical Mobile IPv6

HSPA: High-Speed Packet Access

IAPP: Inter Access Point Protocol

IEC: International Electrotechnical Commission

IMS: IP Multimedia Subsystem

IP Datacast: IP-based Data Broadcast

IP: Internet Protocol

ISO: International Organization for Standardization

ITU: International Telecommunication Union

JPEG: Joint Photographic Experts Group

LAN: Local Area Network

LMDS: Local Multipoint Distribution System

LOS: Line of Sight

LoWPAN: Low power Wireless Personal Area Networks

LTE: Long-Term Evolution

MAC: Medium Access Control

MADM: Multiple Attribute Decision Making

MAN: Metropolitan Area Network

MBMS: Multimedia Broadcast Multicast Service

MDC: Multiple Description Coding

MEHH: Modified Elman Neural Network

MH: Mobile Host

MICS: Media-Independent Command Service

MIES: Media-Independent Event Service

MIIS: Media-Independent Information Service

MIP: Mobile IP

MIPv4: Mobile IPv4

MIPv6: Mobile IPv6

MMDS: Multichannel Multipoint Distribution System

MMS: Multimedia Message Service

MN: Mobile Node

MOS: Mean Opinion Score

MPEG: Moving Picture Experts Group

MPQM: Moving Picture Quality Metric

NGSS: Next Generation Satellite System

NLOS: Non-Line of Sight

PAN: Personal Area Networks

PAR: Picture Appraisal Metric

PCF: Point Coordination Function

PDA: Personal Digital Assistants

PDF: Policy Decision Function

PDM: Perceptual Distortion Metric

PSNR: Peak-Signal-to-Noise-Ratio

PSS: Packet-Switched Streaming

QMS: Quality of Multimedia Streaming

QOAS: Quality Oriented Adaptation Scheme

QoE: Quality of Experience

QoS: Quality of Service

RoI: Region of Interest

RSS: Received Signal Strength

RTP: Real Time Transport Protocol

RTSP: Real Time Streaming Protocol

RTT: Round-Trip Time

SASHA: Smooth-Adaptive Soft Hanover Algorithm

SAW: Simple Additive Weighting

SCTP: Stream Control Transmission Protocol Mobile

SD: Standard Definition

SIF: Standard Input Format

SINR: Signal to Interference plus Noise Ratio

SIP: Session Initiation Protocol

SIR: Signal-to-Interferences Ratio

SMS: Standard Message Service

SNR: Signal to Noise Ratio

SSIM: Structural Similarity Index

TCP: Transmission Control Protocol

TDMA: Time Division, Multiple Access

TFRC: TCP-Friendly Rate Control

TOPSIS: Technique for Order Preference by Similarity to Ideal Solution

UDP: User Datagram Protocol

UMB: Ultra Mobile Broadband

UMTS: Universal Mobile Telecommunications System

UWB: Ultra-Wideband

VDP: Visual Differences Predictor

VITS: Vertical Interval Test Signals

VLC: Variable-Length Codes

VQM: Video Quality Metric

WAN: Wide Area Network

WCDMA: Wideband CDMA

WiMAX: Worldwide Interoperability for Microwave Access

WLAN: Wireless Local Area Networks

WMAN: Wireless Metropolitan Area Networks

WPAN: Wireless Personal Area Networks

WRAN: Wireless Regional Area Network

WSNR: Weighted-Signal-to-Noise-Ratio

WWAN: Wireless Wide Area Networks

Chapter 1

Introduction

1.1 Research Motivation

Mobile devices are increasing their capabilities in terms of processing power, memory, graphical display and most importantly in terms of wireless communications. If early days of mobile communications involved analog voice services only, today digital data processing and communication represent the norm. Consequently mobile computing became extremely popular among users who started to access services and applications which once made the subject of desktop computing only.

This trend is not only sustained by the above-mentioned capabilities of the existing mobile devices but also by the wide range of broadband wireless communication available to mobile users. Various wireless communication technologies have been developed, ranging from Wide Area Network (WAN) technologies like DVB to Personal Area Networks (PAN) based on technologies like Bluetooth and including the very popular Local Area Network (LAN) supported by the WiFi technology and expanding Metropolitan Area Network (MAN) supported by WiMAX for example.

All these wireless networks are rapidly spreading and cover, in an overlapping manner, various geographical areas from individual households to cities and entire regions forming a *heterogeneous wireless networking environment*.

The wide range of capabilities available on the mobile devices combined with the constantly improving data services broadens the types of applications and services offered to the users on their mobile devices. Among these, *multimedia streaming applications* are by far the most popular.

Multimedia applications have several characteristics which make them challenging for any data communication service provider. This type of application requires high bandwidth, imposes real-time constraints and their service quality as perceived by the user is very sensitive to network quality of service (QoS) variations.

The current best-effort Internet access technologies are still facing challenges in providing and most importantly sustaining the requirements of multimedia applications. Moreover the advance of mobile devices and mobile computing opened the door for the development of the future *mobile Internet* which poses more challenges, among which the most important are user mobility and the error prone nature of the wireless environment.

In the context of a mobile Internet the user is roaming through a geographical area covered by various wireless networks. As with the fixed Internet, the user expects to receive a continuous and stable data communication service which ensures an effective access to data. User mobility combined with the dynamics of the wireless networking environment often determines disruptions of the user's active data session impacting negatively on the overall quality of the service received.

All these lead to the necessity of an extra service to be added to the current Internet architecture: *mobility management service*.

Mobility management is responsible for keeping mobile devices reachable by other hosts with Internet connectivity (location management) and most important for providing a continuous and uninterrupted data flow to the existing user data sessions (handover management).

Handover management is the most important component of the mobility management service. It is in charge with offering mobile devices continuous data connectivity and has the greatest impact on user perceived quality of service, extremely important for service's success.

1.2 Problem Statement

In the context of handover management, active research activity determined a wide range of solutions to be proposed. Currently available handover algorithms are such designed to transfer the data sessions from one wireless network to another. There are various types of handover, such as soft-handover, hard-handover, vertical handover, horizontal handover, etc. Each particular solution has different performance in terms of handover latency and packet loss levels. Especially the soft-handover algorithms present good performance in terms of the above mentioned parameters.

However relying on only one network to convey the entire data traffic presents several shortcomings. In certain circumstances due to distance, environmental factors or network congestion, none of the available wireless networks may be able to provide alone the required bandwidth for supporting high user perceived quality, especially in the context of delivering high bitrate multimedia content.

Consequently the traditional handover algorithms will find impossible to select an appropriate network and maintain the expected QoS level. Moreover using only one network may leave valuable communication resources underutilized just because they do not provide the entire amount of bandwidth required. Last but not least aggregating the bandwidth from multiple networks increases the overall throughput and improves the robustness of the data communication service.

Most existing traffic distribution and throughput aggregation techniques proposed for the wired networks and for wireless communications especially in the context of load balancing, do not consider the mobility aspect and user quality of experience.

Consequently the goal of this research project is to propose a handover management solution which efficiently exploits all the available communication resources, provide smooth data transfer with high QoS levels, and most important consider user experience by involving video content quality in the decision making process. Providing an efficient load balancing and cost effective use of wireless resources is also among the goals of this research work.

1.3 Solution Overview

This thesis introduces the Smooth-Adaptive Soft Hanover Algorithm (SASHA), a novel handover management algorithm for multimedia applications in heterogeneous wireless environments.

SASHA performs handover by smoothly distributing and balancing the application data traffic over multiple simultaneous connections. Each connection is established over a separate wireless access network, among the ones available in every location.

The data distribution over the existing networks is made based on the QMS metric, which is computed for each connection separately.

Quality of Multimedia Streaming (QMS) is a flexible and comprehensive metric for evaluating network capacity to efficiently transport multimedia content.

The QMS scores give SASHA an estimation of how much traffic each individual connection can efficiently carry. Using this estimation SASHA calculates the corresponding data rates and splits the main application traffic into streams, each stream being allocated to a distinct connection.

Any variation in network conditions will determine variations in QMS scores, regularly computed by the SASHA server-side, which will further determine a re-balancing of the traffic distribution. In this a way SASHA performs handover, allowing for user mobility, efficiently distributing the load over the existing wireless networks, improving communication resources usage and at the same time aggregating the throughput from multiple networks and providing the user with more data communication resources.

QMS is a flexible and comprehensive metric which incorporates several parameters namely QoS, monetary cost, energy consumption and user quality of experience (QoE). Weights are used to allow for pondering the impact each component has on the overall QMS score computed for a particular connection and on the behavior of the handover management algorithm.

This approach overcomes the shortcoming of the current handover management solutions and efficiently combines mobility management with load balancing and throughput

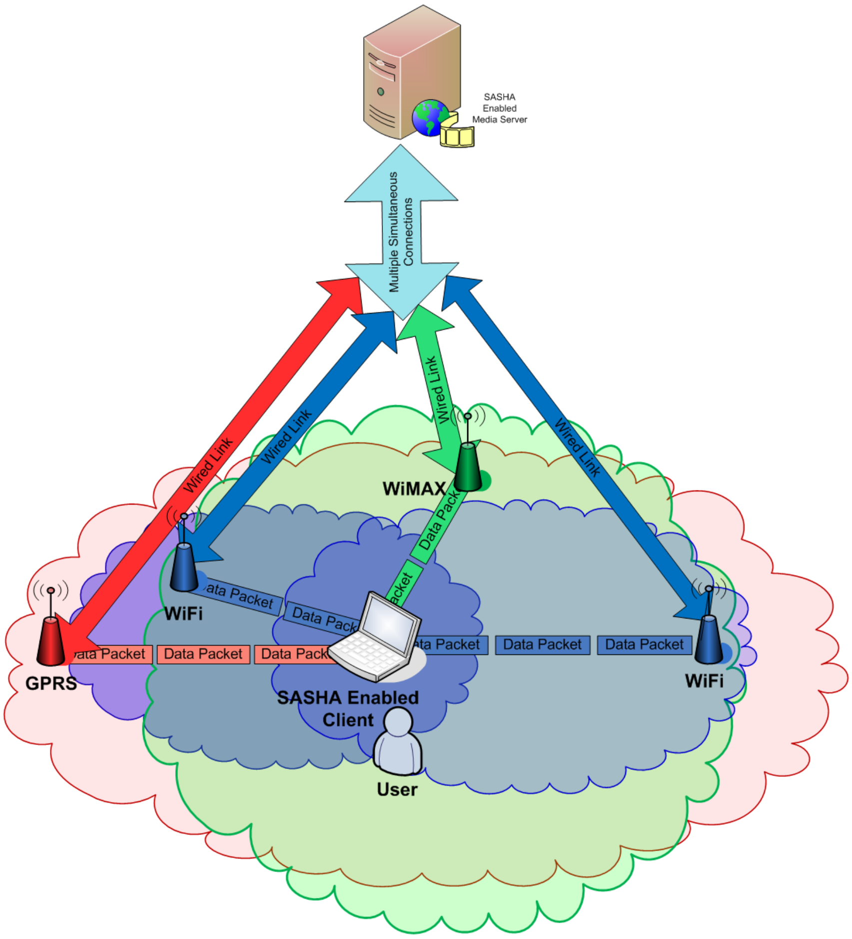

Figure 1.1: Solution conceptual overview

aggregation.

Figure 1.1 schematically presents SASHA handover management concept. A heterogeneous wireless environment is presented were mobile users have for instance four wireless networks available to connect to. Three distinct wireless communication technologies are employed: WiFi, WiMAX and GPRS. SASHA establishes a data connection over each network separately and the traffic is distributed over them. QMS metric is computed periodically for each connection independently and the amount of traffic (bitrate) to be transported by each network is estimated based on the QMS scores.

1.4 Contributions

This thesis and the research activity performed provide the following contributions to the advancement of the current state of the art:

- Proposal of SASHA, a novel handover management algorithm, which
  - Performs soft-handover through load balancing.
  - Makes efficient use of all existing wireless communication resources.
- Introduction of QMS, a novel quality metric, which
  - Is comprehensive and flexible metric combining QoS, QoE, monetary cost and energy components.
  - Represents general purpose metric with a high level of flexibility.
  - Was tested for handover and traffic distribution.
- A new quality-aware mobile data communication architecture for multimedia applications, which
  - Introduces a data communication architecture for multimedia applications including quality aware mobility management.
  - Supports integration with adaptive streaming strategies and algorithms.
  - Provides integration of handover management, load balancing and throughput aggregation into a single modular architecture.

The results of this research activity were properly disseminated and accepted by the international research community through book chapters, journal papers and conference papers.

1.5 Thesis Structure

This thesis is structured as follows. Chapter 2 discusses the background technologies related to wireless data communications, multimedia streaming and mobile device capabilities.

Chapter 3 presents related works in the context of SASHA and the QMS metric. Chapter 4 introduces the QMS metric and discusses its parameters. Chapter 5 presents SASHA, introducing its architecture and algorithms. Chapter 6 evaluates QMS and suggests default values for its parameters and provides a methodology for experimental QMS component weight tuning in different application contexts. Chapter 7 discusses SASHA performance evaluation and analyses testing results. The final chapter presents the conclusions and future work.

Chapter 2

Background Technologies

2.1 Introduction

Ever since they were first tested in the late 19th century [2] wireless radio communications were used to convey information between peers separated by large distances. The benefits of wireless communications in comparison with the wired ones combined with the development of portable computational devices including laptops, netbooks, Smartphones and Personal Digital Assistants (PDA), made the associated techniques increasingly popular ever since their invention more than 100 years ago.

Internet applications and services including e-mail, web browsing and multimedia applications are now present on the portable devices increasing the pressure on the wireless networking technologies in terms of their robustness, availability and Quality of Service (QoS) provision.

Among the above mentioned applications and services the ones involving multimedia content delivery challenge the most wireless networking environments mainly due to their real-time constrains, bandwidth requirements and sensitivity to network QoS variations.

Consequently much research and development effort was spent on algorithms and techniques for transporting high quality multimedia content over wireless networks.

Next standard solutions for content delivery over wireless networks, video compression, video delivery and quality assessment are presented in details as well as typical user devices.

2.2 Wireless Networking Technologies

Various wireless data communication technologies have been developed, each having special characteristics in terms of bandwidth, QoS provisioning, range and mobility support.

Wireless data networks are rapidly evolving and gaining popularity. These networks are mostly used as last mile access network for the wired core networks or to interconnect multiple devices directly or as part of wireless ad-hoc networks. Wireless data networks allow users with wireless enabled mobile devices to access the core network and benefit of Internet connectivity or to interact with other devices in their proximity spaces. Various wireless communication technologies will be presented in the following sections grouped according to their coverage area in the following categories.

Wireless Personal Area Networks (WPAN) cover areas up to tens of meters and are mostly used to interconnect devices over short distances. Wireless Local Area Networks (WLAN) usually cover small areas, in the range of tens to hundreds of meters in radius and are mostly for home and office use. Wireless Metropolitan Area Networks (WMAN) usually cover wider areas such as large as cities. Wireless Wide Area Networks (WWAN) cover even larger areas, larger than a single city, offering support to regions.

2.2.1 Wireless Personal Area Networks

The success of WPANs was sustained by the development of various computational devices starting from desktop PCs and laptops to portable devices and various peripherals which need to be interconnected in order to exchange various types of data.

Several standards and protocols were developed or are under development for short range wireless communication.

Bluetooth [3] specifications were developed by the Bluetooth Special Interest Group (Bluetooth SIG) as a wireless communication technology intended to interconnect various portable devices and their accessories. Bluetooth uses the 2.4 GHz band and the first version offered data rates of up to 1 Mb/s. Bluetooth v2.0 offers increased data rates of up to 3Mb/s, but in the future rates are expected to increase between 53Mbps to 480 Mbps.

IEEE 802.15.1 [4] standard for Wireless Personal Area Networks was developed based on Bluetooth v1.1 [5].

IEEE 802.15.4 [6] specifies the MAC and PHY layer for low-range, low-power wireless network communications. Based on this standard, the Zigbee protocol [7] defines the network layer specialized on ad-hoc networking and the application layer targeting wireless sensor networks as well as other monitoring and control applications. IEEE 802.15.4/Zigbee offers data rates up to 250 Kbps in the 2.4GHz frequency band.

WiMedia Alliance has defined the specifications for ultra-wideband (UWB) wireless communication technology supporting wide range of data rates from 53Mbps to 480 Mbps over short ranges and using low power transceivers [8]. The PHY/MAC protocols developed by WiMedia became the ECMA 368 standard [9] and later on was accepted as ISO/IEC 26907 standard [10].

IPv6 over Low power Wireless Personal Area Networks (LoWPAN), is a protocol suite design to allow IPv6 packets to be transfered over low power wireless networks, specifically IEEE 802.15.4 [11].

Wibree is an ultra-low-power wireless network communication technology with ranges of up to 10 m and data rates of 1Mb/s [5].

Various other protocols and technologies can be found in the literature, some are already commercially available or are currently under development. Although most of them rely on low power, low data rate wireless technologies, some are already targeting high bitrates (Bluetooth - UWB). Additionally as presented above, protocols exist to allow IP networks to be set up using these WPAN technologies.

Consequently WPANs offer good support for wireless multimedia content delivery to mobile devices in a short range environment.

2.2.2 Wireless Local Area Networks

For WLANs the IEEE 802.11 family (WiFi) of standards was developed and includes various extensions which address different issues including higher bit rates, QoS support, and security. This technology gained much popularity mainly because of its low deployment and maintenance costs as well as its high bit rates.

The standards for wireless access networks usually cover the physical layer and the medium access control protocol (MAC) sub-layer.

The original IEEE 802.11 standard first released in 1997 [12] supports data rates up to 2Mbps and was initially developed for best effort traffic only.

Each host connected to a certain IEEE 802.11 access point shares the wireless medium with the other mobile hosts associated with the same access point. This leads to race conditions for medium access which determine high collision rates and consequently low data rates especially when the number of mobile hosts involved in simultaneous data communications increases.

The IEEE 802.11 MAC layer provides mechanisms for medium access coordination, including the Distributed Coordination Function (DCF) and the partially centralized Point Coordination Function (PCF).

A group of mobile stations connected to a single Access Point (AP) form the basic building block defined by this standard as a Basic Service Set (BSS). The geographical area covered by a BSS is called a Basic Service Area (BSA). Connecting several BSSs through a Distribution System (DS) determines an Extended Service Set (ESS).

The first IEEE 802.11 extension, IEEE 802.11b [13] increased the maximum data rate to 11Mbps. The data rate is further increased to 54Mbps in the IEEE 802.11a and IEEE 802.11g standards [14, 15].

Maintaining high QoS levels by using the two coordination methods, DCF and PCF is difficult, thus QoS preservation enhancements for IEEE 802.11 MAC layer were standardized by IEEE 802.11e [16].

Consequently two new mechanisms are described by the new standard, named the Hybrid Coordination Function (HCF) and the Enhanced Distributed Coordination Function (EDCF).

HCF is based on PCF and EDCF relies on its implementation on DCF. Further enhancements brought by this standard are block acknowledgments which allows acknowledging more then one MAC frame by sending only one acknowledge packet and *No Ack* which allows time critical data frames not to be acknowledged. To enhance QoS provisioning for time sensitive and bandwidth hungry applications, traffic prioritisation was proposed for IEEE 802.11 [17]. Four traffic categories are defined:voice, video, best effort and background, and in this order the IEEE 802.11e offers prioritisation support.

The emerging IEEE 802.11n standard [18] aims at providing even higher bitrates, of up to 600Mbps. The data rate enhancement approach of IEEE 802.11n is oriented on improving MAC layer techniques unlike other IEEE 802.11 which aims at increasing the data rates at physical layer. IEEE 802.11n uses the same QoS support techniques proposed for IEEE 802.11e.

The currently under study **IEEE 802.11 VHT (Very High Throughput)** [19] aims at offering data rates of up to 1Gbps for low velocity mobile hosts.

The IEEE 802.11 family does not support host mobility except for the IEEE 802.11s standard [20, 21] which specifies support for mesh networks and which addresses host mobility within the mesh network.

IEEE 802.11p standardizes wireless access in vehicular environments which represents a short to medium range communication service providing high data transfer rates for road-side to vehicle or vehicle to vehicle data communications.

The IEEE 802.11 family groups several other standards addressing various aspects of wireless data networks including security, management and compatibility. A more detailed overview of IEEE 802.11 family of standards can be found in [22].

Table 2.1 and Table 2.2 summarizes the characteristics of the most important IEEE 802.11 standards and extensions, including maximum data rates and frequencies.

2.2.3 Wireless Metropolitan Area Networks

WMANs were developed to cover whole cities and to interconnect LANs or WLANs as well as individual users, both static and mobile.

WMANs use two type of connectivity, **line of sight (LOS)**, when there is not obstacle between the sender and the receiver and **non-line of sight (NLOS)** when the sender and receiver cannot see each other in a straight line.

Standard	Bitrate	Frequency	Description
802.11	1 Mb/s (2 Mb/s)	2.4 GHz	Initial standard
802.11b	11 Mb/s	2.4 GHz	Data rate enhancement
802.11a	54 Mb/s	5 GHz	Data rate enhancement
802.11g	54 Mb/s	2.4 GHz	Backward compatibility
802.11n	600 Mb/s	2.4 and 5 GHz	Data rate enhancement
802.11p	27 Mb/s	5.9 GHz	Vehicular communication
802.11ac (VHT)	1Gb/s	< 6 GHz	Data rate enhancement
802.11ad (VHT)	1Gb/s	60 GHz	Data rate enhancement

Table 2.1: IEEE 802.11 family of standards

Standard	Description
802.11e	Extension for QoS support
802.11aa	Extension for audio/video streaming
802.11r	Handoff support
802.11s	Transparent multi-hop operation (Mesh)
802.11u	Interworking with external networks (cellular)

Table 2.2: IEEE 802.11 family of standards

Companies producing equipment for WMANs have formed the Worldwide Interoperability for Microwave Access (WiMAX) forum concerned with the standardization and technology development in this area of wireless communications.

Specific to WMANs is the IEEE 802.16 family of standards. The IEEE 802.16 is based on two systems, the Multichannel Multipoint Distribution System (MMDS) and Local Multipoint Distribution System (LMDS) [23].

The MMDS system offers better coverage (typical cell radius is 50 km) but the throughput is quite low, between 0.5Mbps and 30 Mbps. With less coverage (3 to 5 km) LMDS provides higher bandwidth, 34Mbps to 38 Mbps with an increase to 36 Gbps for the newer versions.

IEEE 802.16 provides QoS provisioning support. This is achieved mainly trough connections, service flows and service scheduling. QoS provisioning is negotiated at the initiation of the session and QoS requirements are mapped on the QoS parameters in the IEEE 802.16 MAC layer. Mobility is supported in the new IEEE 802.16e standard which permits mobile hosts to change their base station while the data connection is still active. Both soft and hard handover mechanisms are supported, while several enhancement solutions are being proposed [24].

WiMAX becomes increasingly popular, as a wireless broadband solution, with several types of mobile devices being already equipped with WiMAX interfaces.

High Performance Radio Access (HiperACCESS) standardized by ETSI offers LOS broadband wireless access using frequencies between 11 and 43.5 GHz. The typical cell radius is 5km and the data rates per cell ranges between 25Mbps and 100 Mbps [25].

High Performance Radio Metropolitan Access Network (HiperMAN), also standardized by ETSI, offers broadband connectivity targeting residential and small office areas. HiperMAN works in the frequency bands below 11 GHz and offers NLOS connectivity with aggregated data rates of up to 25 Mbps [26].

WiBro is another WMAN solution developed in Korea which offers broadband connectivity to both stationary and mobile users. WiBro operates in the 2.3 - 2.4 GHz frequency band offering data rates of up to 50 Mbps [27]. The major advantage of WiBro over the other WMAN technologies is the mobility feature which is very well developed.

High Altitude Platforms (HAP) [28] use a quasi-stationary aerial platform equipped with wireless transceivers offering broadband wireless access with data rates of 120 Mbps or up to 10 Gbps in some configurations. This type of wireless technology offers good coverage with better LOS connections.

IEEE 802.22 Wireless Regional Area Network (WRAN) offers data rates up to 18Mbps for rural and remote areas using the unoccupied TV channels between 54 and 862 MHz [29].

Cellular networks which initially offered only voice services are already offering broadband Internet access through the current third generation (3G) and the future fourth generation (4G) networks.

The first to provide mobile communication services were the first generation (1G) cellular networks which supported only analog voice calls and very limited data applications. This technology was replaced by the second generation cellular networks (2G) which is entirely digital and apart from voice communication also supports low bit rate data communication in form of Short Message Service, Multimedia Message Service.

The current cellular network technologies can be grouped in two main families: Global System for Mobile Communications (GSM) based on time division, multiple access (TDMA) and code division multiple access (CDMA) [30].

The maximum bit rate in GSM being 9.6 kbps, throughput enhancement solutions have been developed for this standard including the 2.5G General Packet Radio Service (GPRS) and the 2.75G Enhanced Data Rates for GSM Evolution (EDGE).

GPRS supports theoretical data rates around 114 Kbps, but in reality the throughput reaches values around 40 Kbps only. EDGE is the first to open the door for multimedia applications over cellular networks. It supports theoretical throughputs around 400Kbps.

The third generation cellular network (3G) supports voice and continues the improvement of the data communication rates.

In the GSM category the Universal Mobile Telecommunications System (UMTS) makes use of wideband CDMA (WCDMA) and High-Speed Packet Access (HSPA) technologies in order to support bit rates of up to 2Mbps.

The CDMA-based standards for 3G networks include the CDMA2000 family among which CDMA 1xRTT, supports average data rate of 40-80 kbps with peak data rate of 150 kbps. CDMA 2000 1xEV-DO supports only data communications with maximum data rates of 2.4 Mbps.

As the demand for higher bandwidth and QoS support is increasing with the increased popularity of bandwidth hungry, real-time applications, the forth generation network (4G) is in process of being defined and standardized.

The technologies which are principal candidates for 4G networks are Long-Term Evolution (LTE), Ultra Mobile Broadband (UMB) and 802.16m (WiMAX II) [30].

LTE is developed based on the GSM technology with data rates around 250Mbps. LTE will support QoS provisioning for real-time applications like multimedia streaming [31].

UMB is developed based on the CDMA technology and provides data rates up to 288Mbps. UMB incorporates control mechanisms which optimize data transmission in order to meet the QoS requirements of various user applications [32]. UBM also supports inter-technology handover with CDMA2000 standards [32].

IEEE 802.16m (WiMAX II) is developed based on the WiMAX standard with adaptation for cellular networks. 802.11m aims at supporting higher data rates and QoS support for various multimedia services. The data rate is expected to reach 100 Mbps for mobile users and 1Gbps for static users.

2.2.4 Wireless Wide Area Networks

Wireless Wide Area Networks offer the largest coverage area. WWANs can be used as separate networks or as interconnection backbones for WMANs. WWANs are usually satellite networks, but terrestrial versions are also under development.

The terrestrial WWAN is standardized by the IEEE 802.20 [33]. This standard targets high mobility users with speeds of up to 250 km/h. QoS preservation methods as well as handover management schemes will be supported by this technology.

Satellite WWANs have the advantages of global coverage, high mobility support and broadcast capabilities [23]. Initially satellite networks had only broadcast capabilities, but within the Next Generation Satellite System (NGSS) unicast and multicast is also provided.

The Digital Video Broadcasting (DVB) standard family started first by supporting digital video and data broadcasting through the satellite networks. DVB-S (satellite) enables down-link data transfer with rates of up to 45 Mbps only. The newer DVB-S2 increases the downlink rate to 60 Mbps. For uplink DVB-RCS (return channel satellite) standard was developed supporting rates of up to 2 Mbps.

Apart from the satellite versions (DVB-S) DVB has also standardized a terrestrial wireless data service through the DVB-T, and more recently DVB-T2.

DVB-T offers much flexibility in terms of data rates. Depending on the particular configuration of the various parameters specific to the wireless transmission it offers a wide range of bitrates starting from 3.7 Mbps up to 31 Mbps [34].

Although DVB-T broadcasts multimedia content to static and mobile users, including vehicular receivers, it is not optimized for highly mobile handheld devices.

Consequently, DVB team has developed DVB-H (handheld) [35] for multimedia content delivery to mobile devices. DVB-H is developed based on the DVB-T (terrestrial), whose infrastructure it uses. Similar with DVB-T, DVB-H offers one way (downlink) point-to-multipoint data communication over wireless links with indoor and outdoor coverage. Considering the limited radio capabilities of a mobile handheld device as well as the higher error rates due to device mobility, DVB-H incorporates powerful error correction mechanisms. Time-multiplexing technologies are used to improve power consumption to cope with the energy constraints of battery powered handheld devices. Seamless handover between base stations is also supported and loss is highly reduced due to the time-slicing techniques used for power efficiency even with only one radio interface [36].

DVB-H supports mainly downlink communication, interactivity being achieved through separate backward point-to-point channels using other wireless data communication technologies like GPRS or UMTS. Supporting mainly broadcast services DVB-H scales well offering downlink data rates between 3.3Mbps and 31.6 Mbps. DVB-H specifies only the protocol layers below the network layer.

DVB-H provides an Internet Protocol (IP) interface for higher transport layers which is defined by the IP-based Data Broadcast (IP Datacast) specification. IP Datacast also offers the option of accessing an external cellular network for the backward channels and to create the so-called hybrid networks [37].

2.2.5 Conclusion

Table 2.3 presents a comparison between the existing and under development wireless access technologies. The comparison considers the aggregated bit rate, the support for QoS provisioning and mobility support.

The various wireless communication technologies described above demonstrates the heterogeneity of the future wireless networking environment which can be easily anticipated. Each technology has its advantages and disadvantages and the supported bit rates vary from one technology to another as well as their support for application specific QoS levels. In this context maintaining the connectivity with the right network becomes a very important task in terms of high quality level support.

	Category	Aggregate Bit rate	QoS Support	Mobility Support
GSM	Cellular	9.6 Kbps	No	Yes
GPRS	Cellular	114 Kbps	No	Yes
EDGE	Cellular	400 Kbps	No	Yes
UMTS	Cellular	2 Mbps	No	Yes
CDMA 2000 1xEV-DO	Cellular	2.4 Mbps	No	Yes
LTE	Cellular	250 Mbps	Yes	Yes
UMB	Cellular	288 Mbps	Yes	Yes
WiMAX II	Cellular	100 Mbps	Yes	Yes
IEEE 802.11	WLAN	2 Mbps	No	No
IEEE 802.11b	WLAN	11 Mbps	No	No
IEEE 802.11a/g	WLAN	54 Mbps	No	No
IEEE 802.11e	WLAN	-	Yes	No
IEEE 802.11n	WLAN	600 Mbps	Yes	No
IEEE 802.11VHT	WLAN	1 Gbps	Yes	No
IEEE 802.16 (WiMAX)	WMAN	75 Mbps(135 Mbps)	Yes	No
IEEE 802.16e(WiMAX)	WMAN	30 Mbps	Yes	Yes
IEEE 802.20	WWAN	16 Mbps	Yes	Yes
DVB-S	WWAN	45 Mbps	No	No
DVB-S2	WWAN	60 Mbps	No	No
DVB-RCS	WWAN	2 Mbps	No	No
DVB-T	WWAN	- Mbps	No	No
DVB-H	WWAN	31 Mbps	No	Yes

Table 2.3: Wireless access technologies

2.3 Video Compression

Video encoding consists of the process of preparing the video content for the output, which can be storage on a digital data storage medium or transmission over a network paths or real-time playback.

The main task in video encoding process is compression and its goal is to reduce the amount of data to be stored or transmitted. In the following sections various algorithms and standards for video compression will be discussed.

2.3.1 Basic Principles of Video Compression

Raw digitized video produces very high amounts of data which is unacceptable for any data storage medium or data transport network. For example one second of a movie in standard television format would produce 32 Mbytes of uncompressed video data [1]. Consequently 120 minutes of video in standard TV format would require 230 Gbytes of storage space.

Consequently during video encoding, compression algorithms are employed to reduce the amount of data required to store and transport the video data. Video compression relies on a good understanding of the human psycho-visual perception system which allows for the exploitation of redundancies in the video signals.

Based on human visual system's characteristics one of the first steps in video compression is to use the YUV format, where Y denotes and the luminance and U and V are the chrominance components, and to down-sample the chrominance. This is based on the fact that the human visual system is more sensitive to luminance (levels of gray) than chrominance (color differences).

There are three types of redundancies which can be identified in a video signal: spatial, temporal and statistical.

Spatial redundancy refers to the fact that neighboring pixels are more likely to have similar colors. Most of the compression algorithms use a two-dimensional mathematical transformation to differentiate between lower and higher spatial frequencies. As the lower spatial frequencies are more important than the others, they are coded with higher accuracy. The transformation process is usually loss-less and does not achieve any reduction in bitrate. Compression is achieved by *quantizing* the coefficients obtained after the transformation process. Quantization is a lossy operation, consequently the decompressed image will be different to a certain extent from the original image. The most popular transformation is the *discrete cosine transform (DCT)* used by the MPEG compression algorithms and is usually applied on blocks of 8 X 8 pixels. However the size of the blocks can be varied to trade efficiency for complexity and spatial variation.

Statistical redundancy is exploited by assigning shorter codewords to more frequently appearing symbols and longer codewords to less often used ones. Once the quantization is performed each block in the transform domain will contain a large number of zero coefficients. These are mostly corresponding to the higher frequencies. Special coding techniques have been developed to combine a number of zero coefficients with the level of the next non-zero coefficients in a single codeword. Various combinations of zero and non-zero coefficients are assigned variable-length codes (VLC) according to their frequency of apparition. The block in transform domain is scanned from low to high frequencies in such a way that the last non-zero coefficient is reached as soon as possible.

Temporal redundancy relies on the fact that usually there is little difference between consecutive frames in a video sequence. Considering this aspect, by calculating the difference between the frame to be encoded and the previous frame and encode only this prediction error, the bitrate is reduced as the frame difference requires fewer coefficients to encode. However if the motion level is too high and the inter-frame difference is significantly greater, the compression efficiency using prediction error is rapidly lost. To overcome this, a technique called motion compensation is used. Motion compensation compares the currently processed block of pixels with a reference block taken from a different location in the previous frame. The reference frame is searched in a larger area in the previous frame in order to estimate the motion within the video sequence. The position of the chosen reference block is then sent to the decoder as a motion vector.

The motion prediction can also be performed by comparing the current frame with future frames. Consequently there is forward, backward and bi-directional motion prediction.

2.3.2 Video Compression Standards

Various algorithms for video compression have been developed and standardized. Some are in the public domain and are developed by organizations such as International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), and International Telecommunication Union (ITU) and others are proprietary and are developed by industry.

2.3.2.1 JPEG Image Compression Standards

The Joint Photographic Experts Group (JPEG), a joint effort of both ITU and ISO standardization bodies, was established to develop an international standard for compression of color and gray-scale still images. This family of standards is referred to as JPEG.

JPEG [38] is the most popular image compression standard. It offers high flexibility regarding image quality - compression rate trade-off. Still pictures can be compressed at a higher ratio but with a consequent lower image quality or a higher image quality can be achieved by using lower compression ratio. Extremely high compression ratios determine artifacts (blockiness) to appear in the compressed image.

JPEG standard uses the DCT transform and a quantization technique to eliminate redundant information.

Motion-JPEG (MJPEG) is not an official standard, however it is used as the movie mode in some digital cameras [39]. MJPEG does not use the advanced compression features of the video compression algorithms (i.e. temporal redundancy) consequently does not offer a high compression efficiency. However it is more robust due to the fact that there is no dependency between frames in terms of compression and consequently if a frame is lost during transmission the other frames are not affected in any way.

JPEG2000 [40, 41] replaces the DCT transform for the Wavelet transform and increases the compression ratio as compared to JPEG plus reduces the ”blockiness” artifacts by replacing them with slight ”fuzzy” picture. These improvements came with the cost of increased complexity.

Motion-JPEG2000 (MJPEG2000) [42] is similar to MJPEG and presents the same advantages and disadvantages; however it offers slightly better compression ratio.

2.3.2.2 MPEG video compression standards

The Moving Picture Experts Group (MPEG) a working group of ISO/IEC has developed a set of standards for video compression and encapsulation which became one of the most popular solution for multimedia content transmission and storage.

MPEG-1 [43] was the first standard developed by the group and targeted coding of a combined audio-video signal at an average bitrate of 1 Mbps in standard input format (SIF) meaning 352 x 288 pixels at 25 frames/s or 352 x 240 pixels at 29.97 frames/s. Although MPEG-1 was first designed for compressing movies for storage on video CD at SIF resolution, it was also used for interlaced standard definition SDTV for bitrates of 8Mbps to 10 Mbps. Apart from storage MPEG-1 was also used for satellite point-to-point links, this extension being referred to as MPEG-1.5. MPEG-1 was fully replaced by MPEG-2.

MPEG-2 [44] was targeting compression of standard definition (SD) and high definition (HD) interlaced video signals at very high bitrates (up to 20 Mbps) and high picture quality. Although very similar to MPEG-1, the most important improvement brought by MPEG-2 was the compression of interlaced video. MPEG-2 is used in various application scenarios from satellite point-to-point links to direct-to-home (DTH) and of course storage on file servers and DVDs. MPEG-2 is gradually being replaced by MPEG-4 advanced video coding (AVC).

MPEG-3 was planned to address HDTV compression, however the same results could be achieved with minimum modifications of the MPEG-2 standards. Consequently MPEG-3 was discontinued.

MPEG-4 consists of two distinct compression algorithms: MPEG-4 Part 2 (Visual) [45] and MPEG-4 Part 10 (AVC) [46]. Although is built on similar principles, MPEG-4 offers much higher flexibility than the previous MPEG standards. It addresses a wider range of bitrates form those used by low bandwidth mobile multimedia applications to those of extremely high quality and high bandwidth applications. The frame rates supported are also flexible unlike MPEG-2 which allows only 25 frames/s for PAL and 30 frames/s for NTSC. MPEG-4 Part 10 (AVC) introduces more advanced compression algorithms, proposing that a new 4 x 4 transform to replace the 8 x 8 DCT and new entropy-coding algorithm based on arithmetic coding to be used.

2.3.2.3 H.26x video compression standards

The International Telecommunication Union (ITU) through the Telecommunication Standardization Sector (ITU-T) have developed several standards for video and audio compression.

These standards have been developed in parallel with the MPEG working group and sometimes the two organizations have jointly coordinated their activities leading to H.26x standards to be similar and some times identical with the MPEG standards.

ITU-T R. H.261 [47] was developed for teleconferencing and videophone applications targeting the ISDN lines as the transport network infrastructure. The output bitrates range from 40 Kbps to 2 Mbps in multiplies of 64 Kbps. The compression technique employed is similar with the one used by MPEG standards however it offers lower compression ratio and flexibility. The processing delays are lower than the ones involved in MPEG video compression, however the lack of more advanced techniques for bandwidth efficiency is somehow contradictory with their target low bandwidth applications.

ITU-T R. H.262 [48] is identical with the MPEG-2 standard.

ITU-T R. H.263 [49] is similar with H.261 but provides better performance and flexibility. It was initially developed for low bitrate communication (below 64 Kbps), however this target has been relaxed. The novelty brought by H.263 over the H.261 is five standard source formats instead of two, half pixel precision for motion compensation unlike the full pixel provided by H.261, 3-D variable-length coding and median vector prediction. In terms of flexibility H.263 offers a wide variety of optional modes which can be used to improve the performance or to extend the application range. Improved version of H.263 were developed as H.263+ and H.263++.

ITU-T R. H.264 [50] is identical with MPEG-4 (AVC) and has become a key technology for multimedia applications. The goal of H.264 is to provide good video quality while substantially reducing the bit rates and latency. The advanced coding tools adopted by this standard include spatial-domain intra prediction, variable block-size motion estimation, multiple reference frames, and rate-distortion (R-D) optimization. A wide range of applications is targeted including broadcast, satellite, cable, DVD entertainment video, telecom services and streaming and professional film production.

2.3.2.4 Proprietary Compression Solutions

Various video coding techniques have been developed by industry and are mostly proprietary solutions.

VC-1 SMPTE 421M [51] has been standardized by the Society of Motion Picture and Television Engineers (SMPTE) and was adopted as a video codec specification in the next-generation optical media formats, such as HD-DVD and Blu-ray. It is developed by Microsoft and was originally known as the Microsoft Windows Media 9. Due to the standardization of the decoding algorithm most of the compression tools are now in the public domain. VC-1 uses a block-based motion compensation and spatial transform and introduces some techniques which bring good coding efficiency, better frame quality and lower computational complexity as compared to H.264/AVC.

Audio Video Coding Standard (AVS) [52] is developed by the Audio Video coding Standard Workgroup of China. AVS has two separate parts. AVS Part 2 is designed for HDTV, high definition and high quality video services, broadcasting and high-density storage media. AVS Part7 (AVS-P7) aims at low complexity, low picture resolution applications for the mobile environment. AVS-P7 is the most recent and targets performance close to that of H.264/AVC but with lower complexity. AVS-P2 was adopted as the Chinese national standard for digital TV Broadcasting and HD-DVD. AVS is highly efficient while being simpler and easier to implement than MPEG-4 (AVC).

Among the most popular other proprietary solutions are Apple QuickTime developed by Apple and Real Media developed by Progressive Networks.

2.3.3 Conclusion

Various video compression techniques have been developed and standardized. Each technique has its advantages and disadvantages and presents different performance in different application scenarios.

The diversification of application scenarios which involve video content challenges the development of future video coding schemes. Depending on the application scenario the video compression technique has to provide certain compression ratio with reasonable complexity and processing requirements as well as the required flexibility.

In order to achieve this goals various specific techniques are employed by each compression algorithm in each stage of the coding process. The diversity of these techniques are outline in Table 2.4.

Coding Tool	MPEG-1	MPEG-2	MPEG-4 Visual	MPEG-4 AVC	VC-1	AVS
Transform	8x8 DCT	8x8 DCT	8x8 DCT	4x4 8x8 Integer	8x8 8x4 4x8 4x4 Integer DCT	8x8 Integer DCT
Intra prediction	DC	DC	AC Freq. domain	4x4 8x8 16x16 Spatial	AC Freq. domain	8x8 Spatial
Interlace	No	Yes	Yes	Yes	Yes	Yes
Motion comp. block size	16x16	16x16 16x8	16x16 16x8 8x16 8x8	16x16 16x8 8x16 8x8 8x4 4x8 4x4	16x16 16x8 8x16 8x8	16x16 16x8 8x16 8x8
Motion comp. beyond picture	No	No	Yes	Yes	Yes	Yes
Motion vector precision	1/2 pel	1/2 pel	1/4 pel	1/4 pel	1/4 pel	1/4 pel
Multi. Ref. P frames	No	No	No	Yes	No	Yes
Weighted prediction	No	No	No	Yes	Fading comp.	No
Direct mode pred. on B frames	No	No	Yes	Spatial and Temp.	Yes	Temp. and Sym.
Pred. from B frames	No	No	No	Yes	Yes	No
De-blocking filter beyond picture	No	No	Post proc. filter	In-loop filter	In-loop filter overlap smooth.	In-loop filter
Entropy coding	Fixed VLC	Fixed VLC	Fixed VLC	CABAC	Adapt. VLC	Adapt. VLC

Table 2.4: Comparison of coding tools [1]

2.4 Video Content Delivery: Techniques and Protocols

Multimedia content delivery is usually associated with broadcasting in the traditional TV broadcasting services. Currently, with the development and proliferation of other multimedia based services like video-on-demand or video conferencing, unicast and multicast content delivery becomes widely associated with multimedia applications.

Broadcasting has the main advantage of being cost effective in terms of bandwidth. Delivering the same content to multiple users in the same time reduces resource utilization. As the users are more likely to prefer video-on-demand-like services [53], this may decrease the popularity of broadcasting.

Unicast has the main advantage of supporting video-on-demand services as well as broadcast services. Another benefit of unicast is that network resources are used only when there is a user requiring a certain service [53].

In the context of multimedia delivery to mobile devices unicast presents the benefit of supporting content adaptation for each user separately in order to meet their device capabilities as well as networking resources. **Multicast** is beneficial especially for group content delivery in applications like video conferencing. However the management of multicast groups is difficult and complex.

Various wireless solutions have been proposed to address multimedia content delivery to mobile users. Three typical categories can be identified: DVB-based solutions which enhance the Digital Video Broadcasting standard for mobile devices (handheld), solution enhancing the third generation cellular networks including UMTS with Multimedia Broadcast Multicast Service (MBMS) support and content delivery solutions exploiting the widespread popularity of WLAN (802.11 - WiFi).

2.4.1 Multimedia Streaming in Cellular Networks

In cellular networks multimedia delivery was introduced starting from the 2.5G technologies and continuing with current 3G technologies. The packet-switched streaming (PSS) standard developed by the Third Generation Partnership Project (3GPP) provides the means of content transportation for streaming and downloading applications. PSS uses various protocols for content delivery and information exchange. Content is delivered using Real Time Transport Protocol (RTP) over User Datagram Protocol (UDP). Other types of media including text and graphics are delivered over HTTP.

The Real Time Streaming Protocol (RTSP) is used for control information exchange. PSS supports user quality of experience monitoring which permits content adaptation for improving user satisfaction.

Apart from the transport mechanisms, PSS also specifies a set of media codecs including Adaptive Multirate (AMR), H.263, H.264, MPEG-4 Advanced Video Codec (AVC) and MPEG-4.

For further improvement of user perceived quality PSS also includes an Adaptive Streaming feature [54] which is useful for adapting the content to network condition variations as well as variations in network characteristics due to handovers between systems [53] like GPRS to WCDMA and vice-versa.

The **IP Multimedia Subsystem (IMS)** [55] was developed within 3GPP as a service platform to provide multimedia services over 3G networks. IMS uses the Session Initiation Protocol (SIP) for signaling and session control, RTP for media transport and IPv6 at the network layer. The IMS platform is not directly involved in media transport (only in session control) [55], but QoS is maintained by collaboration between the IMS platform and the transport network. The Policy Decision Function (PDF) is the IMS sub-module which is responsible for QoS negotiation according to the application requirements.

3GPP also proposed the **Multimedia Broadcast/Multicast Service (MBMS)** for UMTS [56]. MBMS delivers multimedia content to a group of users in a point-to-multipoint manner using UMTS MBMS transmission bearer. MBMS is composed of two modules, the MBMS bearer service which deals with transmission procedures below the IP layer and the MBMS user service which manages streaming and downloading methods and procedures. The streaming methods used by MBMS are similar to PSS in terms of transfer protocols (e.g. RTP) and codecs.

Broadcast and Multicast Services (BCMCS) [57] protocol is similar with MBMS but was developed by 3GPP2 for the CDMA2000 protocol family for 3G networks. Similar to MBMS, BCMCS provides point-to-multipoint content delivery and guarantees QoS for two way multimedia applications. Content adaptation may be performed using SVC which was also introduced in MPEG-4 standard as **FGS (MPEG-4 FGS)** [58].

2.4.2 DVB-based Multimedia Delivery

The Digital Video Broadcasting (DVB) standards offer point-to-multipoint data services with high data rates for multimedia (especially TV) content delivery to end users. Apart from the satellite versions (DVB-S), DVB also standardized a terrestrial wireless data service through DVB-T. Although DVB-T broadcasts multimedia content to static and mobile users, including vehicular receivers, it is not optimized for highly mobile handheld devices.

Consequently, Digital Video Broadcasting (DVB) Project has developed DVB-H (handheld) [35] for multimedia content delivery to mobile devices.

DVB-H specifies only the protocol layers below the network layer, consequently an Internet Protocol (IP) interface for higher transport layers which are defined by the IP-based Data Broadcast (IP Datacast) specification was introduced.

IP Datacast specified the protocols for higher layers in concordance with the Internet protocol stack. For transport layer UDP was chosen with RTP for real-time media broadcasting and File Delivery over Unidirectional Transport (FLUTE) for non real-time data transfer like download-based media delivery.

For media encoding DVB-H aims for high compatibility between network components and terminals, therefore MPEG-4/H.264 was chosen for video encoding as well as the Microsoft Windows Media 9 based VC-1 codec. For audio MPEG 4 AAC+ is recommended.

IP Datacast was developed towards a hybrid interoperability of several types of networks in order to benefit of their advantages and balance their disadvantages. DVB-H has the advantage of being highly scalable with the number of users unlike cellular networks or WLANs which are basically point-to-point and suffer from severe QoS drops when congested.

On the other hand DVB-H interactivity is quite limited. Considering these aspects using a hybrid solution where the content management layer decides which network to use for delivery of a certain service depending on its popularity (the number of users requesting the same service at the same time) or its interactivity requirements may improve the overall quality of the service as perceived by the user.

2.4.3 Multimedia Delivery Over WLAN

Wireless LAN is probably the most successful and cost-effective option for multimedia delivery. With encouraging link layer data rates, IEEE 802.11 based WLANs popularity is constantly increasing.

Despite all the positive aspects WLANs suffer from the same unpredictability of the wireless links as well as from severe QoS drops when the wireless medium gets congested with the increase in the number of simultaneous hosts engaged in data sessions.

IEEE 802.11 standards describe only the physical and MAC layer. There are several enhancements proposed for supporting multimedia applications including the IEEE 802.11e and IEEE 802.11n QoS support features in the MAC layer. There are also various prioritisation schemes [17] which allow traffic differentiation depending on the type of traffic (priority class).

WLANs use the Internet protocol stack for higher layers, consequently all the multimedia streaming solutions designed for the Internet may be used in scenarios including WLANs. Although WLANs are compatible with the IP based network paradigm their particularities especially regarding the error prone and highly dynamic wireless links has to be considered by the higher layers in order to provide a high quality multimedia streaming service.

2.4.4 Conclusion

As outlined in the previous sections, delivering multimedia content to mobile users is an active research and development area.

Various systems, algorithms and protocols have been developed and deployed over the existing transport networks including the cellular networks, broadcasting networks and wireless LANs.

Considering the diversity of the available technologies and the increasing capabilities of the mobile devices in terms of communication, managing the multimedia content delivery process becomes an extremely challenging task.

2.5 Audio-Video Quality Assessment

Video and audio quality assessment is extremely important in the development and operation of multimedia applications. The quality of the video and audio content delivered to the end-user has a great impact on the user Quality of Experience (QoE) and consequently on the success of the multimedia-based service or application.

Video quality assessment tools and methods are used in the performance evaluation process of multimedia applications. Video quality assessment metrics are also employed in monitoring the content distribution networks for maintenance purposes. These metrics are also used by various adaptation schemes which are involved in adjusting multimedia delivery process to the contextual particularities of the application scenario. They are particularly useful for assessment of the effects variable network conditions have on user perceived quality.

If analogue video quality can be easily measured and monitored, for example by using four vertical interval test signals (VITS), assessing the video quality of digitally compressed content poses certain challenges.

Digital video quality assessment requires an extensive knowledge of the human psycho-visual system and usually makes use of complex algorithms and models.

From the point of view of user involvement in the assessment process, video quality assessment methods can be classified in two categories: **subjective methods** and **objective metrics** [59].

The main goal of video quality assessment tools is to detect the various impairments, denoted artifacts, which may appear in the video content presented to the user.

Audio content quality assessment is equally important, various algorithms being developed and some standardized by ITU.

2.5.1 Video Impairments

Visual impairments may appear in the video content delivered to the user from various sources including video sensor noise, encoding artifacts and transmission errors.

Random noise consists of a random dot pattern which appears in the captured image due to noisy video sensor or low-light conditions. This kind of video artifact affects the entire image area and has a negative effect on the user perceived quality. The same negative effect random noise can leave on video processing tools which may consider it as fine grained motion within the video sequence like rain drops or tree leafs swinging in the wind.

Blocking effects or tiling consists of false blocks appearing in the decoded image and is specific to block based encoding techniques like MPEG. This artifact is determined by the independent quantization of each separate block in each frame.

Blurring consist in reduced sharpens around the edges and of spatial details. It is determined by the suppression of high-frequency coefficient by using coarser quantization factors.

Mosquito effect is a form of edge busyness determined by time-varying sharpness of the edge of objects and is determined by different coding of same area of the image in consecutive frames.

Jagged motion is determined by poor motion estimation while **Jerky motion** is caused by lost motion when video is transmitted at lower frame rate. Jerkiness is perceived as a series of discontinuous images.

Color bleeding appear as a smearing of colors between areas of strongly different chrominance and is determined by suppressing the high-frequencies in the chroma space.

Misplaced macro blocks with respect to their own color and the color of the surrounding area is called **Chrominance mismatch**. This is determined by the use of only luminance information in the motion estimation process.

The transmission process negatively affects the video content quality through loss of data packets. Lost data prevents the decoder to fully decompress the entire image area. Consequently some blocks are not properly decoded leading to visual impairments in the frame area. This is basically **spatial propagation of transmission errors**.

Motion prediction determines an interdependency between blocks in subsequent frames. Under these circumstances, if error blocks appear in a frame, the decoding process of all other blocks in other subsequent frames which depend on lost blocks will be negatively affected. This represents **temporal propagation of transmission errors**.

Video artifacts are specific to the compression algorithm used. For example, block related artifacts (e.g. blocking) appear when block-based compression schemes are employed and do not affect video content compressed using wavelet-based algorithms. In the same time transmission errors affect differently video content compressed using distinct compression algorithms.

For more details regarding the video impairments and the various types of artifacts specific to digital video data please refer to [60].

2.5.2 Subjective Assessment of Video Quality

Subjective video quality assessment is performed using human observers involved in the testing process [61].

Subjective testing represents the reference in video quality assessment and is considered to offer the most accurate rating of the quality level as perceived by the user.

In order to make subjective testing as reliable as possible and repeatable, methodologies and recommendations such as those from ITU-R BT.500 [62], ITU-T R. P.910 (one way video test methods) [63], and ITU-T R. P.911 (quality assessment methods for multimedia applications) [64] have been defined.

Subjective testing can be used only in the development and evaluation stage of a multimedia application, they are not suitable for in-service quality monitoring. High number of observers have to be involved in order for the test to be valid. This brings the main shortcoming of this assessment method which is the amount of time required to complete it. Other disadvantages would be the complexity of the setup and the cost.

These combined with the wide range of available methods and test parameters to be considered make subjective perceptual test quite challenging.

The main advantage of subjective perceptual tests is that they offer good insight into the way human subjects perceive the quality of the video sequences and consequently the quality of the offered service, representing a benchmark in video quality assessment.

During subjective tests the viewers are shown sets of short video sequences (e.g. 10 seconds) which are then asked to rate on a predefined scale. The environment is controlled and so are the test equipments. The average of all valid scores represents the mean opinion score (MOS) of human observers related to a particular test. More information on the testing methods can be found in the above mentioned recommendations.

2.5.3 Objective Video Quality Assessment Metrics

Objective video quality assessment methods involve algorithms designed to achieve a good correlation with the human visual perceived quality and approximate as good as possible the subjective mean opinion score (MOS).

Objective methods can be classified from the point of view of usability in conjunction with adaptive streaming solutions as **out-of service** methods (i.e. the original sequence is available and no time constraints are imposed) and **in-service** methods (i.e. performed during streaming without original sequence and with strict time constraints) [65].

From the point of view of the existence of the original multimedia stream during the quality assessment [66] the objective methods can be classified into **full reference** methods (i.e. use comparisons with the reference stream), **reduced reference** solutions (i.e. employ feature extraction) and **no reference** methods (i.e. no original stream is required for quality assessment).

From the point of view of analyzing the decoded video two categories can be distinguished. **Data metrics**, which estimates the fidelity of the signal without considering the content and **picture metrics** which consider the visual information contained in the decoded picture [67].

From the transmission's point of view there are metrics which look at the packets and the encoded bitstream without decoding the video [67] and those which require decoding.

Among the most popular data metrics or mathematical metrics are the Peak-Signal-to-Noise-Ratio (PSNR) [68] and Weighted-Signal-to-Noise-Ratio (WSNR).

Their popularity is mainly due to the simplicity of the formulas which determine ease of implementation and fast computation. However research have shown that their correlation with the human visual system is poor [69].

Picture Appraisal Metric (PAR) [70] was introduced as a non-reference metric by Snell & Wilcox. Being a non-reference metric it is suited for in-service used. However it is based on PSNR and consequently suffers from the same poor correlation with the human visual system perception particularities.

The shortcomings of the data metrics have determined the development of more complex metrics which model more accurately human perception.

Among the most well known metrics which model various components of the human visual system are Visual Differences Predictor (VDP) [71], Sarnoff JND [72], Moving Picture Quality Metric (MPQM) [73] and Perceptual Distortion Metric (PDM) [74].

Other types of metrics rely on extracting and analyzing certain features or artifacts in the video image [67]. Among the most popular such metrics are Structural Similarity index (SSIM) [75] and Video Quality Metric (VQM) [76]. Although they do not fundamentally model the vision system, these metrics incorporate aspects related to the human psycho-visual perception.

With the increase in popularity of multimedia streaming applications deployed over IP networks several packet- and bitstream-based metrics have been developed. These metrics measure the impact of transmission losses on video quality and require no or little decoding of the actual video content.

Examples of such metrics are the ones developed by Verscheure et al. [69] which investigate the impact of packet loss and bitrate on video quality and the one developed by Kanumuri et al. [77] which uses bistream parameters to predict video quality in MPEG-2 and H.264 encoded video content. V-Factor developed by Winkler et al. [67] is a metric from the same category.

Few studies that compare the performance of the objective metrics exists and a clear hierarchy in terms of video quality prediction accuracy ca not be determined. Consequently subjective testing still remains a benchmark for video quality assessment.

2.5.4 Audio and speech quality assessment

Subjective human listening tests have been used for evaluating audio quality of various devices and applications based on the ITU recommendations [78]. However, as is the case of the subjective video quality assessment audio and speech quality tests are expensive and time consuming. Consequently objective metrics and algorithms have been developed to replace the subjective methods such as the E-model [79].

The ITU standard for objective audio quality assessment used in the development of multimedia devices, codecs and streaming applications is PEAQ [80]. Similar with the case when objective video quality metrics are used for modeling the human visual perception system, PEAQ models the psychoacoustic characteristics of the human auditory system.

The PEAQ model has been further enhanced with new functionality and increased performance. Hubers novel assessment model[81] and Barbedo's novel cognitive model [82] are important enhancements of PEAQ.

An equivalent standard algorithm for speech quality assessment is Perceptual Evaluation of Speech Quality (PESQ) and its wideband extension WB-PESQ [83]. The main disadvantages of this standard from the point of view of multimedia applications are the facts that it only supports limited bandwidth signals such as narrow band speech (4 kHz bandwidth) and it does not work with high bandwidth applications which are common in the most modern audio systems.

2.6 User Devices

Mobile devices have gone through a rapid evolutionary process in the last decade. If the main goal in the 1990s and early 2000 was to reduce the size of the mobile terminals, since 2002 an ever increasing set of features started to appear in the mobile devices.

Voice telephony and Standard Message Service (SMS) were the first applications available on the mobile phones. Since then new functionalities have been added including cameras, Multimedia Message Service (MMS), e-mail and mobile Web browsing.

These applications are supported by additional hardware which includes high-resolution color displays, high processing power, increased memory and storage space.

Since all these technological advancements have taken place, the traditional mobile phones have evolved into several other categories of mobile devices.

Smartphones are a combination of a mobile phone and a Personal Digital Assistant (PDA). A smartphone has all the basic features of a standard mobile phone but also include more advanced features. Among these we can mention high-resolution camera, bigger and higher resolution color screen and various networking interfaces including Bluetooth, WiFi and cellular modem.

Personal Digital Assistants (PDA) have evolved from devices with little connectivity focused on personal information management to highly connected mobile terminals. Nowadays the PDAs are equipped with all the networking interfaces which can be found on a smartphone including the cellular modem. This aspect narrows the gap between smartphones and PDA to a minimum.

Internet tablets are small computers with applications like Web browsers and even operating systems ported from the desktop computing world. Internet tablets are equipped with a high resolution color touch screen and advanced wireless connectivity including WiFi, 3G and WiMAX.

Ultra mobile PC's and Netbooks are bigger then Internet tablets but smaller than a notebook PC. These devices run desktop operating systems with minimal adaptation and provide most of the functionally a standard notebook would provide, however the processing power is traded for low energy consumption.

Laptop computers are a portable version of personal computers providing all the functionality a desktop PC would provide. They are mostly equipped with full size keyboards and screen sizes between 13" to 17" however bigger versions can be found on the market. Laptop computers are basically equipped with wired LAN and WiFi interfaces however Bluetooth and 3G modems are rapidly becoming a common feature.

2.6.1 Conclusions

Mobile computational devices have greatly evolved in the past decade with a wide range of types and form factors being available on the market.

Their functionalities have also evolved especially from the point of view of processing power, wireless connectivity and graphical screen resources.

These aspects combined with the development in wireless networks and their wide availability to users have opened the door for the future mobile Internet and among other applications to multimedia content delivery.

The diversity of available mobile devices and the multitude of wireless networks available combined with the various user preferences and expectations makes high quality service delivery over the mobile Internet an extremely challenging task.

2.7 Chapter Summary

This chapter presented the state of the art in various technologies related to the subject of the thesis. The current developments in the area of wireless communications have been introduces. Various types of networks such as WPAN, WLAN, WMAN and WWAN, have been discussed the currently available or under development technologies to support them.

Technologies and algorithms supporting multimedia content delivery to mobile users have been discussed as well. Encoding standards and compression algorithms have been discussed together with various systems and protocols developed for supporting delivery of encoded multimedia content to users.

And last but not least, the development of wireless mobile devices was approached. Mobile device evolution and its technological characteristics are very important for the success of the user application running on it.

Chapter 3

Related Works

3.1 Introduction

Delivery of Internet-based services to mobile users is sustained by the latest developments in wireless communications and mobile device technologies. However, technological diversity raises certain challenges in maintaining high level of user satisfaction while ensuring an efficient and cost effective usage of the available resources.

Regarding multimedia streaming, which represents the application context of this thesis, various modules, algorithms and strategies are involved in encoding, encapsulating and transporting multimedia content to its destination which, in this particular case, is a mobile user.

For a better understanding of the various processes involved in high quality multimedia content delivery to mobile users (targeting an increased level of user satisfaction is crucial for the success of any service), a real life scenario will be used. Consider a user receiving multimedia content on a mobile handheld device (e.g. Smartphone). Assume this device is equipped with several wireless interfaces allowing it to connect to various networks including, WIFI, GPRS and WiMax.

In the context of a heterogeneous wireless environment, the mobile user, depending on their location, will be able to access Internet services over one or several wireless networks via the interfaces mentioned above. If the mobile device is on the move, the availability of the wireless networks will become highly dynamic. This is determined by the limited range of the wireless links and variability of the environment characteristics (e.g. presence of obstacles between mobile devices and base stations and/or interferences).

Under these circumstances one of the highly required services is **Mobility Management**. This service has the role of keeping the active data sessions alive and provide the mobile host with the best possible connectivity. Mobility management is also responsible for keeping the mobile devices reachable to any devices which intend to initiate communication sessions with them.

The increasing popularity of Internet applications run on mobile devices are a clear sign that network load in terms of mobile host number as well as the number of data sessions initiated by a mobile device will dramatically increase in the future.

This brings forth another important issue in data communications which is congestion. To accommodate the increasing number of mobile users and efficiently exploit networking resources available, **Load Balancing** strategies have to be employed.

Load balancing combined with mobility management becomes a very powerful component of the current data communication systems. It provides an optimal distribution of mobile users and theirs data sessions over the available networks in order to support high QoS levels for the users as well as efficiency and cost effectiveness for the network operators.

Traffic Distribution Over Multiple Connections have been used before for throughput aggregation and load balancing. This approach presents several advantages including higher reliability and robustness as well as more efficient utilization of radio resources. However it also raises several technical challenges including rate allocation and traffic splitting over multiple connections and traffic merging.

The following sections will discus in detail each of the three services briefly introduced above.

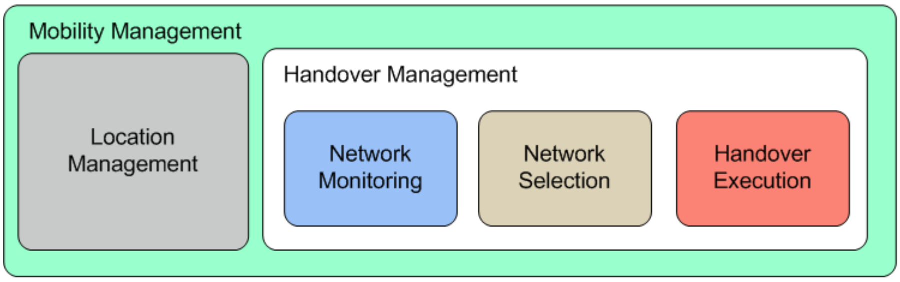

Figure 3.1: Mobility management services and components

3.2 Mobility Management

3.2.1 Services, Requirements and Components

User mobility, heterogeneity of the wireless networking environment and the impact of wireless network dynamics on the QoS level received by the mobile device (and consequently the impact on user perceived quality) make mobility management a crucial component of the future mobile Internet.

Starting with the main goals, which are to accommodate user mobility within an ever changing wireless networking environment and provide mobile host with the desired QoS level according to applications requirements, there are two main tasks that a mobility management module has to perform.

As schematically represented in Figure 3.1 the first task is to maintain the mobile host reachable while roaming trough the heterogeneous wireless environment. The service in charge with this goal is **location management**. Location management keeps track of mobile host's movements and maintains an updated location information database. A different mobile or fixed host wishing to establish a data session with the mobile host will query the location management system for the current location of the mobile host.

After the data session is initiated the mobile host will encounter variations in QoS parameters' values or potential severe disconnections due to wireless network environment dynamics. Wireless network characteristic dynamic variations are determined by environmental conditions, host movements and network congestion levels. In this context maintaining a certain level of QoS requires a constant monitoring of the network conditions and a seamless traffic switching or distribution over the most appropriate currently available network. This task is performed by another mobility management component, the **handover management** service.

In order to be effective these mobility management components have to fulfill various requirements. These requirements may be summarized as follows:

- Security: mitigate additional security risks generated by mobility procedures.
- Preserving compatibility with IP routing: extremely important in the context of the all-IP convergence trend of the future wireless networks.
- Scalability: increased number of nodes performing handover simultaneously should not decrease the performance of the mobility management solution.
- Robustness: mobility management solutions should be able to function correctly with good performance in the highly dynamic wireless environment.
- Facile and cost-effective deployment: as an additional component of the already largely deployed Internet protocol stack, the mobility management service should involve as little changes as possible to the existing infrastructure.
- Power efficiency: as mobility is mainly associated with portable battery-powered devices, power efficiency is mostly desirable.
- Transparency: regardless of its position within the Internet protocol stack, mobility management should be as less disruptive as possible to the higher levels.

It is desirable to have a mobility management solution which does not interfere with the network infrastructure (require additional support from the network infrastructure). This is mostly valid when not all the available networks implement mobility support or the network administrators are not willing to provide mobility support. Consequently mobility management solution design should minimize the interference with the network infrastructure.

Considering all the requirements summarized above, providing a quality-oriented mobility management for the Internet represents a very challenging task. Each additional requirement may impact on the mobility management's capability to sustain a certain level of QoS and to preserve high user perceived quality. For example, security related procedures are most important in the context of wireless mobile networking to avoid intrusion but they involve a certain level of delay which may impact QoS levels during handover. However the most important aspect of mobility management with the greatest impact on QoS and user perceived quality is handover management.

Handover management allows mobile devices to roam freely within a geographical area covered by several wireless access networks (wired network connectivity may also be available at some points) while preserving their ongoing data sessions. The mobile devices will change their point of attachment from one network to another depending on network availability and capability to offer required data communication services.

As presented in Figure 3.1 handover management consists of three different sub-services.

Network monitoring is responsible for gathering data related to current network conditions as well as network availability. Network monitoring detects any context changes and triggers the handover when the required QoS level is no more sustainable or a drop in QoS level is foreseen.

The network monitoring sub-module has to collaborate with another service which may not be part of the handover management system. This is the host/user profile monitoring. This module harvests information about user preferences, applications running on the mobile device and their QoS requirements in order to give **handover execution** information for decision making.

Network selection decides to which network the data session should be transferred to in order to maintain the required QoS level. The network selection algorithms have a great impact on the overall quality and performance of the handover management. Choosing the wrong network will negatively impact the quality as perceived by the user. The network selection sub-service also needs to collaborate with the host/user profile monitoring module in order to retrieve application and user related reference information. The decision module

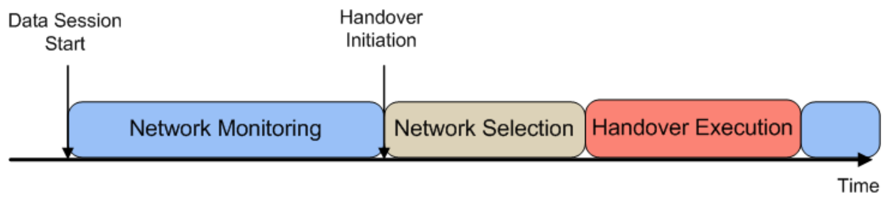

Figure 3.2: Handover management process

also needs information about the available networks and their provided service levels.

Once the new network is selected, the handover is actually performed by the **Handover Execution** sub-service and the data session is transfered to the selected network. Figure 3.2 presents the timeliness of these operations.

Handover may also be used to efficiently distribute traffic load over the available networks. Handover mechanisms have an important impact on the overall quality of the handover management solution and have to be performed seamlessly and non disruptive from the point of view of the upper layers.

3.2.2 Network Monitoring and Context Information

Maximizing user perceived quality in the context of a heterogeneous wireless environment requires an efficient and effective handover decision mechanism. In order for such a mechanism to perform well, a set of context information regarding the mobile devices and users as well as the available networks is required.

This context information can be classified according to [84] in the following categories.

Network-related information includes **radio link quality** (e.g. Received Signal Strength (RSS), Bit Error Rate (BER), signal to noise ratio (SNR)), coverage, bandwidth, delay, data transfer cost, security policies and QoS provisioning. **Terminal-related information** comprises velocity, battery power, location, wireless communication capabilities and for more advanced adaptive schemes, display and user interface capabilities as well as processing power. **User-related information** includes user profile and preferences. Service-related information includes service capabilities and QoS requirements.

This entire set of context related parameters has to be monitored and updated values have to be delivered to the handover decision process. Apart from ensuring updated and relevant data, the real-time constraints required by the handover timing and the dynamics of the context have to be met as well.

Context information gathering involves collaboration between several entities on the network side as well as at the mobile host level. Each of the context parameters identified before are discussed next.

Radio link quality can be usually determined by interrogating the lower network layers, more precisely physical and MAC layers. The information obtained may be used locally for channel quality assessment or it can be reported as feedback to distant decision entities [85]. **Network coverage** estimation and reporting is mainly the operator's task. The network operator may estimate the coverage of each access point or base station under his management and this information may be provided to the handover decision modules [86]. A Network Traffic Monitor [87] may be used to monitor and evaluate bandwidth, delay as well as other QoS related network parameters for each available access point or base station, information which can be queried and used by the handover decision modules.

Data transfer cost and **security policies** are received from the network and are not expected to change very often (at least not during a data session). **QoS provisioning mechanisms** are specific to each network technology and network operator. By using access control mechanisms each network can decide if a new data session may be accepted at the required level of QoS, or not.

Terminal-related parameters are usually harvested locally through specializes services. **Battery power monitoring** process may be used to determine mobile device's current power level and current power management policy. Wireless communication capabilities as well as display, user interface capabilities and processing power may be determined by interrogating the specific services of the operating system or by directly interacting with the corresponding device driver.

Global Positioning System (GPS) receivers are becoming a common feature of mobile devices as their capabilities are frequently extended with navigation services. Consequently **velocity** and **geographical position** may be obtained from the built-in GPS receiver. As the GPS signal is not received indoors, various other localization mechanism have been proposed using triangulation based on strength of the signal received from the access points or base stations or using specialized indoor localization systems [88].

User profile and **preferences** may be determined locally by providing specialized user interfaces where users may set their preferences. Alternatively more complex user profiling may be achieved and stored in specialized repositories [87] from which user-related information may be retrieved.

The application layer should be able to provide application related information including **QoS requirements** and **real-time constraints** in order to allow the decision making modules to maintain the required level of QoS and connectivity. An interaction between the application layer and the mobility management layer is most desirable especially in the presence of adaptive multimedia streaming applications where a content adaptation scheme may work with the handover decision module to effectively match application needs to available networking resources.

The diversity of parameters which may be involved in handover decision making and the difference in location between the point of information harvesting and information consumption require an effective mechanism for context information dissemination. The emerging IEEE 802.21 standard provide such a support for inter technology handover [19]. The draft specification [89] defines the link-layer services and focuses mainly on the pre-handover phase comprising decision and preparation processes.

Three main services are defined by the IEEE 802.21 standard: Media-independent information service (MIIS), Media-independent command service (MICS) and Media-independent event service (MIES). MIIS provides information about the available heterogeneous wireless network environment (available networks, network topology, network proprieties and available services). MICS is in charge with the management and control of the link interfaces. Querying the available networks about varying parameters and resources is also supported by MICS. MIES provides link layer triggers and measurement reports.

IEEE 802.21 is a significant framework which provides necessary means to perform a smooth, seamless and efficient handover with minimal data session disruption and QoS degradation. However there are some implementation/deployment issues. The first is to integrate IEEE 802.21 support in the link layer as well as network component support for the IEEE 802.21. The second issue is more of an administrative matter and is related to collecting and maintaining the information database describing the networks available in a heterogeneous environment. It is difficult to make all network owners and intermediate operators to subscribe and contribute to monitoring such a service without benefit for them.

3.2.3 Network Selection

The future of wireless communications is marked by the development of various wireless technologies, each technology having its own particularities in terms of coverage, provided QoS levels, reliability, monetary cost and power consumption of the network interface, etc. Under these circumstances meeting the requirements of the user applications relies on another important component of the mobility management which is **network selection**.

Network selection strategies target the **Always Best Connected (ABC)** principle which states that the user applications should always get the best connectivity anywhere and at anytime.

The complexity and consequently the impact on the overall system performance of the network selection algorithm depends on the networking scenario and the type of handover performed. In the following sections various types of network selection algorithms will be presented and discussed.

3.2.3.1 Necessity of Advanced Decision Strategies

In a homogeneous networking environment like GSM, network selection is mainly based on radio link quality measured by parameters which include Received Signal Strength (RSS) , Carrier-to-Interferences Ratio (CIR), Signal-to-Interferences Ratio (SIR), and Bit Error Rate (BER). In this environment handover is mainly determined by user mobility and the consequent fading of the current link. Apart from the radio characteristics the network can be chosen also considering load balancing strategies [90].

The criteria mentioned before is sufficient in homogeneous networks with mobile devices switching networks using the same technology, managed by the same operator and running low bandwidth applications like voice services in GSM networks.

However when the homogeneous network environment becomes a heterogeneous one with several types of networks being available for the mobile user to connect to, each network having different characteristics which include monetary cost, bandwidth or services provided, advanced network selection algorithms which consider all these new parameters in the selection process have to be employed.

Moreover if the range of applications running on the mobile device is also becoming heterogeneous, including no-real-time as well as real-time processes, the network selection algorithms have to consider the characteristics and requirements of user applications in order to make the right decision, which is going to impact on the level of user perceived quality.

3.2.3.2 Network Selection Strategies

Handover initiation and network selection relies on various factors in the decision making process. These factors, denoted decision criteria, represent the qualities that are measured and give the necessary information for handover decision making.

These criteria can be grouped in several categories as follows:

- Network related: bandwidth, coverage, latency, link quality (RSS, CIR, SIR, BER), monetary cots, energy consumption, security level.
- Device-related: velocity, battery power, location information, functionality.
- User-related: user profile and preferences.
- Service-related: service capabilities, QoS levels, application requirements.

Based on these criteria decision policies are established and consequently network selection strategies are developed. Policies define how the criteria are interpreted and how these criteria influence when and where the handover will be performed. The network selection strategies may be grouped in several distinct categories [84] based on the decision policies as follows:

- Function-based strategies: use a weighted function comprising various parameters which estimates the benefit of handing over to a certain network.
- User-centric strategies: consider user preferences in terms of QoS and cost in the decision making process.
- Multiple attribute decision strategies: the selection is make from a limited set of options with respect to multiple distinct criteria.
- Fuzzy Logic based strategies: use Fuzzy Logic (FL) in conjunction with multiple criteria to develop advances decision policies.
- Neural Network based strategies: use Neural Network (NN) in conjunction with multiple criteria to develop advances decision policies.
- Context aware strategies: use context information related to terminal and network and context change to decide on when and where the handover has to be made.

In the following sections various strategies proposed in the literature in each category will be discussed.

3.2.3.3 Function-based Strategies

The function-based decision strategies use a weighted cost function which evaluates the benefits of connecting to a certain network. The cost function is evaluated for each available network separately. The parameters used in the cost function may vary from solution to solution. The network which presents the best cost rating according to the predefined scale is selected.

Wang et al. [86] present a selection strategy which uses bandwidth, power consumption and monetary cost for decision. Based on these parameters a cost function is evaluated for each network separately and the network with the lower cost is chosen as the target network. A stability period is used in form of a waiting period before performing handover after a better network was discovered. This stability period has the role of determining if the selected network is really better than the current one and if it keeps its low cost.

Chen et al. [91] propose a similar approach, which further improves the solution presented in [86]. This solution introduces an adaptive stability period based on network resources and application requirements. The stability has the role of avoiding false handover or unnecessary handover as well as the ping-pong effect. The evaluation is based on bandwidth and movement speed as decision factors. The available networks are discovered using the ideal coverage concept combined with node localization.

Zhu and McNair [92] introduce an optimized network selection strategy based on policy-based network architecture. A trade-off between the user satisfaction and network efficiency is achieved. In order to reduce the delay and processing complexity network elimination constraint were introduced. A prioritized multi-network scheme is also employed to improve the throughput of the mobile host with multiple sessions.

Tawil et al. [93] present a vertical handover decision scheme based on a network quality estimation function. The Simple Additive Weighting (SAW) ranking method is used to introduce a quality function which involves various parameters including QoS, service cost, security, power consumption and the list is open allowing for application specific set of parameters to be added. The approach distributes the decision making process between the network and the mobile devices. The network quality rating is evaluated by the network were the access control scheme is also involved but the decision itself is made by the mobile host. This approach reduces the processing requirements at the mobile host level, but it complicates the deployment of the mobility scheme.

Shen and Zeng [94] propose a cost function based on a network selection strategy which approaches the decision making process from a systemic perspective. Traffic load (available bandwidth), received signal strengths (RSS) and monetary cost. The cost function used has weighted parameters (bandwidths, RSS and cost) and by setting different values to these weights the decision making process can be tuned to achieve different goals. By relying on bandwidths only, load balance is achieved while by using RSS only the best quality link is selected. However in a heterogeneous wireless environment comparing signal strengths of links using different technologies is relatively difficult to achieve.

3.2.3.4 User-centric Strategies

User centric network selection approaches consider user priority regarding policy parameters. Users are becoming central in system design in various areas of data communications. For instance 4G networks will support personalized services where the user will always receive the best connectivity. Is a paradigm shift from ”Always Connected” which is specific to 2G and 3G networks to ”Always Best Connected”, with the user in the center of decision making process.

Calvagna and Modica [95] developed a model for handover decision making which considers monetary cost and QoS. The decision is made so the user will receive an optimal combination of cost and QoS according to their preferences. Two policies are defined, one which offers better connectivity for the users who are willing to pay more for services and a second one which offers a low monetary cost but with no connectivity and QoS guarantees. This solution as presented by the authors is limited to vertical handover between GPRS and WiFi as wireless networking technologies.

Ormond et al. [96] introduced a network selection strategy for non-real-time data services (i.e. FTP file transfer) based on a consumer surplus value. The propose scheme is focused on user satisfaction from an economical point of view. If the user pays for the data transfer less than what they are willing to pay for that particular data (file) which is transfered they make profit. Based on the estimated transfer time on each particular network and the corresponding cost the appropriate utility function is chosen using user risk attitude as a decision metric. Three distinct user risk attitudes are considered: neutral, risk seeking (users prefers the alternative of less delay to money saving) and adverse (users prefer to be certain of paying less). In this work no multimedia traffic was considered.

Niyato and Hossain [97] study the dynamics of network selection in heterogeneous wireless networks using the theory of evolutionary games. They propose two evolutionary game theory based network selection algorithms. The first approach uses payoff information for each separate user calculated using a throughput-based utility function and the monetary cost. This info is maintained by a centralized controller. The system tends to reach an equilibrium state where all the users receive an equal share of the available bandwidth. In the second approach each user tries different networks and based on the received throughput and the monetary cost decides if the network has to be changed. This last approach uses reinforcement learning technique.

Rehan et al. [98] present a cross-layer user-centric approach to network selection for vertical handover. This approach is based on two sets of criteria. One is user preferences in terms of which network to be used for which type of application. The second is the link layer triggers provided by the IEEE 802.21 Media Independent Handover standard. This cross-layer approach has the advantage of bringing together information from different important points of view the application layer and the user and the lower link layer. However it relies too much on user preferences who may not always make the best choice.

Sehgal and Agrawal [99] propose a QoS-based multi-attribute network selection algorithm for 4G networks. The proposed algorithms consider the cost of service, bandwidth utilization, call drop probability during handover, power requirements as well as security level and area covered. A distance function (i.e. Spearman footrule) is used to calculate an optimally ranked list of available networks. This method uses a list of priorities assigned by the user to some of the parameters considered in the decision making. These priorities are scaled so their sum is 1 and give weights to the distance function. A modified version of Borda's method of rank aggregation is used for network selection.

3.2.3.5 Multiple Attribute Decision Strategies

A typical Multiple Attribute Decision Making (MADM) problem is to choose an alternative from a set of options, each one characterized by a set of attributes. This is exactly what a handover decision making does. A certain network has to be chosen from a list of available networks, each network being characterized by a set of attributes (bandwidth, cost, power consumption, etc.).

There are various MADM methods including Simple Additive Weighting (SAW), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), Analytic Hierarchy Process (AHP), and Grey Relational Analysis (GRA). An evaluation of three of these methods can be found in [100].

Next some MADM solutions are discussed in details.

Song and Jamalipour [101] introduced a user-centric network selection algorithm based on two mathematical methods to analyze and make the tradeoff between the multiple attributes considered. The attributes considered include availability, throughput, timeliness, reliability, security and cost. All this information is classified into three categories: user preferences, service class and network performance. The mathematical techniques used are AHP for processing the subjective information by pairwise comparison and GRA for prioritisation of the available networks and decision making. The solution was tested and showed good performance in UMTS/WLAN environments.

Godor and Detari [102] introduced a similar approach to network selection. The multiple attributes used are grouped in three distinct categories: service characteristics, user profiles, network facilities. User profiles include service requirements, monetary cost and throughput. Service characteristics include QoS classes, minimal necessary resources and sensitive parameters (real-time service, data lost). Network facilities include network-related parameters like QoS parameters: throughput, jitter, delay, BER. The mathematical techniques used are again AHP for processing the subjective information and GRA for prioritizing the available networks and decision making.

Bari and Leung [103] proposed a multi-attribute network selection approach based on iterative TOPSIS for heterogeneous wireless environments. The following are the attributes considered to be representative for this decision making process: Cost per Byte (CB), Total Bandwidth (TB)(e.g. 54 Mbps for IEEE 802.11a), Allowed Bandwidth (AB)(bandwidth allowed by the access network on a per user basis), Utilization (U), Packet delay (D), Packet Jitter (J), Packet Loss (L). Using this information and based on TOPSIS the network that performs the best in the current context is selected.

Sheng-mei et al. [**?**] proposed a Signal to Interference plus Noise Ratio (SINR) and Analytic Hierarchy Process (AHP) based SAW(SASAW) decision making strategy for vertical handover. This solution uses the combined effects of SINR, user required bandwidth, user traffic cost and available bandwidth of the available networks as decision attributes. The presented solution shows good performance when compared with other two similar algorithms.

3.2.3.6 Fuzzy Logic and Neural Network based Strategies

Fuzzy Logic and Neural Networks-based strategies may also be used in the network selection process prior to handover execution. These techniques are combined with the multiple attribute concept for developing advanced network selection strategies. As the data involved in a handover decision making is often imprecise and the multiple attributes decision making methods present reduced efficiency, fuzzy logic may be employed to improve the performance. Next several fuzzy logic and neural-networks-based solutions are presented and analyzed.

Pahlavan et al. [104] proposed a neural network-based algorithm. When a RSS drop is detected the handover decision is made by a three-layer back propagation neural network which employs pattern recognition. In the context of WLAN/GPRS handover based on a Mobile IP, the output of the decision making algorithm is a binary signal: 0 means that the mobile device should continue communicating with the WLAN AP and 1 means that the host should handover to the GPRS BS. The proposed scheme performs better than traditional RSS-based handover decision algorithms however, it requires prior knowledge of the radio environment and involves complicated deployment.

Chan et al. [105] proposed a solution based on fuzzy logic for decision making in wireless environments where terrestrial (GPRS and UMTS) and satellite mobile networks operate alongside one another. The handover decision algorithm aims at selecting a network for a particular service that can satisfy objectives such as low cost, good RSS, optimum bandwidth, low network latency, high reliability and long battery life. The preferred access network is considered and the input from the user is used to determine the weighting given to each of these criteria. For example the user may decide that the cost is more important than the received QoS.

Xia et al. [106] proposed a fuzzy logic-based handover decision making algorithm for WLAN/UMTS environments. This scheme includes a pre-decision unit which uses mobile device velocity and RSS to decide if a handover is necessary. After the prediction, fuzzy logic is used to decide on handover and new network choice based on three inputs: RSS, predicted RSS and bandwidth. This approach reduces the number of unnecessary handovers and avoids the ping-pong; however fixed weights are assigned to the input attributes reducing the flexibility and adaptability to network variability.

Guo et al. [107] introduced a decision making algorithm which combines both neural networks and fuzzy logic approaches. The algorithm includes a Modified Elman Neural Network (MENN) for predicting the number of users after the handover (input of the adaptive multi-criteria decision) and a Fuzzy Inference System (FIS) for analysis of relevant criteria and final decision making. Simulation results in UMTS/WLAN scenarios demonstrated that the adaptive multi-criteria vertical handover decision algorithm proposed performs better in terms of QoS in comparison with conventional algorithm based on RSS.

3.2.3.7 Context-Aware Strategies

Context-aware strategies consider mobile host-related and network-related context information in the handover decision process. Changes in context trigger the network selection procedure and consequently the handover. The context information includes network-related parameters (e.g. bandwidth and coverage), user-related (e.g. user profile and preferences), terminal-related (e.g. location and power resources) and service-related parameters (e.g. QoS requirements).

Balasubramaniam and Indulska [108] presented a framework for context-aware handover decision making. The framework consists of two parts: the context repository and the adaptability manager. The context repository gathers, manages and evaluates context information while the adaptability manager decides on adaptation to context changes and handover execution. The decision making process is based on the AHP method and aims at maximizing user device preferences and application bandwidth and at the same time minimizing jitter, delay, loss and bandwidth fluctuations. The proposed solution was evaluated using several types of access networks (Ethernet, WLAN, GPRS) in the context of streaming applications. The results showed good performance, offering smooth adaptation to context changes. However the main drawback is represented by the single point of information gathering, i.e. context repository.

Wei et al. [87] proposed a similar approach; however in this case the context information is distributed over several repositories. A user profile repository, a location information server and a network traffic monitor are used to store the context information. The deployment process was simplified as well as the decision process which presents low latency. The technique used is based on software agents which prepare the collected context data and the handover algorithm needed at the collection point. These agents are then downloaded in advance at the decision point which is basically the mobile terminal and invoked when necessary. This approach presents good performance and higher flexibility.

Ahmed et al. [109] developed an intelligent handover decision algorithm based on AHP which includes the session transfer. This is a mobile-initiated and controlled solution and the context-aware decision algorithm is performed for each service type separately. The main objectives are defined as lowest cost, preferred interface and best quality (i.e. maximizing throughput, minimizing delay, jitter and BER). For simulation purposes the authors have considered a multi-mode mobile terminal with two WLAN interfaces and an interactive service (web browsing application) as the active application. The decision algorithm proved to be highly versatile with low delays in both decision making and session transfer which makes it suitable for delay sensitive applications.

Mokhesi and Bagula [110] proposed a novel context-aware handoff decision process based on a multi-criteria decision making model. The approach aims at providing a QoS-based mobility service in wireless environments. Artificial intelligence methods based on the Bayesian Belief framework are used in the decision making process. The system is based on a distributed middleware infrastructure running on fixed hosts in the wired network, proxies, and on the mobile device, client stubs. The proxy agent gathers context information specific to the network while the client stubs consider the client side context information (i.e. Service Level Specification, User preferences, Device characteristics and properties). Based in this information, the handover decision is made. The proposed solution shows good performance and flexibility.

3.2.4 Handover Execution

Host mobility with data session preservation may be achieved by using dynamic routing protocols. Although such a solution may provide efficiency and a reasonable level of performance, its main drawback is scalability. Consequently dynamic routing protocols may provide mobility for small scale networks [111], but not for large scale networks and consequently not for the Internet.

Host mobility in an IP network environment may occur at different layers of the hierarchical organization of the Internet. Mobility can be at micro level, when a mobile device moves to a different attachment point in the same sub-network, at macro level when the mobile device switches to a new sub-network, and at global level when the new sub-network is a different administrative domain [112].

The complexity of the handover operation varies from one situation to another and each case should be treated individually. From the point of view of communication technology mobility may occur between two networks using the same technology in which case the handover is horizontal or using different technologies, in which case the handover is vertical.

Although the layer which is the best suited to accommodate mobility is subject to dispute [113] handover solutions were proposed at different layers of the protocol stack. The handover mechanisms and handover support solutions are further discussed for each layer separately.

3.2.4.1 Lower Layers

Link layer handover procedures are already standardized in the IEEE 802.11 family (IEEE 802.11 a/b/g). The signaling procedures are implemented in the MAC layer and allow a mobile host to attach to different access points within the same subnet. A wireless access point (AP) and its attached mobile stations form a Basic Service Set (BSS). Several BSSs may be connected to a common distribution system (DS), which is basically a wired network, forming an Extended Service Set (ESS). The handover occurs when a mobile node changes its point of attachment to the network from one AP (BSS) to another. Depending on the level of AP congestion, the security policies the handover may be faster or slower. Although packet loss is avoided trough buffering at the mobile host and AP, the higher layers may detect an increase in packet delay. There are various solutions which improve the 802.11 handover performance [114, 115].

In order to facilitate high quality, loss free handover the IEEE 802.11f standard proposes the Inter Access Point Protocol (IAPP). IAPP facilitates inter AP communication which permits a better synchronization in order to perform a faster and more reliable handoff and to preserve the QoS context of the mobile node [116, 117].

Link layer handover works only for intra subnet mobility. Inter subnet mobility involves acquiring a new IP address which require interaction and support from the higher layers especially the network layer (IP). Although the link layer handover is quite fast and may be performed without any packet loss, but with an increase in packet delay, the higher layer may still react negatively to these QoS variations. QoS variations determined by the wireless link are different in nature than those determined by congestion and should be treated differently by the higher layers (especially transport layer congestion control). Consequently there are several link layer solutions which aid the higher layer protocols, mainly the TCP, to cope better with the link layer QoS characteristics. Such enhancements are Snoop TCP [118] and SPLADE [119], the later being designed for real time multimedia applications, which use buffering and local retransmission techniques to shield the TCP layer from the wireless QoS variations.

3.2.4.2 Network Layer

The network layer is where the IP protocol resides in the Internet architecture. The IP protocol is responsible for routing data packets from their source to the destination based on the IP addresses. The IP addresses are allocated by an administrative entity and are specific to sub-networks of the Internet. The IP addresses have a dual role, both as a host identifier and as a location tag within the Internet.

There are two approaches to provide mobility at the network layer. One is to use host-based routes and update them as the host moves. The other uses indirection agents to reroute the traffic to the mobile host through the visited sub-networks.

The main mobility management solution for Internet using the indirection agent approach is Mobile IP [120, 121]. Although the host-based mobile routing approach is note scalable enough for an Internet level mobility, this approach is used for domain level mobility (micro-mobility).

Mobile IP is a network layer mobility solution, first developed as an extension of IPv4, Mobile IPv4 (MIPv4) [120] and then as part of the IPv6 standard, Mobie IPv6 (MIPv6) [121].

In **MIPv4**, the mobile node (MN) is assigned a fixed IP address corresponding to its home network which uniquely identifies the MN. While roaming through different foreign networks the MN will use a home agent (HA) to keep track of its current location and to intercept and tunnel the data packets originating at the corresponding nodes (CN) to its current location. The mobility mechanism of Mobile IP is schematically presented in Figure 3.3.

The data traffic originating at the MN uses the home address as the source address and is routed directly to the CN or reverse tunneled in case ”ingress filtering” is implemented at router level. While roaming away from the home network, the MN will constantly send binding updates to inform the HA about its current location-specific address (denoted ”care-of-address”).

The main advantage of this solution is that the CN is not required to implement any protocol extension. The main disadvantages are the necessity of deploying a home agent in the home network and eventually foreign agents in the foreign networks. Routing the data traffic through the HA (triangular routing) is also a major efficiency drawback of MIPv4.

MIPv6 [121] was developed as part of the IPv6 protocol. The mode of operation is similar to MIPv4, but it presents certain differences and offers several improvements. The main improvement offered by MIPv6 is route optimization. In MIPv6 the MN is able to

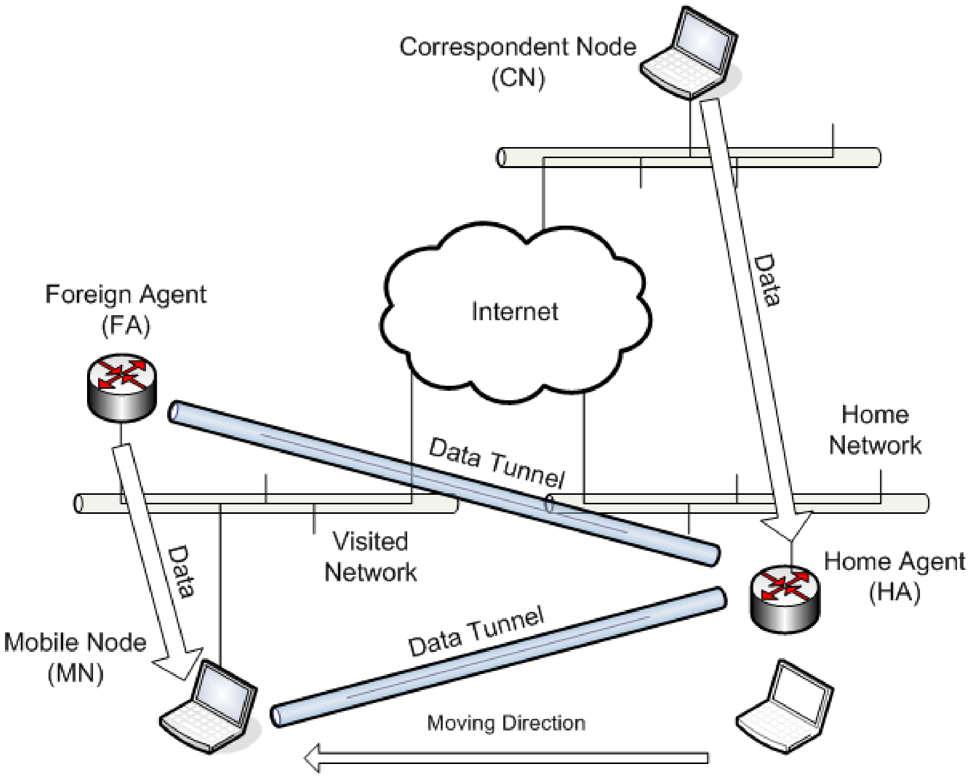

Figure 3.3: Mobility management services and components

register its current care-of-address with its CN. Using this feature the corresponding nodes are able to send data packets directly to the mobile node's care-of-address avoiding the inefficient triangular routing when relaying the packets through the home agent. MIPv6 also provides good security mechanisms to mitigate the risks involved by node mobility.

Although MIPv6 presents better performance than IPv4, it still suffers from quality degradation due to high handover delay determined by new care-of-address registration, this is especially evident when the HA or CN are far away from the current MN location.

Several extensions for Mobile IP were proposed to enhance the performance of the basic mobility model. For example IP micro-mobility protocols enhance the basic Mobile IP by managing local mobility (i.e. within a domain) offering fast and seamless handover [122, 123] while improving efficiency and power consumption.

Cellular IP [124] integrates host localization and handover techniques with the routing mechanisms. It assumes that the wireless access network is connected to the Internet via a gateway node and micro-mobility is performed by Cellular IP at domain level. Macro-mobility is performed by Mobile IP at the Internet level.

Cellular IP presents good handover performance for local mobility, but its main drawback is the necessity to be implemented by the majority of routers within the wireless access network. Another significant drawback is represented by the constraints imposed on the access network architecture. The use of a common gateway which acts as a foreign agent may have a major impact on the reliability of this solution as it represents a single potential point of failure.

Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [125] offers intra-domain mobility support, while relaying on Mobile IP for inter-domain mobility. HAWAII aims at enhancing efficiency, scalability, reliability and QoS while reducing data traffic disruption. While roaming through HAWAII-enabled domains, the mobile hosts retain their IP address; the packets are routed to destinations based on routes established using specialized path setup schemes. Each administrative domain is assumed to have a gateway router called the domain root router. The domain root router is responsible for receiving the packets addressed to the MH and routing them to destination based on the dynamic host-based routes established in some of the intra-domain routers.

Hierarchical Mobile IPv6 (HMIPv6) [126, 127] uses a network organization based on domains. Each domain contains several access routers (AR) and a Mobility Anchor Point (MAP) which connects the domain to the Internet. Any MN has a regional care-of-address which is registered with the HA and CNs and represents the location of the MN at higher level. The MN also has an on-link care-of-address which is registered with the MAP and represents the location within the domain. The HA and CNs send the data packets to the regional care-of-address which is basically the address of the MAP. The MAP receives the packets and tunnels them to the on-link care-of-address of the MN.

This solution reduces the handover delay and loss by performing a micro-level address registration which takes less time for binding updates. There is still the macro-level handover latency (when MNs passes from one domain to another) which involves high handover latency.

Although by using the route optimization option (Mobile IP RO) the pressure on the home agent is reduced, it still represents a single point of failure and a possible bottleneck leading to performance and reliability problems. A few solutions have been proposed to avoid relying on a single home agent or home network.

Mobile IP with location registers (MIP-LR) [128] uses multiple distributed home location registers (HLR) which keep track of node locations. Each host trying to reach a mobile host queries a HLR for the current location of the mobile node. The main disadvantages of this solution are the additional infrastructure requirements and lack of transparency for the corresponding nodes [19].

A **homeless extension to MIPv6** [129] was also proposed. Using this solution a mobile host may operate without a unique home address. For each connection the homeless mobile host maintains a host cache which holds a list of local addresses and a foreign cache which holds a list of destination addresses. The connections are no more bound to IP addresses but to the host cache. In order for this solution to work both peers involved in a data connection have to support homeless operation, option which has to be negotiated at the connection initiation.

3.2.4.3 Mobility at a New Layer

One of the main impediments in maintaining high QoS while allowing full host mobility is the dual role of the IP address and its association with the transport end-points. The IP address contains both the host identifier and the host locator (subnet). Being part of the end-point bounding in a transport session, changing the IP address will definitely disrupt the ongoing data transfer if appropriate transport layer measures are not taken.

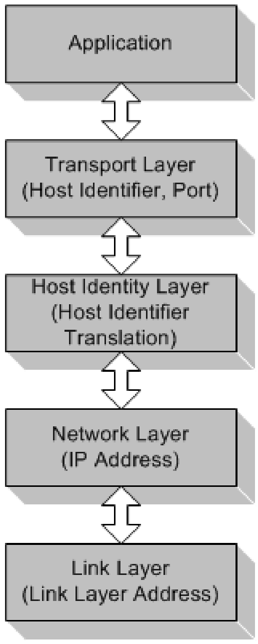

Figure 3.4: Host Identity Protocol paradigm

The **Host Identity Protocol (HIP)** [130] proposes a separation of the identifier-locator role of the IP address. Figure 3.4 schematically presents the principle of HIP.

In a HIP compatible host, the IP address will be only the locator while the identifier will be a public key called Host Identifier (HI). The HI will be responsible for host identification and authentication and will replace the IP address in transport layer associations. HIP defines a new layer between the network and transport layer where the HI will be associated with an IP or a set of IPs. From the host mobility point of view HIP allows for a more natural operation of multiple addresses allowing relatively easy address change. When the mobile host changes its point of attachment for the outgoing data messages, not much interaction is needed as the corresponding node will identify the received messages based on the HI and not based on the source IP address which may have changed. Regarding the incoming traffic the corresponding node has to be informed about the IP address change through HIP Readdress message. The corresponding node performs a host reachability test before starting communication. Location management may be obtained through DNS or rendezvous servers.

HIP may provide seamless mobility mechanism through its more natural and efficient use of multihoming. The main disadvantages are related to deployment (requires changes in APIs or development of new APIs) and the handover overhead which may be prohibitive for frequent network changes.

3.2.4.4 Transport Layer

The packet loss and packet delay induced by the error prone nature of wireless links and mobility procedures negatively impact the solution, especially the congestion control mechanisms implemented at the transport layer. This is mainly the effect of the lack of information received by the transport layer regarding the wireless link conditions or the presence of handover. By providing mobility at the transport layer the two features, handover management and congestion control, can work in conjunction to provide seamless mobility management.

Several advantages may be outlined in favor of transport layer mobility. There is no triangular routing; each connection can benefit from the route optimizations features of the network layer. There is no single point of failure in terms of home agents or foreign agents. There is also little dependency on a certain infrastructure as opposed to the network layer solutions. The mobile host uses its current IP address in the source field of data packets (the correct address for the current sub-network) avoiding the possible security-related packet blocking at router level.

Transport layer mobility suffers from lack of location management support. The solution is to rely on dynamic DNS which can solve the problem, although performance issues may arise in relation to timely efficient update of the DNS servers.

One of the most popular transport protocols in the Internet is the **Transmission Control Protocol (TCP)**. TCP is a reliable connection-oriented protocols; its congestion control and congestion avoidance mechanisms have been proved to be very effective in wired networks.

Unfortunately the standard TCP implementation does not present the same effectiveness over the wireless networks. This is mostly because packet loss and packet delay, which in case of wired networks are clear sign of network congestion, may appear in wireless networks due to channel error, or host mobility or various other environmental interferences.

TCP enhancements for wireless networks can be divided in two categories, solutions for improving the behavior of TCP in the presence of wireless networks and TCP-based handover algorithms [131].

Among the mobility solutions proposed for TCP are **TCP-Redirection (TCP-R)** [132], **MSOCKS** [133], **Mobile TCP** [134] and **TCP-Migrate** [135].

TCP enhancements for wireless networks include **Lightweight Mobility Detection and Response (LMDR) TCP** [136], **Indirect TCP (I-TCP)** [137], **Snoop TCP** [118] and **Freeze TCP** [138]. TCP versions supporting real-time applications are **TCP with real-time mode (TCP-RTM)** [139] and **TCP-MR (Minimum Rate)** [140].

Mobility support solutions with improved reliability have been also proposed for the **User Datagram Protocol (UDP)** protocol as well. **Mobile UDP (M-UDP)** [141] uses a similar approach to I-TCP in splitting the connection at a supervisor host (SH) which is responsible for managing host mobility within the cells (controlled by the SH) and retransmission of lost packets within the available bandwidth to maintain the packet loss as low as possible.

The newly developed **Stream Control Transmission Protocol Mobile (SCTP)** [142] offers a very important feature for mobility which is multihoming. The mSCTP handover mechanism is schematically presented in Figure 3.5.

In SCTP the MN and CN negotiate a SCTP association where both end-points maintain a list of IP addresses corresponding to the MN and CN. The handover algorithm proposed for SCTP, **Mobile SCTP (mSCTP)** [143] uses the ADDIP SCTP extension [144] which allows the two pairs involved in an association to manage the list of IPs associated with each end-point. When the MN changes its location and acquires a new IP address it can send an Address Configuration Change (ASCONF) Chunk with an Add IP Address parameter to inform the CN of the change in IP address. If the MN wishes to redirect the traffic on the newly acquired IP address it can ask the CN to set the new IP address as the primary address. In this manner the traffic flow originating at the CN will be transferred to the new IP address.

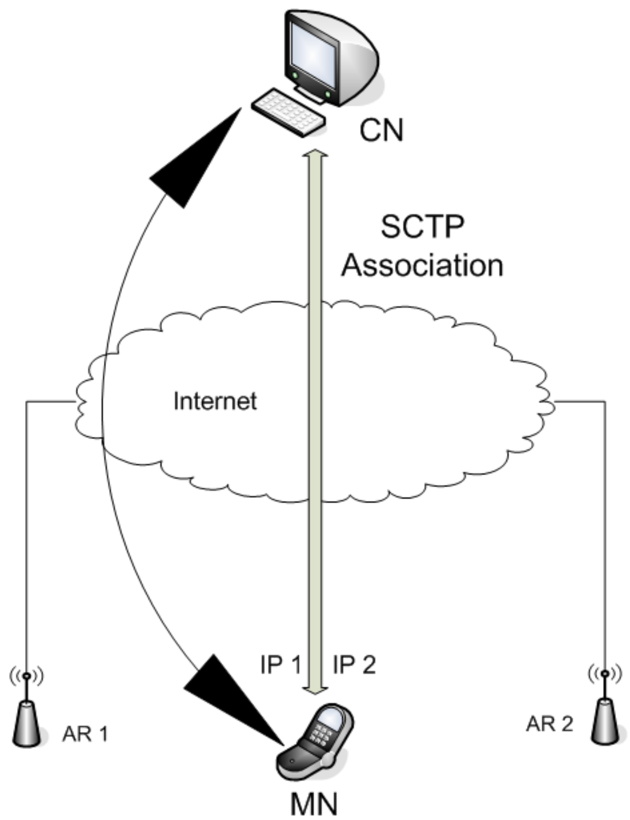

Figure 3.5: SCTP handover mechanism

The operations sequence performed during mSCTP handover is presented in Figure 3.6.

Cellular SCTP [145] improves SCTP-based handover by sending a duplicate stream to the old network as well as to the new network. This solution aims at minimizing traffic disruption at the cost of efficiency when the new and the old network are topologically far.

A similar approach is the **Transport Layer Seamless Handover (TraSH)** [146] which also offers location management through a location manager which maintains a list of associations between the mobile node and its current IP address.

A mobility extension for the **Datagram Congestion Control Protocol (DCCP)** [147] which is another recently introduced transport layer protocol is presented in [148]. Mobile DCCP uses a generalized connection which groups several normal DCCP connections. When the MN changes its location a new connection is added using the new IP address. The traffic is transferred to this new connection while the old one is deleted from the generalized connection.

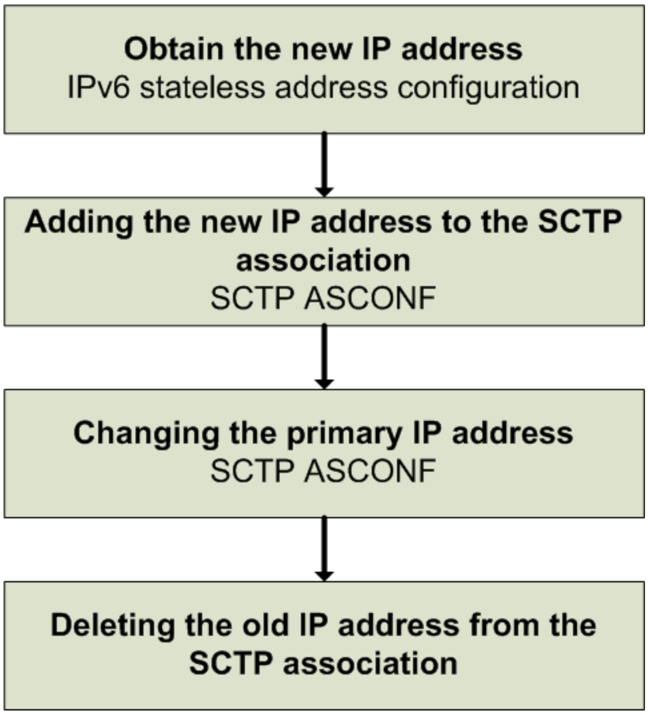

Figure 3.6: mSCTP handover operations sequence

The mechanism of Mobile DCCP is presented in Figure 3.7.

Figure 3.8 presents the sequence of operations involved in handover when using DCCP generalized connections. This solution can provide seamless handover although an efficient algorithm for managing the traffic over this group of normal connections is not specified.

3.2.4.5 Session and Application Layers

Providing mobility at higher layers in the Internet protocol stack may present the same advantages as the transport layer mobility including some more outcomes. The higher layers like the application layer have a better view of the current data sessions and the application QoS requirements. Providing mobility at the application layer would decouple even more the mobility management from the network infrastructure by avoiding changing the possible mature and well defined transport protocols.

On the other hand more consistency may be achieved with higher layer mobility as it may be problematic to have various transport protocols running concurrently on the same

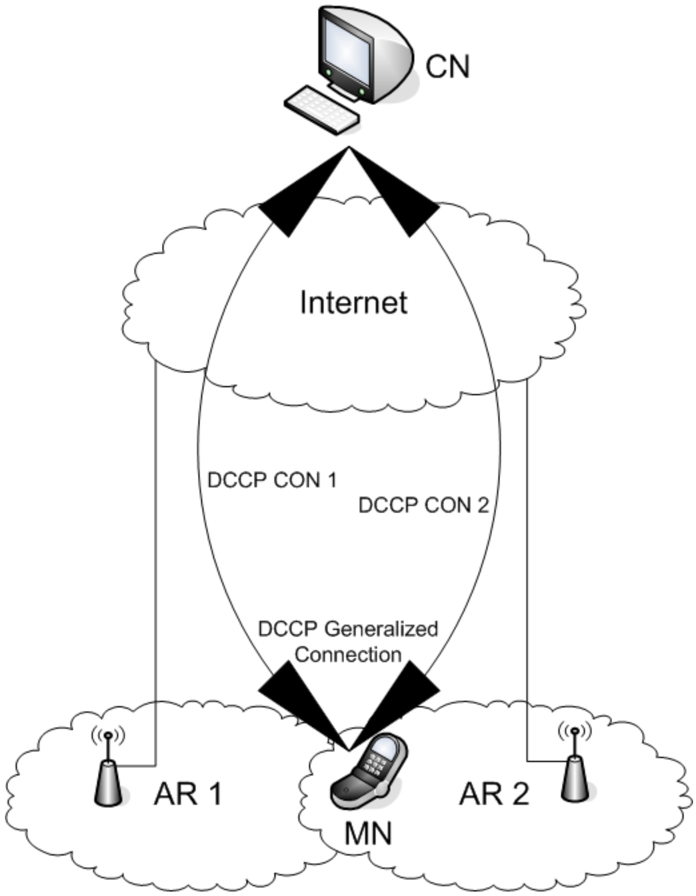

Figure 3.7: Mobile DCCP handover mechanisms

host machine and dealing with mobility in a different way especially related to network state evaluation and decision making as well as location management updates.

Session Layer Mobility Management (SLM) [149] is a session layer mobility framework which operates above TCP. SLM supports mobility and QoS using a session management module which switches TCP streams between different connections. The application traffic is intercepted by the session manager which is responsible for managing the TCP connections and the creation of sockets making the whole process transparent.

The **Session Initiation Protocol (SIP)** [150] was developed for multimedia signaling at the application layer. Several mobility support solutions using SIP have been proposed in

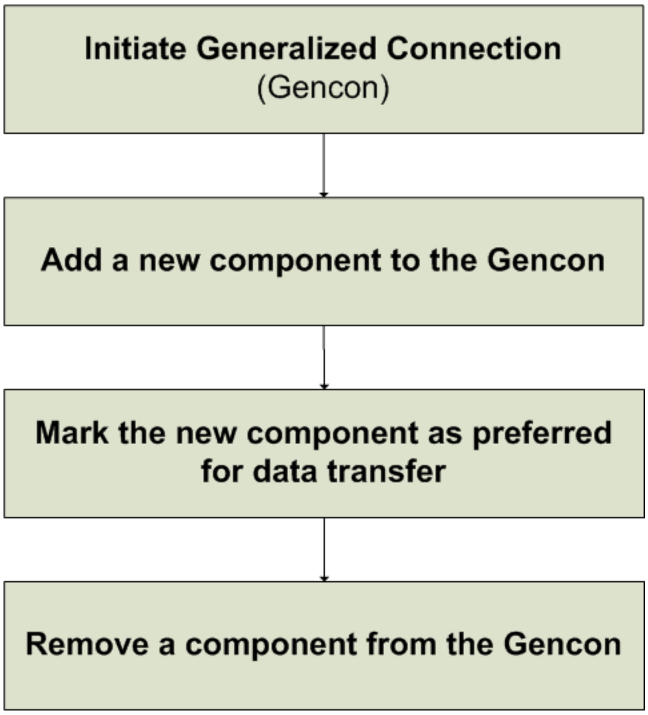

Figure 3.8: DCCP generalized connection handover operations sequence

the literature [131]. The main idea of handover management using SIP is based on sending a RE-INVITE message to the corresponding node informing it about the new IP address of the mobile node. Upon receiving a RE-INVITE message the corresponding node starts sending the data to the new IP address.

Handover management using SIP may suffer from increased latency due to signaling messages procedures and the overhead for IP encapsulation. Performance improvement solutions have been proposed for SIP mobility by using intra-domain optimization and Predictive Address Reservation SIP (PAR-SIP) which allows for proactive address reservation and session updates.

MOBIKE [151] represents a mobility extension for the Internet Key Exchange version 2 (IKEv2) [152] signaling protocol. MOBIKE maintains the security associations active to avoid re-initialization. When the mobile user is changing is point of attachment the MOBIKE detects user mobility by its mechanisms to detect dead peers, and initiates an IP address update notification.

3.2.5 Application Specific Integrated Solutions

Integrated mobility management solutions for specific applications like multimedia streaming have been proposed. These solutions support horizontal, vertical or both horizontal and vertical handover, also integrating handover initiation and network selection procedures.

M4: MultiMedia Mobility Manager [153] uses Multihomed Mobile IP for handover support and a simplified version of Relative Network Load (RNL) for network selection. The proposed solution supports vertical handover and was evaluated using WLAN and CDMA2000. The RNL-based networks selection algorithm uses the round-trip time (RTT) and RTT jitter values of binding updates to compute the grade for each available network. The main drawback of using only RTT for decision making is the lack of information related to network bandwidth and loss level which may negatively impact on multimedia quality levels. The multimedia mobility management solution proposed in [154] uses proactive buffering to perform seamless handover and select networks based on the received signal strength (RSSI). The main drawback of using signal strength only for network selection is impossibility of detecting network's level of congestion and bandwidth capacity as well as amount of delay and packet loss.

The **Unified Mobility Scheme (UMS)** for multihomed mobile nodes is presented in [155]. This mobility solution uses SIP for multihoming support and location management. When an interface of the multihomed node acquires a new IP address the corresponding node will be informed about this through a binding update. UMS supports both horizontal and vertical handover. UMS interacts with the application for QoS requirements and performs traffic distribution over multiple links if the QoS requirements exceed the capacity of only one link.

3.3 Load Balancing in Wireless Networks

Accommodating the increasing number of wireless network users with their demanding QoS requirements represents a challenging task for network operators. The heterogeneous and ubiquitous wireless environment which is now available, increases the pressure on the network management system (NMS).

Load balancing is among the most important component of a network management system. Load balancing has a great impact on both network resource usage and user perceived quality of service.

Mobile users have a fairly random behavior with respect to mobility and bandwidth requirements. Although certain patterns can be detected by studying mobile user behavior, initial network deployments and later short-term maintenance can be performed accordingly. However there is no long term guaranteed network stability.

In order to ensure an efficient network resource utilization and high user perceived quality, load balance strategies have to be employed to cope with the more or less often situations when the entire network or parts of the network get overloaded. This is important for two main reasons.

First, from the operator's point of view, wireless networking resources are scarce and expensive. Wireless spectrum, except for the unlicensed bands, requires licensing which is very expensive. Under these circumstances an efficient utilization of the wireless spectrum is extremely important for the cost effectiveness of the entire system.

Second, unbalanced network load leads to situations when the capacity of some access points or base stations is exceeded resulting in mobile users being denied connection or current users experiencing severe disruptions of their current data sessions. All these have a negative impact on user Quality of Experience (QoE) and consequently on users desire to pay for that specific service.

Load balancing algorithms can be grouped in two main categories: network-driven approaches and user-driven solutions [97].

Network-driven approaches are centralized, requiring a central controller which gathers load information for the entire network and makes the appropriate decisions for balancing the load. The main disadvantage of such an approach is the single point of failure represented by the centralized controller as well as the high communication overhead involved. Additionally there are serious scalability limitations. The advantage is more efficient decision making which is based on excellent overall view of the entire network status and load distribution.

User-driven approaches are distributed; the algorithms are implemented on the user side. These approaches are more robust and involve less communication overhead. However being distributed, the long term stability is not guaranteed and the decision may not be optimal for the entire network. This is determined by the lack of information regarding the overall status and load distribution over the entire network.

There are three main situations involved in load balancing. **Admission control** decides if a new mobile host is allowed or not to associate itself with an access point or base station based on the load of that particular wireless network.

Once a mobile host is connected and started its data sessions the admission control policies cannot be applied anymore. However due to variations in applications traffic requirements and migrations of other users toward other wireless networks, unbalanced load distribution may appear. In this context load balancing can be achieved by **forcing handover** of some users from one network to another in order to evenly distribute the load among adjacent and/or overlapping networks.

Load balancing can also be achieved by efficient **traffic distribution over multiple networks** in the case of multi-modal devices capable of parallel communication with multiple networks simultaneously. Although this approach offers a better and more efficient utilization of the existing network resources it involves various challenges related to traffic splitting, data distribution over the available networks and traffic merging.

In the following sections, various load balancing strategies will be presented and discussed.

3.3.1 Admission Control in Load Balancing

One step in keeping the network balanced is to control the association between the mobile host and the access point or base station. The decision can be made either by the network (network-centric) or by the user (user-centric).

The classic user-AP association is based on the received signal strength indicator (RSSI) level and the AP with the highest RSSI is preferred. This type of association does not provide any support for load balancing among the available APs. Consequently some manufactures have upgraded their wireless LAN APs to broadcast load information represented by the number of mobile nodes associated with each particular AP. This improves the load balancing in the network; however mobile hosts generate variable amounts of traffic as well, and this is not considered. Consequently the number of mobile hosts associated with an AP and RSSI level do not accurately represent network load level and other solutions are required.

Balachandran et al. [156] presented a centralized load-balancing scheme for congestion relief in public-area wireless networks. A central access server collects the load level of each AP and decides based in this information to which AP the new mobile host should connect to. The proposed scheme also offers assisted user-AP association by explicit channel switching. This method allows for signal strength tradeoff for reduced load. Network directed roaming is also proposed, where users are directed toward the less loaded areas of the network. This algorithm increases the load balance in the wireless network by 30% and bandwidth allocation by 50%.

Bejerano et al. [157] proposed a set of algorithms for load-balancing and fair bandwidth distribution in wireless networks. The proposed scheme is centralized and as novelty it considers both the load level of the wireless network and the load of the wired backbone link. This scheme controls user-AP associations targeting max-min fair bandwidth allocation. The simulation-based testing of the proposed scheme showed that it reached near optimal load balancing within the wireless network.

Scully and Brown [158] introduced a load balancing technique based on genetic algorithms. They proposed two versions of this strategy, one based on a standard genetic algorithm denoted MacroGA and one using a micro-genetic algorithm refereed to as MicroGA. The algorithms aim at allocating users to access points in such a way that the network throughput is improved and the users get maximum QoS. The users are organized in classes according to their application requirements. Each user is assigned a certain bandwidth depending on his/her location related to the access point and the class of users he/she belongs to. The two load balancing techniques proposed proved to be more efficient than the traditional RSSI-based ones with MicroGA being a good choice for time-critical scenarios.

Various other similar solutions have been proposed in the literature; however although access control involves to a certain extent a decision making process similar to the one used in handover, it usually does not use a complex set of metrics which is usually required for an efficient handover management process.

3.3.2 Handover in Load Balancing

Handover plays a very important role in load balancing. Handover allows active data sessions to be transfered from one network to another. Under these circumstances if a certain access point or base station becomes overloaded, some of the currently attached users can be handed over to neighboring networks according to a predefined load balancing strategy.

Nunzi et al. [159] introduced a self-configuring (decentralized) algorithm for load balancing in wireless networks. The load balancing is performed in two stages. In the first stage the load level of each AP is determined relative to three predefined thresholds: Max Load, High Load and Light Load. In the second stage the adjacent APs are negotiating their power level which will later determine mobile nodes to handover from one wireless network to another. Oscillation avoidance techniques are employed to maintain system stability. The algorithm aims at minimizing the number of users who are transfered to another AP in order to reduce the overhead. The algorithm presents a good trade-off between efficiency and control overhead; however under high load circumstances unpredictable behavior can negatively impact on user perceived quality of service levels.

Ha el al. [160] proposed a load balancing strategy for heterogeneous wireless environments based on the concept of community. A community consists of several wireless networks (using the same or different technologies) covering a certain geographical area. Overlapping between communities both in terms of networks and geographical area is considered. Information about the existing networks within the community and their status is collected and disseminated by a specialized entity within the community. Load balancing is performed based on available bandwidth and received signal to noise ratio. The proposed strategy improves radio resource utilization and reduces the call blocking probability. However its performance is tightly related to the performance of the vertical handover algorithm employed.

Yun et al. [161] introduced a load balancing strategy based on cell breathing. Cell breathing is a technique that allows the access points to vary their power and consequently their coverage areas in order to adjust the number of mobile hosts attached. Considering the traditional handover decision making mechanism which is based on the received signal strength, two or more access points can distribute the users and force handover by adjusting their beacon broadcasting power. In the proposed load balancing strategy a central operation agent (COA) collects load information from the access points and based on this information it calculates the power level for each access point in order to achieve an optimum load distribution. The proposed solution improves the overall performance of the network, while reducing the management overheads. The main disadvantage of this approach is the impossibility to select precisely the users which should be handed over to a different network.

Ishizu and Harada [162] developed a framework to collect bandwidth information from terminals, process the network load and feed it back to the wireless terminals. The bandwidth information is processed on the AP Resource Advertisement Server. The information is then unicast to the AP Side Advertisement Translator which then broadcasts it to the mobile terminals. Based on the load information received, mobile hosts handover to the most vacant AP withing their range or not. This technique improves the throughput received by the terminals.

Ong and Khan [163] introduced a handover-based load balancing technique for wireless LAN. The proposed scheme offers both admission control and handover services. Maintaining high QoS levels is the main goal of this approach. Admission control is performed using the packet delay as the metric aiming at preserving the QoS received by the users currently associated with an AP. The handover is triggered by the packet loss rate, and a threshold of 2% is considered for VoIP applications. This solution offers QoS guarantees during and after handover while improving the overall fairness and resource utilization of the network. The main disadvantage is lack of flexibility in terms of handover threshold which is set for VoIP applications only.

Velayos et al. [164] proposed a completely distributed load balancing scheme. Agents running in each AP are gathering load information and determine if the AP is balanced, overloaded or underloaded. Overloaded APs force the handover of some mobile stations to neighboring APs to reduce the load. The load metric used is the throughput relayed by the AP and the system balance is estimated using a β index. The handover stations or new mobile hosts are accepted only by underloaded APs in the pursuit of reducing handovers rates. Experimental results show an improvement in throughput and a decrease in cell delay when the proposed load balancing scheme is used.

3.3.3 Traffic Distribution in Load Balancing

Distributing the main data traffic generated or received by the host over multiple simultaneous connections has been used for throughput enhancement and improved reliability. Although beneficial, this approach presents a set of challenges. Multiple simultaneous connections require multiple interfaces which in turn increase power consumption, hardware complexity and cost. Splitting the traffic at the sender side and merging the received data at the receiver side also involves time, computational complex and power consuming algorithms especially when the active connections present high values of delay and delay jitter. However, traffic distribution over multiple networks has the benefit of improved quality and efficient utilization of radio resources.

Chen et al. [165] have proposed an iterative load balancing algorithm for time and bandwidth allocation among access points and mobile hosts. The aim is to achieve heterogeneous fairness and satisfy application requirements. The algorithm can be used both in a centralized approach (using a central processing unit) or in a decentralized manner. Every mobile host is associated with several access points within its range. The mobile device is allocated a time share of the AP channel availability and based on this and the available total throughput that an AP can provide to its users, the individual user bandwidth ca be calculated. Utility functions are used for each host with the user bandwidth as the main parameter and time share allocation is made by maximizing the sum utility. The proposed solution offers good performance in maximizing user received throughput however it assumes a quasi-static mobility pattern for the mobile users.

Sarkar and Sarkar [166] introduced a load balancing technique for wireless LAN which uses concurrent association of the mobile node with all the access points in its range. The mobile device uses the functionalities offered by the IEEE 802.21 standard to associate with multiple access points simultaneously and then request services from these access points sequentially. The authors propose a solution for access point utilization, expected queue length and waiting time for the mobile host. However the proposed scheme relies on the existence of a concurrent multipath transport protocol at the transport layer. The methods used for traffic distribution and rate allocation are not specified.

3.4 Traffic Distribution Over Multiple Connections

Transporting data over multiple simultaneous connections involves several technical challenges. First, each independent connections has to be monitored and the amount of traffic which is going to be relayed by it has to be estimated. This is basically the rate allocation process. Second, the main application traffic has to be split in several bit streams which are later allocated to independent connections. Each data stream has to have the right bitrate which was previously estimated during the rate allocation step. In the following subsection various techniques involved in multi-connection data transport are presented.

3.4.1 Rate Allocation for Traffic Distribution

Rate allocation is an extremely important step in traffic distribution among multiple paths. In order to achieve minimum distortion of the application traffic and to avoid network overloading and consequent decrease in user perceived quality through low QoS levels, the capacity and status of each distinct path has to be precisely estimated and the right amount of traffic load has to be allocated.

Rate allocation when multiple flows share the same network is a well investigated problem and various solutions were proposed. The most popular is the TCP congestion control [167] scheme which is used by most of the Internet applications. For multimedia applications using UDP as the underlying transport protocol TCP-friendly rate control (TFRC) [168] is the most popular solution.

Various other solutions as well as improvements for the above mentioned techniques have been proposed in the literature. However the situation when multiple simultaneous connections are shared by one or several streams is not considered.

Shakkottai et al. [169] presented a mechanism for splitting the traffic among multiple IEEE 802.11 APs to which the mobile hosts are simultaneously associated. The proposed solution is based on non-cooperative game theory and claims to maximize the throughput justifying the use of multihoming. An operation method for traffic splitting is not presented though and user mobility and traffic characteristics are not considered.

Zhu et al. [170] presented a rate distribution mechanisms for multi-homed video streaming in heterogeneous wireless environments. Multiple streams are allocated to multiple networks in a distributed manner. The rate allocation for each path is made based on observed available bitrate, the round-trip time of each access network and video distortion-rate characteristics of the video content. The goal is to reach an optimum traffic distribution which minimizes the distortion of the video stream. The problem is formulated as a convex optimization framework and the sum of expected distortions for all the streams is minimized. This solution proposes a rate allocation for distributing traffic over multiple connection, but although network dynamics are considered, user mobility is not addressed.

3.4.2 Traffic Splitting and Distribution

Traffic splitting and distribution is a very important step in multi-connection data transfer. It has the role of partitioning the main application data stream into sub-streams and to allocate the data packets to the corresponding connection. In the case of multiple application data streams allocated to multiple access networks the solution is straightforward. Each stream is assigned a network according to the rate allocation procedure. However, when a single data stream (e.g. multimedia stream) has to be distributed over multiple simultaneous connections, traffic splitting procedures and later sub-stream distribution are essential.

Traffic splitting and distribution can be made in three different ways: packet-based splitting, object-based splitting and stream-based splitting. Related to multimedia content the following methods can be used [171].

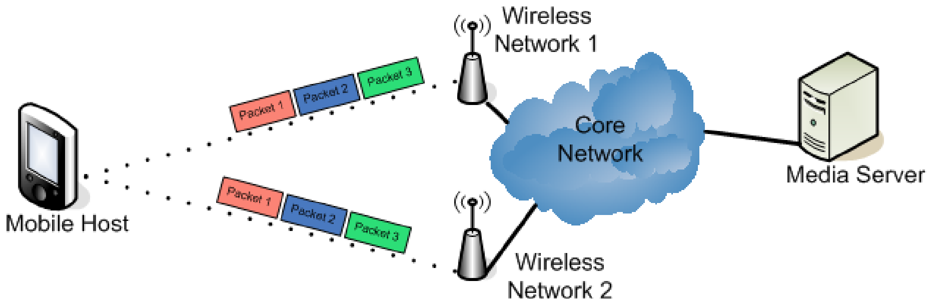

Figure 3.9: Packet-based traffic splitting

Packet-based splitting takes each data packet from the main application data stream and allocates it into different queues, each queue corresponding to a different access network. The length of the queue and the number of packets allocated depends on the predetermined rate allocation. Figure 3.9 presents schematically the packet-based traffic splitting.

Round robin and weighted round robin are two simple methods of packet-based traffic dispersion. The first equally distributes the packets over the existing connections while the other considers the available path bandwidth to weight the allocation rates.

El Al et al. [172] proposed an extension for the SCTP protocol to support load sharing among multiple simultaneous connections. The traffic is allocated based on the previous estimation of bandwidth and the round trip time in order to compensate the differences in delays. Flow control is performed at the association level while congestion control is implemented at the individual path level. Although this solution considers the dynamics of available paths it does not explicitly consider user mobility. Moreover the rate allocation and traffic distribution do not consider parameters like cost of data transfer and energy consumption.

Packet-based traffic distribution suffers from the variable delays of the distinct networks used for data transport.

Frame-based splitting groups all the packets corresponding to a frame into one single logical entity and allocates these entities to the available connections. Figure 3.10 presents schematically the frame-based traffic splitting.

Frame-based splitting suffers from error propagation, especially when MPEG-based

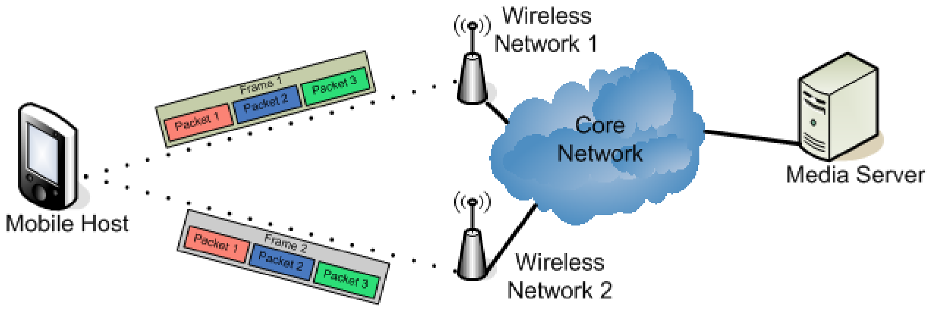

Figure 3.10: Frame-based traffic splitting

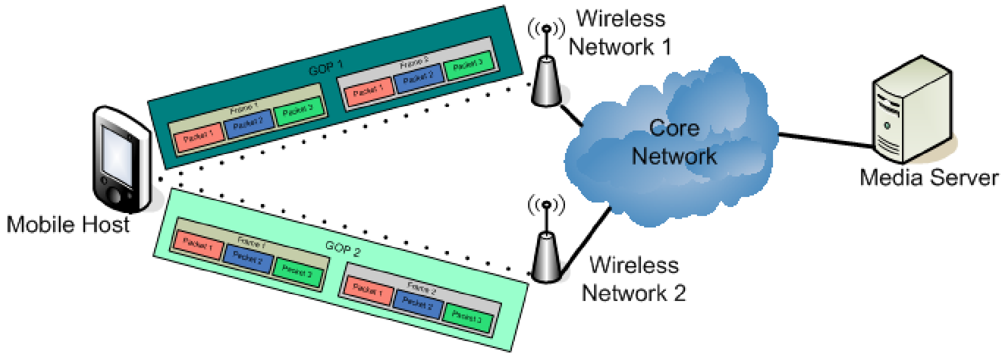

Figure 3.11: GOP-based traffic splitting

encoding scheme are used. The dependency between frames in a group of frames/pictures (GOP) determines a high impact of packet loss corresponding to a single frame. More exactly if a frame is affected by packet loss, all the frames within the GOP which depend on that frame for decoding will be affected. Consequently this type of traffic dispersion is not suitable for multi-connection data transfer in wireless environments.

GOP-based splitting groups all the packets corresponding to the frames which are part of a GOP and allocates these logical data blocks to the existing communication paths. Figure 3.11 presents schematically the GOP-based traffic splitting.

Because there is no dependency between the GOP within the video stream there is less error propagation which makes this technique more suitable for multi-path data transfer in wireless networks.

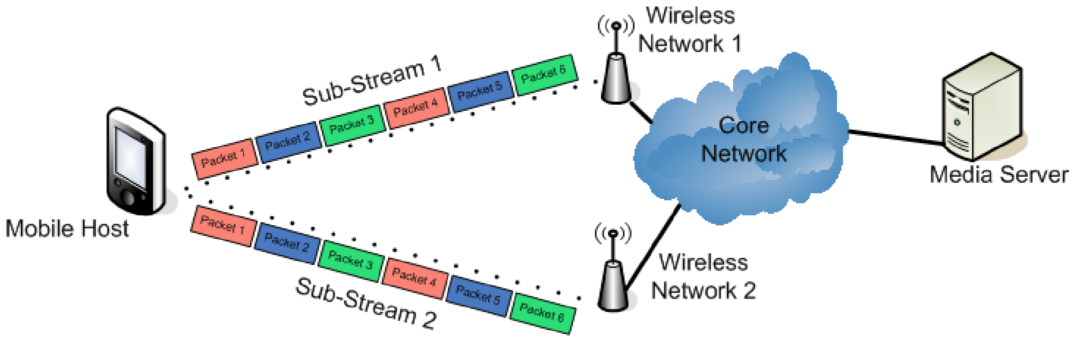

Figure 3.12: Stream-based traffic splitting

Stream-based splitting uses multi-stream video encoding which becomes increasingly popular. The main video stream can be encoded into multiple dependent or independent streams which can be delivered to the receiver over independent paths. Figure 3.12 presents schematically the stream-based traffic splitting.

Independent sub-streams are the most resilient to transmission errors and consequently are the most suitable for multi-connection video content transport.

Wang et al. [173] developed a bitrate allocation strategy based on H.264/SVC. Scalable Video Coding (SVC) encodes the video content into multiple streams. One is the base stream or the base layer and offers the minimum video quality when decoded. This layer has the maximum priority and has to be delivered to the receiver in order to decode the video content. The enhancement layers have less priority and each layer when decoded offers higher video quality over the base layer. This technique can be used to transport the video content over multiple simultaneous connections. However the dependency between layers represents an impediment in the context of wireless links where a connection can be easily lost in the presence of user mobility. Because the base layer is mandatory for decoding the video content this encoding approach is not suitable for mobility management using multiple simultaneous connections.

Multiple description coding (MDC) is more appropriate for multi-path multimedia streaming. In a MDC approach the video content is encodes in multiple independent sub-streams (descriptions). If the multiple descriptions are independently sent over distinct paths and any of the several descriptions is received the decoder can reconstruct the video content with a certain level of distortion. The minimum distortion is achieved when a all descriptions are received and decoded.

The simplest way to produce multiple descriptions is to split the source data into several sets and then compress each set separately to obtain the sub-streams [174].

Vaishampayan and John [175] proposed a multiple description coding scheme for video based on the multiple description scalar quantizer (MDSQ), first proposed by Vaishampayan in [176].

Reibman et al. [177] applied a multiple description method based on correlating transforms to motion compensated multiple description video coding. These methods present good performance however they were developed as stand alone codecs and consequently are not compatible with any mainstream video compression standard.

Tillo et al. [178] proposed a multiple description coding scheme which exploits the redundant slice coding option of H.264. H.264 allows both primary and redundant slices to be encoded within the video stream. In case the primary slice cannot be decoded or is highly affected by transmission errors, the redundant slices can be decoded to compensate the lost data. The proposed MDC scheme forms descriptions by interlacing primary and redundant slices. The two streams obtained are perfectly compliant with H.264; any of them can reconstruct the video content with only drift distortion generated by the redundant slices.

3.5 Chapter Summary

This section discussed the requirements and presented the main components involved in mobility management and traffic distribution over multiple simultaneous connections. Two main components of a mobility management system can be identified which are **location management** and **handover management**. The handover management module is directly involved in preserving data session and maintaining QoS levels. The required QoS, in the context of multimedia applications, represents the condition which need to be met by the quality parameters of the transport network including throughput, packet loss, packet delay and delay jitter. User QoE is also and important factor and in the context of multimedia applications is highly dependent of the quality of multimedia content delivered to the user.

There are three important operations in terms of maintaining mobile device's connectivity and guaranteeing high QoS levels: **network monitoring**, **network selection** and **handover execution**. Each of these operations has a great impact on the overall QoS lever received by the user. Consequently, each of them has to be carefully designed and parameterized in order to efficiently interact with each other toward the main goal of achieving maximum user perceived quality and satisfaction.

Although there is much arguing on the protocol stack layer where mobility should be accommodated, by choosing session and application layers, most benefits can be obtained. Application layer mobility solutions benefit from the well defined and mature features of transport protocols including congestion control, reliability and security, involve less interference with the underling infrastructure and are IP compatible. Application layer mobility solutions also benefit of a better interaction with the application and the user. Deploying mobility at application layer has the additional benefit of cost-effectiveness. This is mainly determined by the reduced interference with already largely deployed entities, including transport layer and network layer protocols.

The most popular handover solutions developed at network, transport and session layer are Mobile IP, Mobile DCCP and Mobile SIP respectively. These protocols represent a good comparison base to evaluate the performance of application layer handover solutions and outline their benefits in comparison with mobility solutions developed at other layers of the protocol stack.

Network monitoring and **handover decision** making are also extremely important various solutions for these being proposed in the literature. The emergent IEEE 802.21 Media Independent Handover is a useful tool in network information gathering and management as well as handover decision making.

Handover decision making solutions proposed in the literature include cost function-based, context aware, Fuzzy Logic and Neural Networks-based as well as user-centric solutions.

Multi-path or multi-connection data transfer has been already used for throughput aggregation and load balancing in wired and wireless networks. Various solutions have been proposed in the literature for rate allocation and traffic splitting over multiple connections.

However to the best of our knowledge a multiple simultaneous connections approach has not been used in mobility management solutions and consequently the existing solutions are not fit for handover management in heterogeneous wireless environments.

Chapter 4

Quality of Multimedia Streaming Metric

4.1 Introduction

Estimating the capacity of a network to efficiently transport multimedia content or any other data traffic is extremely important for delivering high quality services to the user. This chapter introduces the Quality of Multimedia Streaming (QMS) metric, a novel capacity estimation mechanism for wired and wireless networks.

QMS estimates how much traffic can be transported over a certain network in order to efficiently deliver high quality multimedia content to mobile devices.

QMS combines in an innovative way both a network selection strategy and a traffic allocation model. An additional novelty introduced by QMS is the inclusion of a QoE parameter in the decision making process. QoE is estimated from the video quality assessment of the delivered content which is performed by making use of stream-based video quality metrics.

The proposed SASHA performs handover by distributing and balancing the load between multiple networks simultaneously. The existing network selection strategies for handover in heterogeneous wireless environments are not fit for this approach to mobility because they are developed to choose a new network and not perform intelligent allocation of traffic data rates to multiple networks. At the same time, the rate allocation methods designed for load balancing or traffic distribution are not developed with the heterogeneous wireless environment in mind and consequently they do not consider the various parameters required for an efficient traffic distribution in such an environment.

QMS combines the two aspects, network selection and rate allocation, in a cost function-based approach to handover management through load balancing. QMS incorporates various parameters which are relevant for network selection in a heterogeneous wireless environment. This set of parameters include: QoS, QoE, monetary cost, energy consumption and user preferences. The QoS parameter is a cost function in itself which incorporates throughput, packet loss, delay and delay jitter as QoS related components. QoS component illustrates the characteristics of each from the point of view of network QoS. QoE component represents the quality of the video content delivered to the user and estimates his perceived QoE. Monetary cost parameter denotes the charges applied for transferring a certain amount of data of the target connection (network). Energy, similar with monetary cost component, represents the energy consumed by the transceiver to receive or transmit the data over a certain network.

This QMS metric is highly flexible. Each parameter is weighted. Weight normalization is applied for consistency.

The general QMS function and each of the QMS main parameters will be presented and discussed in the next sections.

4.2 QMS Architecture

The QMS architecture consists of two main modules: the server-side module and the client-side component. This is schematically illustrated in Figure 4.1.

The client-side module is the most complex and consists of four submodules. QoS Monitor submodule is in charge with measuring QoS related parameters such as received throughput, packet loss, packet delay and delay jitter.

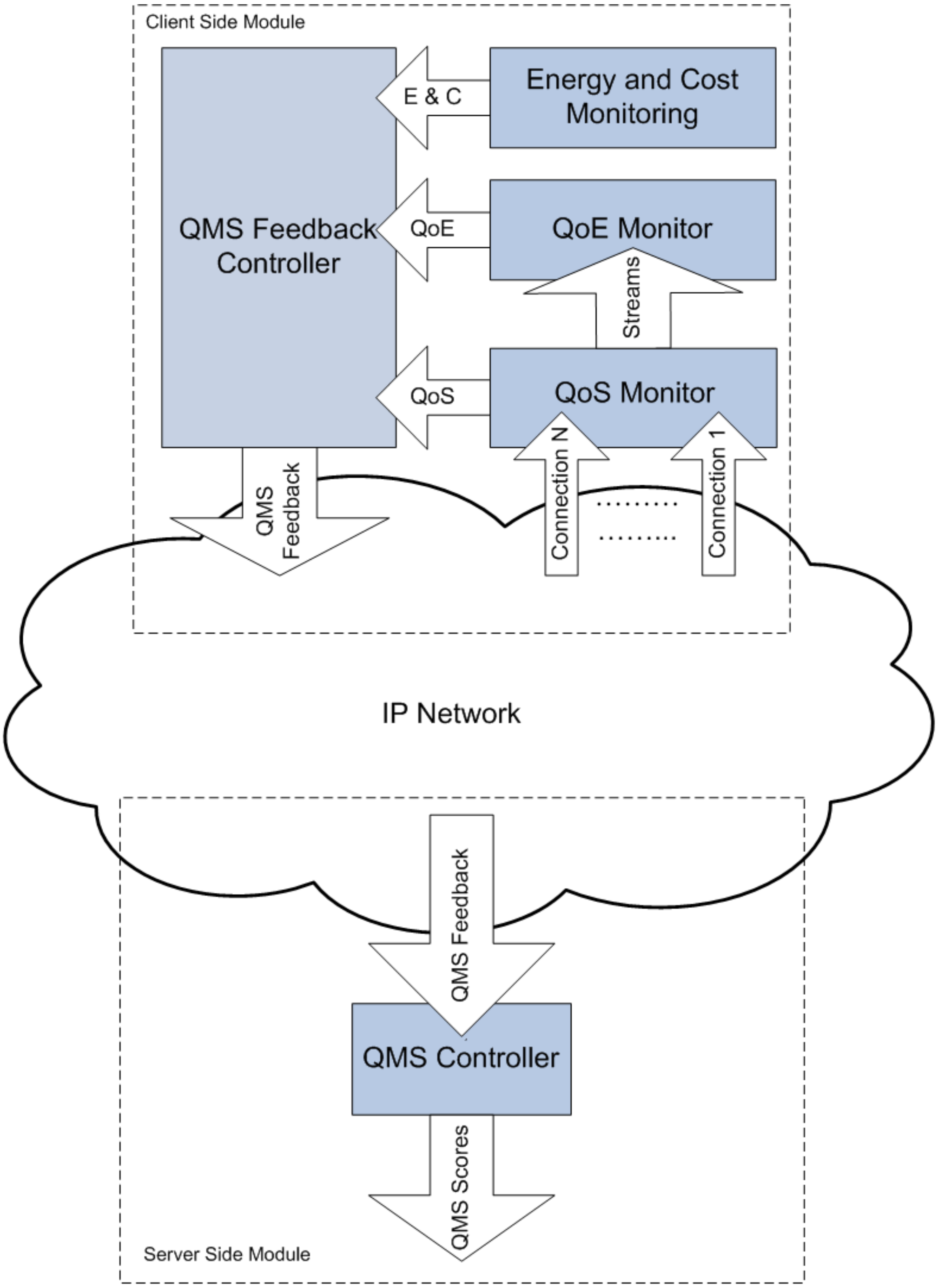

Figure 4.1: QMS Architecture

QoE Monitor is responsible for estimating user QoE based on video quality assessment for each connection separately. This module implements a non-reference stream-based video quality assessment metric.

Energy and Cost Module harvests information regarding the monetary cost of transferring data over each of the existing networks separately and the energy consumed by networking interface corresponding to each network to transmit and receive a unit of data (e.g 1Mb).

QMS Feedback Controller periodically gathers data from the three submodules introduced above and sends the corresponding feedback to the server-side QMS module.

The server-side module has only one component and this is the QMS Controller. QMS Controller periodically receives feedback from the client-side QMS module and based on it, it computes the QMS scores for each connection individually.

4.3 QMS Function Parameters

4.3.1 QMS General Function

The Quality of Multimedia Streaming (QMS) metric can be used to evaluate the capacity of a communication link to transport a certain amount of multimedia traffic. As network related parameters (i.e. QoS, Cost) and device related parameters (i.e. Power consumption) impact differently on user satisfaction depending on the application characteristics, user preferences and other aspects of user interaction with the system, the QMS metric is designed for maximum flexibility and efficiency.

The QMS metric is described by the function from Equation 4.1 and is dependent on the characteristics of the communication channel i.

$$\begin{aligned} QMS^i = w_{QoS} * QoS^i_{grade} + w_{QoE} * QoE^i_{grade} \\ + w_C * Cost^i_{grade} + w_P * PEff^i_{grade} \end{aligned} \tag{4.1}$$

The components, expressed on a 100 point scale from 0 to 100, with maximum grade 100, include QoS parameters, QoE component, monetary cost and energy consumption.

For maximum efficiency and flexibility weights are associated with each component. These weights are set based on user preferences and application requirements. Weight normalization is required, so the condition from Equation 4.2 has to be respected.

$$w_{QoS} + w_{QoE} + w_C + w_P = 1 \tag{4.2}$$

4.3.2 QoS Grade Component

$QoS_i grade$ represents the grade which assesses the network QoS for the communication channel i and is described by the formula from Equation 4.3.

$$\begin{aligned} QoS^i_{grade} = w_T * Throughput^i_{grade} + w_L * Loss^i_{grade} \\ + w_D * Delay^i_{grade} + w_J * Jitter^i_{grade} \end{aligned} \tag{4.3}$$

QoS component is the most important parameter of the QMS metric as QoS has a great impact on the quality of the delivered content and consequently on the user perceived quality. The QoS component includes four different parameters: received throughput, packet loss, packet delay and delay jitter.

The components of the $QoS_i grade$ are also weighted to offer maximum flexibility. This is done in order to meet the various requirements and sensitivities of multimedia encoding and transport schemes which may be used.

For example if an error concealment technique is used in the encoding and streaming process then packet loss becomes a less important parameter so its weight can be reduced in the favor of other parameters.

For accurate results weight normalization is also required, so the condition from Equation 4.4 needs to be respected.

$$w_T + w_L + w_D + w_J = 1 \tag{4.4}$$

The network QoS components are also expressed on a 100 point scale, from 0 to 100, with maximum grade of 100. This grade represents the estimated percentage of the whole application traffic which can be transported by the network for which the QoS component is computed.

The components of $QoS_i grade$ are computed by the server-side module using the statistical information collected by the client-side module and are periodically reported to the server. The client also provides the server with information related to the application requirements in terms of target streaming bitrate and sensitivity to network QoS parameters.

The client module monitors the throughput, loss, delay and jitter for each communication channel i and sends reports to the server module every t seconds. Consider the $[t_n, t_{n+1}]$ time interval of length t, we analyze the QMS components for each communication channel i in each time interval t.

Throughput component is computed using the formula presented in equation 4.5.

$$Throughput^i_{grade} = \frac{MaxGrade * Throughput^i_t}{SRate} \tag{4.5}$$

$Throughput^i_t$ represents the throughput received by the communication channel i in the time interval t as reported by the client-side measured in Mbps. $SRate$ is the average total data rate of the application considered in the same time interval and is expressed in Mbps. $MaxGrade$ represents the maximum grade which is 100.

The dependency of $Throughput^i_{grade}$ on the throughput of network i, as it can be seen in Figure 5.3 is linear, which in this context generates a conservative behavior. The maximum throughput of 5Mbps presented in the figure is for illustrative purposes only and does not represent a limit in itself.

$Throughput^i_{grade}$ component aims at keeping network i load at its current level. This kind of behavior avoids network overloading and consequent packet loss, but at the same time tends to leave bandwidth resources unused.

Throughput is a very important parameter; it has been used in many handover decision making strategies and load balancing rate allocation techniques. This parameter shows the load level of each communication channel separately.

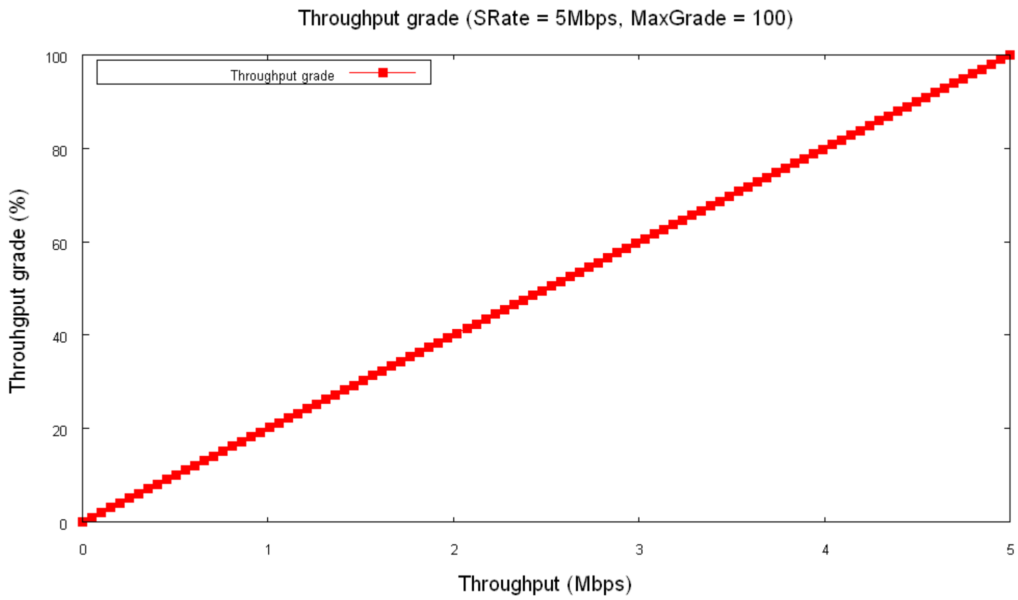

Figure 4.2: Throughput grade

Loss component is described by the functions from Equation 4.6 and Equation 4.7.

$$Loss^i_{grade} = \frac{MaxGrade}{1 + \left(\frac{LossRate^i_t}{MaxLossRate}\right)^{Q_L}} \tag{4.6}$$

$$LossRate^i_t = \frac{Loss^i_t}{SRate * RateShare^i} \tag{4.7}$$

$Loss^i_t$ represents the average loss recorded by the client on communication channel i on the time interval t and is expressed in Mbps.

$SRate$ is the application's current streaming rate (measured in Mbps), $RateShare^i$ represents the percentage of the $SRate$ that is currently allocated to communication channel i. $LossRate^i_t$ represents the loss as a fraction of the total data rate transported by channel i. $MaxLossRate$ represents the maximum allowed loss expressed as a fraction of the streaming rate. Q_L is a quality factor and is set by the application to adjust the reaction to loss.

As it can be seen in Figure 4.3 the loss component has an optimistic behavior. It tends

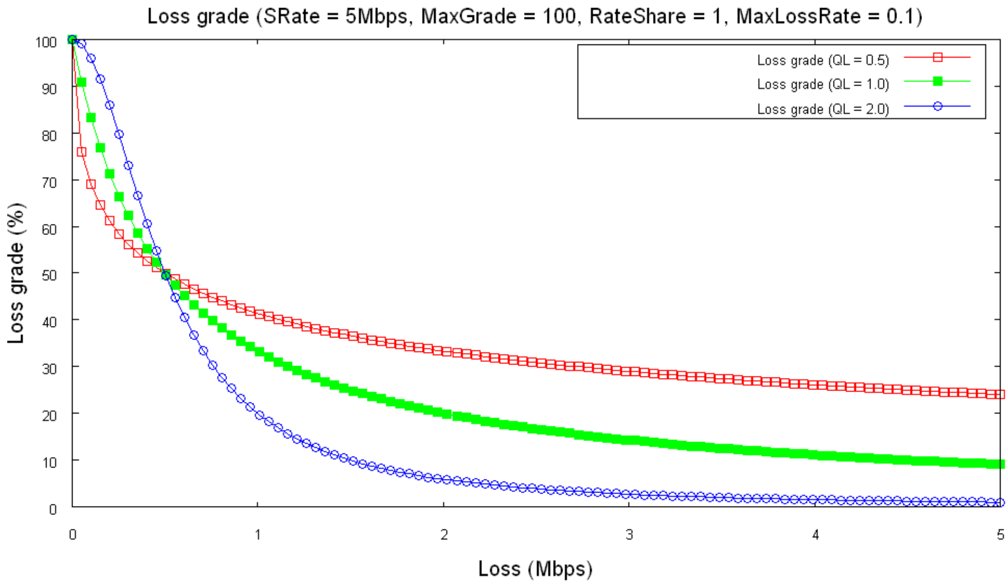

Figure 4.3: Loss grade

to load the current network especially when the loss rate is within the accepted boundaries, but also when high loss rates are recorded.

The default value for Q_L is set to 1, this being the neutral value. However, as it can be observed in Figure 4.3, Q_L values greater than 1 (e.g. 2) determine the loss component to be less optimistic for loss rates greater than $MaxLossRate$ while Q_L values less than 1 (e.g. 0.5) increases the optimism of loss component for the same loss level.

To exemplify for a loss rate of 0.2 and maximum loss rate of 0.1 the loss grade will suggest a 33% traffic load when Q_L is set to 1, 41% when Q_L is 0.5 and 20% for Q_L equal to 2.

When the loss rate is 0.05, below the maximum rate of 0.1, the loss grade indicates a 67% traffic load when Q_L is set to 1, while Q_L set to 0.5 determines the load rate to decrease to 58%. When Q_L is set to 2, the same circumstances, loss grade suggests an 80% load rate.

Delay and **Jitter** components are described by the functions in Equation 4.8 and Equation 4.9.

$$Delay^{i}_{grade} = \frac{MaxGrade}{1 + \left(\frac{Delay^{i}_{t}}{DThreshold}\right)^{Q_D}} \tag{4.8}$$

$$Jitter^{i}_{grade} = \frac{MaxGrade}{1 + \left(\frac{Jitter^{i}_{t}}{JThreshold}\right)^{Q_J}} \tag{4.9}$$

$Delay^{i}_{t}$ and $Jitter^{i}_{t}$ represent the average delay and jitter measured by the client on communication channel i during the time interval t. $DThreshold$ and $JThreshold$ are thresholds specified by the application and represent the maximum delay and jitter accepted while still preserving a minimum multimedia quality. The two thresholds depend on the playback time and buffer size and are dictated by the application.

Q_D and Q_J represent quality factors which denote application's sensitivity to delay and jitter respectively. The two parameters affect the behavior of the Delay and Jitter components in the same manner Q_L affects Loss component. The default neutral value is 1.

The dependencies between the two grades and the values of delay and jitter recorded on network i are similar with the dependency between loss grade and packet loss.

4.3.3 Energy Component

The energy component of QMS aims at controlling and reducing the overall amount of energy consumed by the mobile device engaged in wireless network communication.

$\mathbf{PEff^{i}_{grade}}$ represents the energy efficiency score of communication on channel i with respect to the mobile device power usage. This parameter is detailed in Equation 4.10.

$$PEff^{i}_{grade} = \frac{MaxGrade}{e^{PMb*P}} \tag{4.10}$$

PMb is the power consumed by the transceiver to receive 1 Mb of data. P is a power efficiency factor and denotes the application's sensitivity to power consumption.

The dependency between power consumption per Mb of data and the energy grade is presented in Figure 4.4

An exponential relationship between the power efficiency grade and the power con-

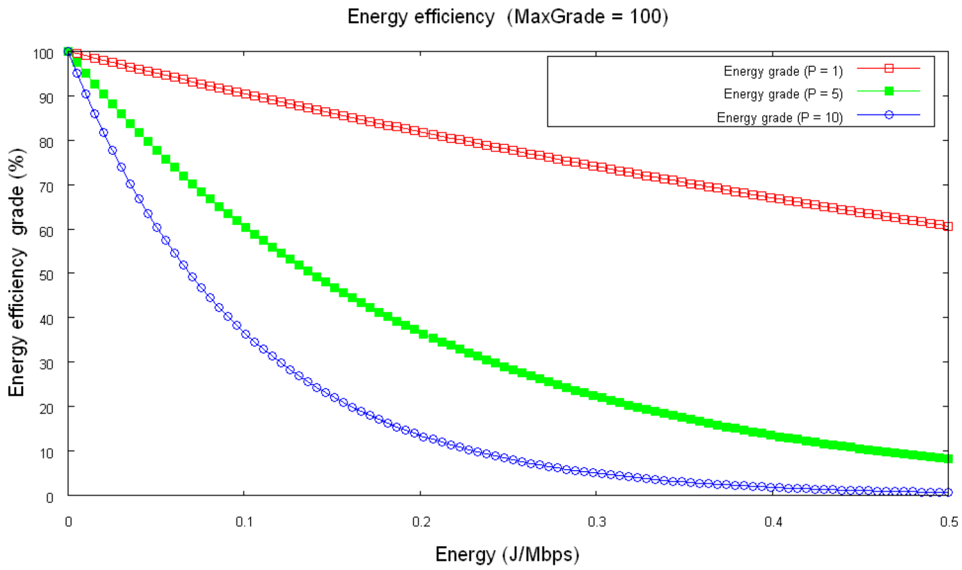

Figure 4.4: Energy grade

sumption was chosen based on the fact that battery lifetime decreases exponentially with the increase in load [179]. P is set in such a way that the dependency between load and battery lifetime is followed. This dependency can be obtained through simulations [179] or experiments and does not represent the subject of this theses.

The use of energy grade is justified by the fact that in a heterogeneous wireless environment different networks may have different energy consumption characteristics.

4.3.4 Monetary Cost Component

$\mathbf{Cost^{i}_{grade}}$ is a monetary cost related component and is computed based on the user cost-utility rating of the provided service.

This component is described in Equation 4.11 and Equation 4.12 with $MaxC$ representing the maximum cost that the user is willing to pay for viewing the specified multimedia content and C the total cost of streaming multimedia content over channel i at the application's current streaming rate, $SRate$. CMb represents the monetary cost of transferring one Mb of data over channel i. $EstPlayTime$ is basically the length of the video clip or an estimated play time for the streamed content expressed in seconds. $MaxGrade$ represents the maximum grade.

$$Cost_{grade}^{i} = \begin{cases} MaxGrade - \frac{MaxGrade}{MaxC} * C & ,C \leq MaxC \\ 0 & ,C > MaxC \end{cases} \tag{4.11}$$

$$C = SRate * CMb * EstPlayTime \tag{4.12}$$

Monetary cost is another parameter which may differ from one network to another. In order to allow control over the cost of a multimedia session this parameter has to be included in the rate allocation process.

4.3.5 QoE Component

$\mathbf{QoE^{i}_{grade}}$ represents user perceived quality measurement and is computed based on the received content's video quality, assessed using objective video quality metrics. Throughout this theses the QoE will be estimated by video quality of the multimedia content delivered to the mobile user.

In order to reduce the computational complexity of this process, QoE is estimated using stream-based metrics when multiple description coding is used for traffic distribution.

Such a metric is the PSNR metric proposed by Lee in [180]. The formula used is presented in Equation 4.13 where $MAX\_Bitrate$ represents the average bitrate of the multimedia stream after the encoding process, EXP_Thr is the average throughput expected when delivering the multimedia stream over the network and CRT_Thr is the actual throughput measured during delivery.

$$PSNR = 20 \cdot log_{10}\left(\frac{MAX\_Bitrate}{\sqrt{(EXP\_Thr - CRT\_Thr)^2}}\right) \tag{4.13}$$

If packet-based traffic distribution is used then QoE grade for each communication channel is determined by distributing the overall QoE according to the channel's corresponding rate share. The function described in Equation 4.14 outlines the procedure of assessing the contribution of each of the currently used communication channels to the overall QoE grade $QoE_{overall}$.

$$QoE^i = QoE_{overall} * RateShare^i * LossRate^i \tag{4.14}$$

$SRate$ is the application's current streaming rate and $RateShare^i$ represents the percentage of the $SRate$ that is currently allocated to communication channel i and $LoassRate^i$ represents the packet loss on channel i expressed as percentage. Equation 4.14 is based on the assumption that the more traffic a connection transports the more it will impact on the overall video quality and consequently the user perceived quality (QoE) at the same loss rate.

Consequently, the QoE^i_{grade} formula is presented in Equation 4.15; the dependency on the estimated video quality score, (QoE_i) is linear. $MaxQoE$ is the maximum score that can be achieved with the video quality metric used. $MaxGrade$ represents the maximum grade which is 100.

$$QoE^i_{grade} = \frac{QoE_i}{MaxQoE} * MaxGrade \tag{4.15}$$

4.4 Weights Settings and User Profiles

The weights presented in Equation 4.2 and Equation 4.4 give QMS its flexibility and have a great impact on the performance of the application or mechanism which uses it, in this case SASHA. In order for the user of this metric to benefit from its potential the weights need to be set according to various needs.

However extensive testing of the impact of weight values on the performance and the outcome of the application is necessary for best performance. For SASHA deployment, the testing has been performed in Chapter 6 and default values have been proposed.

In some situations, as it is the case of QoS, energy and monetary cost, default values cannot be proposed. This is because the application's outcome in terms of QoS, energy and monetary cost is inverse proportional. For this purpose, development of user profiles and classes of devices are required. The weights will be then set in order to reach the goals of each user profile or device class.

To exemplify, consider the weighting of the energy component and the QoS component. Three categories of users can be determined.

One category groups the users which are more concerned on the quality of the received multimedia content and consequently the QoS than the battery life. For these users the weights will be set in such a way that maximum importance is given to the QoS scores rather than the energy grades.

The second category groups the users who have a balanced approach. For them the weights are set in such a way that a fair tradeoff between energy consumption and QoS levels is achieved.

The third group consists of users which are mostly concerned with energy saving and consequently the battery life. For these users the weights are set in the favor of the energy component and not on the QoS grades.

User profiling and device classes do not make the subject of this work, consequently it will not be further discussed.

Details regarding the dependency between QMS weights setting and the behavior of a QMS-based handover mechanism are presented in Chapter 6.

4.5 Chapter Summary

This chapter introduced the Quality of Multimedia Streaming (QMS) metric, a novel general purpose mechanism for estimating the capacity of a network to efficiently carry multimedia traffic.

QMS is a flexible and comprehensive metric incorporating several parameters such as QoS, QoE, monetary cost and energy consumption. For maximum flexibility each separate parameter has a weights which can be adjusted in order to modify the impact of that particular parameter on the overall behavior of QMS.

QMS's architecture with its two components, the server-side module and the client-side module, is introduced followed by the mathematical model of QMS. Each of QMS component is separately introduced and detailed discussion and analysis is presented.

Finally QMS tuning, based on setting particular weight values, was discussed in the context of user profiles and device classes.

Chapter 5

Smooth Adaptive Soft-Handover Algorithm

5.1 Introduction

In the currently deployed wireless environment, a mobile node passes through the coverage area of various wireless networks, a diversity of wireless communication technologies being used. In order to maintain a high level of network quality of service (QoS), and consequently high user perceived quality of service, in the context of a highly dynamic heterogeneous wireless environment a quality-oriented mobility management system has to be employed.

As discussed in the previous sections handover management part of the mobility management system has the greatest impact on the user perceived QoS and consequently user quality of experience (QoE). Existing solutions have several limitations which make them less suitable in such an environment.

Consequently in this section, the Smooth Adaptive Soft-Handover Algorithm (SASHA), a novel handover management solution, is proposed to overcome the identified issues of the existing handover management techniques.

5.2 Heterogeneous Wireless Environment

Many of the wireless communication technologies developed over the last years are already in their deployment phase, others are fast approaching that stage. This determines a high heterogeneity of the wireless environment which now provides multiple alternative communication resources for mobile devices equipped with the corresponding network interfaces.

Although this may lead to the conclusion that mobile devices will be receiving high bandwidth data communication services, the QoS may not be constantly maintained at high levels leading to a decrease in user perceived quality. This is mainly determined by the dynamics of the wireless network resources in a heterogeneous wireless environment.

A heterogeneous wireless environment consists of multiple wireless networks, using a set of different wireless communication technologies, which cover a certain geographical area in a cellular overlapping manner.

For a more detailed illustration of the issue, Figure 5.1 schematically presents a wireless heterogeneous environment which covers a certain geographical area. Various mobile users, with various communication requirements and device capabilities access the Internet through this wireless environment.

The characteristics of each wireless technology employed in such a heterogeneous environment (i.e. resilience of the wireless link to environmental interferences, coverage and bandwidth) determine a certain level of dynamics of the QoS level provided in a particular location at a certain time. User mobility is further increasing network dynamics as the QoS level received by the mobile device depends on the distance to the access point or base station. Network load, both in terms of the number of mobile stations serviced and the amount of traffic has an additional negative impact on the received QoS level.

In this context it is hard to guarantee constantly a high level of QoS to a mobile device roaming through a geographical area covered by multiple wireless networks. The main issues impacting quality provisioning are increased load, increased interference, poor signal strength due to environmental factors or user mobility. In particular problems occur when moving from the coverage area of one network to another.

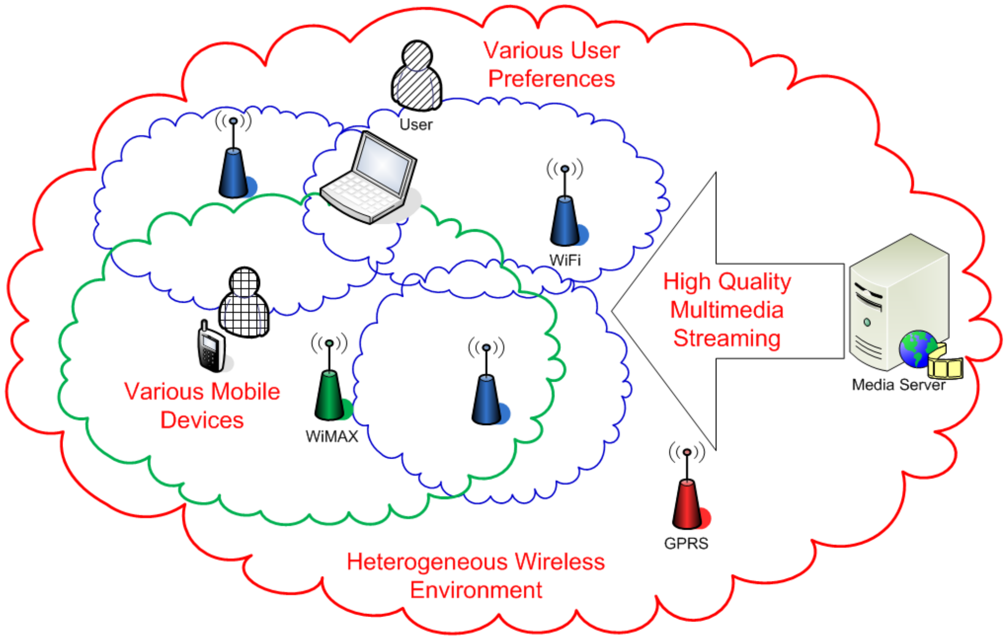

Figure 5.1: Heterogeneous wireless environment

An simple example including two networks is presented in Figure 5.2 and Figure 5.3. Assume that a mobile device is receiving and average 1.5Mbps of multimedia traffic over wireless links. The mobile node travels through the coverage area of two distinct wireless networks as depicted in Figure 5.2.

An illustration of the possible throughput received from each network by a mobile device passing through the coverage areas of the two distinct networks is presented in Figure 5.3.

It can be seen that within the overlapping area of the two networks the throughput drops to 0.9 Mbps for both networks due to the distance between the mobile host and the access points or base stations. In such a situation, none of the two networks is capable of providing the required throughput,(in this case is 1.5 Mbps) leading to an inevitable drop in user perceived quality if no action is taken.

The handover management solutions proposed in the literature, when the current network cannot satisfy communication needs of the mobile device, usually transfer the entire data flow to a new network, in this case Network 2 from Figure 5.2. This approach to mobility is not capable to cope with situations such as the one presented in Figure 5.2 due to

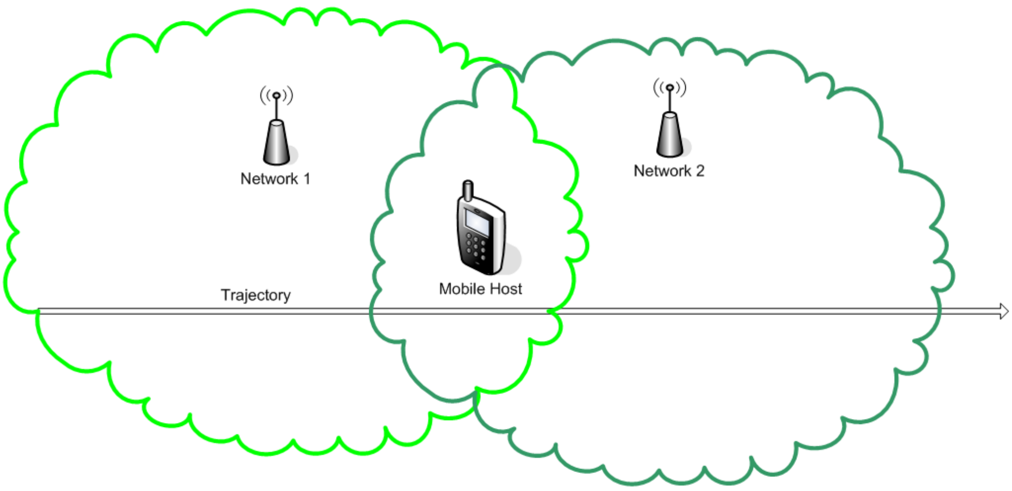

Figure 5.2: Network Coverage

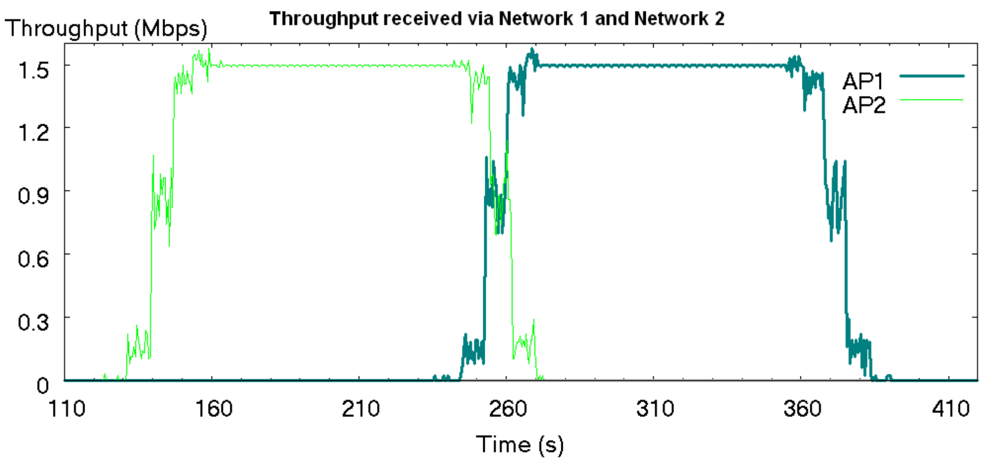

Figure 5.3: Received throughput

the impossibility to choose a network which can provide the entire required bandwidth by itself. This represents a major drawback of the existing handover management solutions.

The same situation may be encountered when all the available networks are highly congested and none is capable of supporting the required level of QoS. Additionally many handover management solutions proposed in the literature do not have any quality oriented approach to handover. The level of QoS and more important the user perceived quality are not considered in the decision making process and in the handover operation process.

In this context there is a need for a quality-oriented handover management scheme

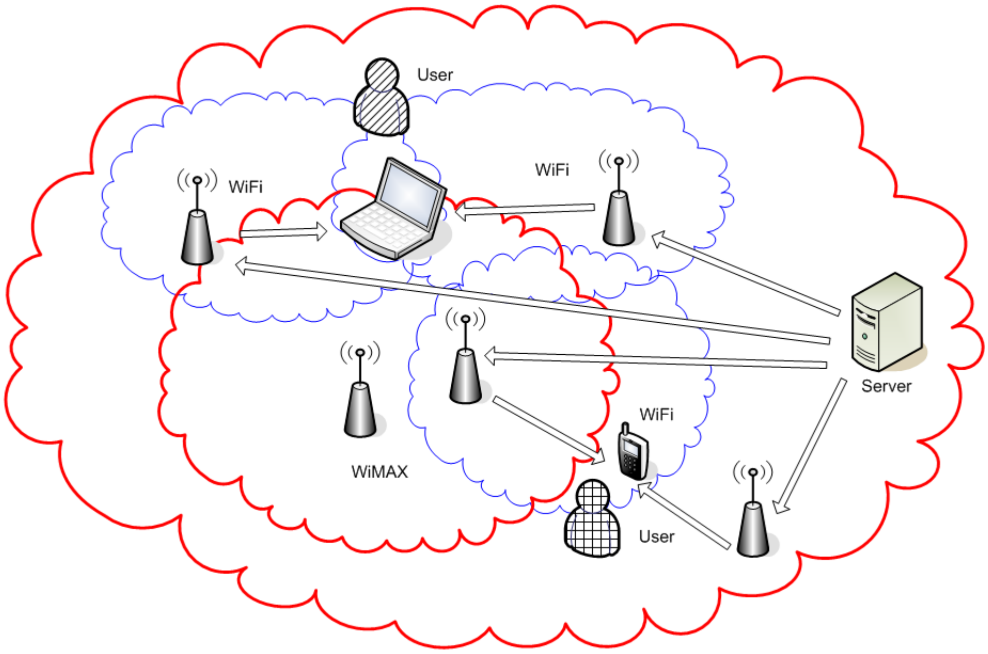

Figure 5.4: SASHA principle

which efficiently exploits all the available wireless resources and offers high levels of user perceived quality.

5.3 SASHA Overview

To overcome these drawbacks of the current handover management schemes, a novel Smooth Adaptive Soft-Handover Algorithm (SASHA) is introduced to exploit all available network resources simultaneously. The main application traffic is split and distributed over multiple simultaneous connections. Host mobility is allowed by balancing the load from one network to another. If the signal from one network is fading due to distance and consequently the throughput is decreasing, a new network may be available and the traffic that cannot be conveyed by the first network can be transmitted over the second one.

SASHA's basic principle is presented in Figure 5.4 where several mobile hosts are located within a heterogeneous wireless network environment. Each mobile device is capable of communication with multiple networks simultaneously and require transmission of multimedia traffic. The multimedia traffic is split and distributed to the mobile devices over a subset of available networks.

The amount of traffic transported by each network is variable and depends on the efficient capacity of each network. The efficient transport capacity is estimated based on the available bandwidth and other factors including monetary cost, energy consumption and user QoE.

The architecture of the proposed system and the handover algorithm will be detailed in the following subsections.

5.4 SASHA Architecture

The heterogeneity of the wireless network environment provides mobile users access to the Internet through various wireless networks. Considering the various characteristics of the current wireless technologies with all their advantages and disadvantages as well as the dynamic nature of network QoS due to factors such as load and environmental influence, mobile users should employ an intelligent mobility management mechanism which exploits all the available network resources in order to maintain high QoS levels and consequently very good user perceived quality of experience.

The Smooth Adaptive Soft-Handover Algorithm (SASHA) was proposed as a handover management solution for multimedia applications. SASHA aims at preserving user perceived quality through intelligent network resource management and traffic distribution. SASHA is an application layer mechanism which provides multimedia applications with multimedia content delivery support over heterogeneous wireless networks.

SASHA distributes the application traffic over multiple simultaneous connections, each connection using a different wireless network for the actual data exchange. Traffic distribution is made based on each network's capacity to deliver multimedia content at high quality. Mobility is achieved by **smoothly transferring the traffic** from old fading wireless links to new ones.

Unlike other handover enhancement solutions which duplicate the traffic over both new and old network, SASHA **splits the main traffic and distributes the sub-streams over**

multiple parallel connections.

For efficient traffic distribution and smooth handover management each active connection has to be monitored and its capacity to deliver high quality multimedia traffic has to be estimated. Therefore the newly introduced Quality of Multimedia Streaming (QMS) metric was employed, which was discussed in details in the previous chapter. QMS has the role to estimate the amount of traffic each existing communication channel can efficiently carry. Based on the QMS scores SASHA computes the data rates for each channel separately.

Experimental simulation-based results show that by exploiting all the communication resources available, better QoS and consequently higher user perceived quality can be provided in comparison with the case when only one network is used to carry the whole multimedia traffic. Moreover performing a smooth handover through gradual traffic balancing presents a less disruptive impact and less negative effect on multimedia quality than the solutions which abruptly switch the traffic from one network to another.

Figure 5.5 presents the SASHA architecture illustrated based on protocol placement on the TCP/IP layered model. SASHA resides at the application layer and provides the multimedia application with a middleware framework for mobile data transport. SASHA opens multiple simultaneous data sessions (connections) over each active network.

Any transport protocol can be used; however for multimedia applications UDP or DCCP protocols are considered because of their best suitability for this type of applications. The application traffic is distributed and transported by these connections. The traffic is not duplicated; it is split, transported over the multiple connections and then aggregated at the client side.

In order to efficiently distribute the application traffic over the existing wireless networks, various information is required, including network conditions, monetary cost, energy consumption, etc. These information is harvested by the mobile device by monitoring the existing networks and the data sessions or by various network entities or agents which provide the required information (e.g. IEEE 802.21).

A more detailed block-based architecture of SASHA is presented in Figure 5.6. SASHA comprises two components: a server side module, which uses SASHA to stream real-time

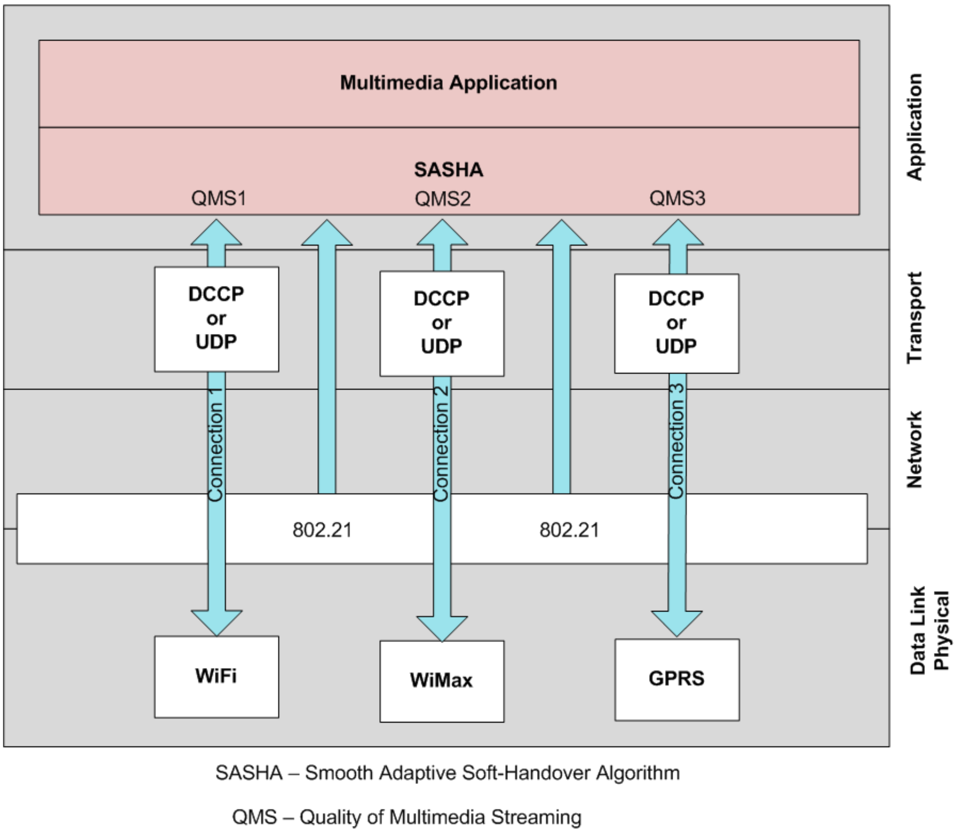

Figure 5.5: SASHA overview architecture

multimedia content over wireless networks and a client side module which attaches to the multimedia client application and receives the streamed content.

SASHA server-side component is composed of three sub-modules: SASHA Controller, SASHA Connection Manager and SASHA Traffic Splitter and Allocator.

SASHA Controller receives feedback information, containing statistical data regarding received throughput, packet loss, packet delay and delay jitter as well as energy consumption and monetary cost of data transmission from the mobile host monitoring modules and is responsible with QMS score computation and data rate allocation over the available connections is performed accordingly. Based on the QMS scores computed for each connection separately the SASHA Controller estimates the amount of application traffic which can be transported by each connection.

Figure 5.7 presents the data structure used for describing the feedback information. It

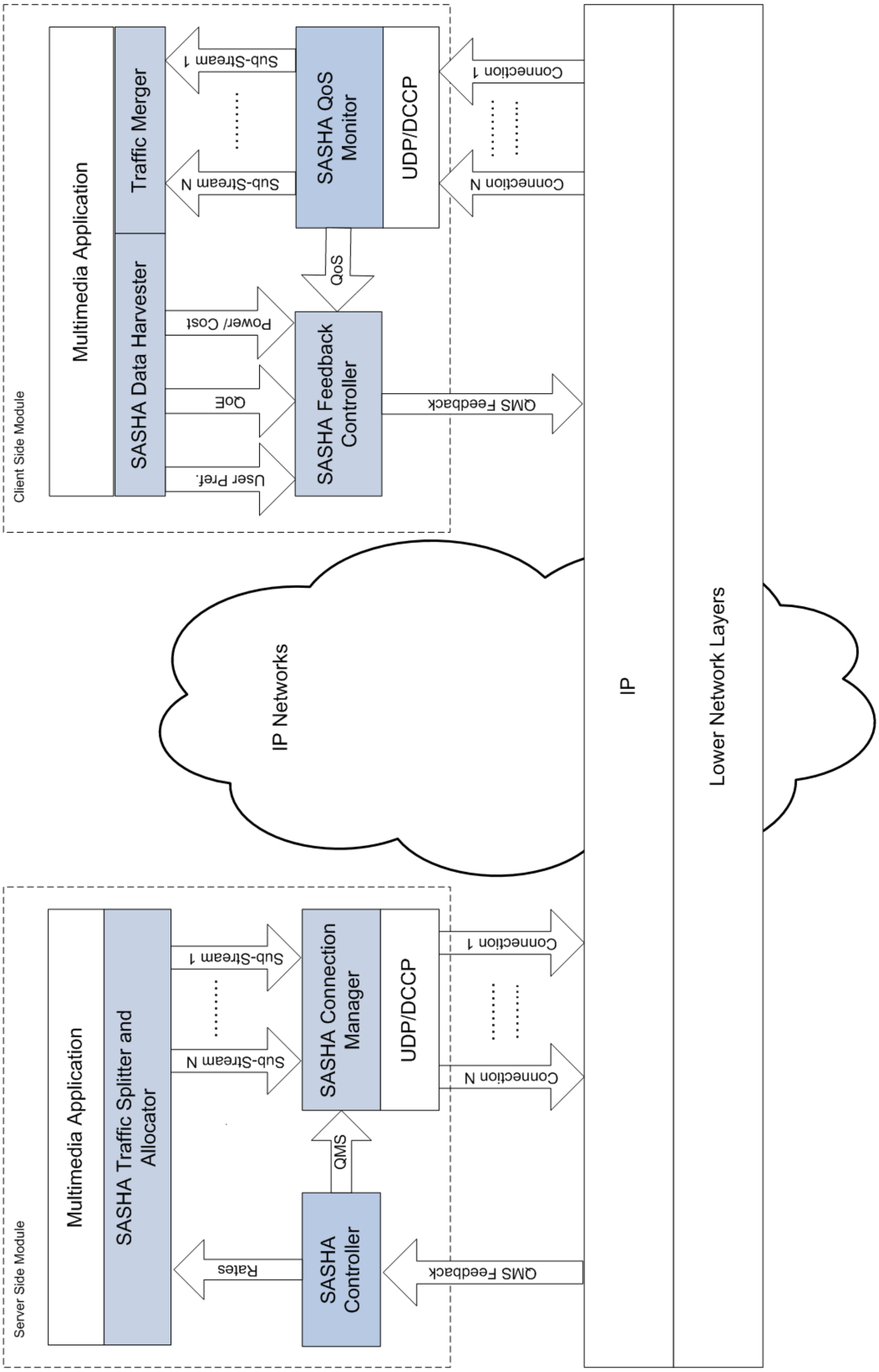

Figure 5.6: SASHA architecture

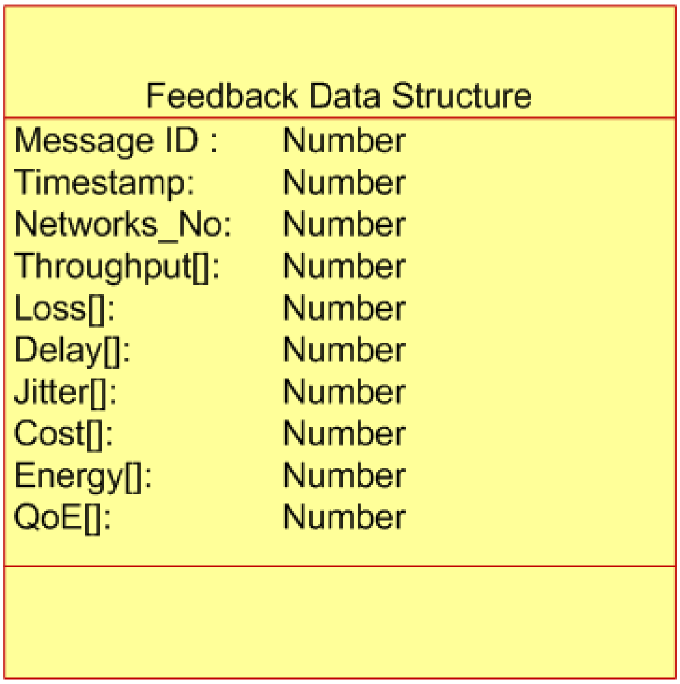

Figure 5.7: Feedback Data Structure

contains all the data required for computing QMS for each communication channel separately.

The estimated connection data rates are used by the **SASHA Traffic Splitter and Allocator** sub-module to split the main data traffic into several sub-streams, each stream having the bit rate corresponding to the transport capacity of its allocated connection as computed by the SASHA Controller.

SASHA Connection Manager is responsible for maintaining the connection pool from which the active connections are chosen according to their QMS scores and the traffic requirements. The connection manager accepts the incoming new connection initiated by the client side module which detected a new available network. The inactive connections are sampled by the manager using a low bitrate sampling traffic to determine their availability. The inactive connections which are not capable of delivering at least the low bitrate sampling traffic are considered to be dead and are closed by the manager.

Figure 5.8 presents the sequence diagram involved in adding a new connection to the list of connections used by SASHA.

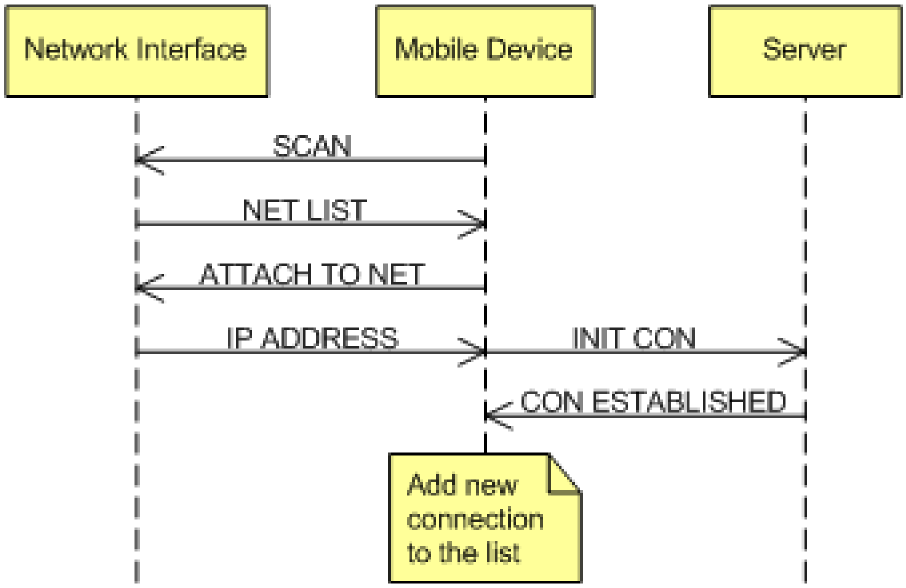

Figure 5.8: Add a new connection to the pool

The above mentioned low bitrate sampling rate is used to test availability of an inactive connection. Inactive connections are the ones which were not allocated any application data traffic. SASHA Connection Manager periodically sends dummy messages (gathering a data rate of no more than 1kbps) over these connections. If these messages reach the client, they will appear as a nonzero throughput for their corresponding connection. This will tell the SASHA Connection Manager that the particular connection is still alive. If the reported throughput is 0 for a predefined period of time (10 seconds) the connection is considered to be dead and consequently is removed from the connection pool.

SASHA client-side component comprises four sub-modules: SASHA QoS Monitor, SASHA Data Harvester and SASHA Feedback Controller.

The **SASHA QoS Monitor** is responsible for monitoring the QoS parameters on each connection separately. SASHA QoS monitor intercepts each packet received on each separate connection. Each received packet is counted and it contributes to the average throughput value computed and stored by the SASHA QoS Monitor. Based on the packet timestamp and sequence number, packet loss, packet delay and delay jitter is calculated. These values also contribute to the average values stored by the monitoring component. The average values of the QoS parameter mentioned above are periodically reported to the SASHA Feedback Controller. After the report is delivered all the counters and average values are reseted and the monitoring continues for another predefined time interval of length t ($[t_n, t_{n+1}]$).

Algorithm 1 presents in pseudo code the mechanism for monitoring the QoS parameters at the client-side module.

Algorithm 1 QoS Monitoring Algorithm

Procedure:

$Time\_interval \Leftarrow 1s$
$Current\_time \Leftarrow Get\_System\_Time()$
$Next\_report_time \Leftarrow Current\_time + Time\_interval$
for all $Message = Receive\_Message()$ **do**
  if $Message.Type == Data\_Message\_Type$ **then**
    $Received\_Message\_Count = Received\_Message\_Count + 1$
    $Total\_Throughput = Total\_Throughput + Message.Size$
    $Total\_Loss = Total\_Loss + (Message.Seq - Last\_Sequence - 1)$
    $Total\_Delay = Total\_Delay + (Current\_time - Message.Timestamp)$
    $Total\_Jitter = Total\_Jitter + absolute((Current\_time - Message.Timestamp) - Last\_Delay)$
    if $Current\_time >= Next\_report_time$ **then**
      $AVG_Throughput = Total\_Throughput\ Received\_Message\_Count$
      $AVG_Loss = Total\_Loss\ Received\_Message\_Count$
      $AVG_Delay = Total\_Delay\ Received\_Message\_Count$
      $AVG_Jitter = Total\_Jitter\ Received\_Message\_Count$
      $Send\_Report(AVG\_Throughput, AVG\_Loss, AVG\_Delay, AVG\_Jitter)$
      $Received\_Message\_Count = 0;$
      $Total\_Throughput = 0$
      $Total\_Loss = 0$
      $Total\_Delay = 0$
      $Total\_Jitter = 0$
    else
      $Last\_Sequence = Message.Seq$
      $Last\_Delay = Current\_time - Message.Timestamp$
    end if
  end if
end for

The **SASHA Data Harvester** provides three types of information which are included with the QoS parameters in the evaluation of QMS scores.

Energy consumption is specific to each interface and is a parameter that changes less often than the QoS parameters. The energy consumption is expressed in J/Mb (Joule per

Megabit) and is ether calculated using data provided by the interface driver (if this data is available) or it can be preset as average estimations in the device settings. The latest is the simplest method however it is not highly accurate. The former is more accurate but it depends on the network interface driver to avail of the required information.

Monetary cost involved by transmitting data over a certain network is also harvested by this module and is expressed in Euro/Mb. In the same manner as energy, monetary cost can be obtained by interrogating corresponding services at the provider side or it can be present on the mobile device as estimated values.

The QoE represents the user Quality of Experience and in the context of video streaming applications is estimated by a video quality metric. For this purpose a non-reference video quality metric can be used which has to be supported by the multimedia application. As non-reference video quality metrics are computationally intensive leading to performance degradation when run on battery powered mobile devices, user QoE may be estimated at the server side using the streaming bitrate and the loss rates measured and reported by the client QoS monitoring module. Such a metric is the stream-based PSNR introduced by Lee in [180].

Although the first two parameters discussed here are not changing their values very often, for consistency reasons they are updates as often as the rest (i.e. QoS and QoE).

The **SASHA Feedback Controller** gathers average QoS parameter values, QoE, power and cost and sends them through a control channel to the SASHA Controller at the SASHA server-side module. The feedback message contains the data structure schematically illustrated in Figure 5.7.

SASHA Controller and SASHA Feedback Controller may use any reliable transport protocol to communicate although considering the real-time nature of the application some real-time constraints are applied to the control information as well. For this purpose the feedback messages are duplicated and sent over two distinct simultaneous connections. This improves the reliability at the cost of doubling the control overhead.

Figure 5.9 presents the sequence diagram describing the feedback procedure. As it can be seen in the diagram, SASHA QoS Monitor and SASHA Data Harvester monitor QMS

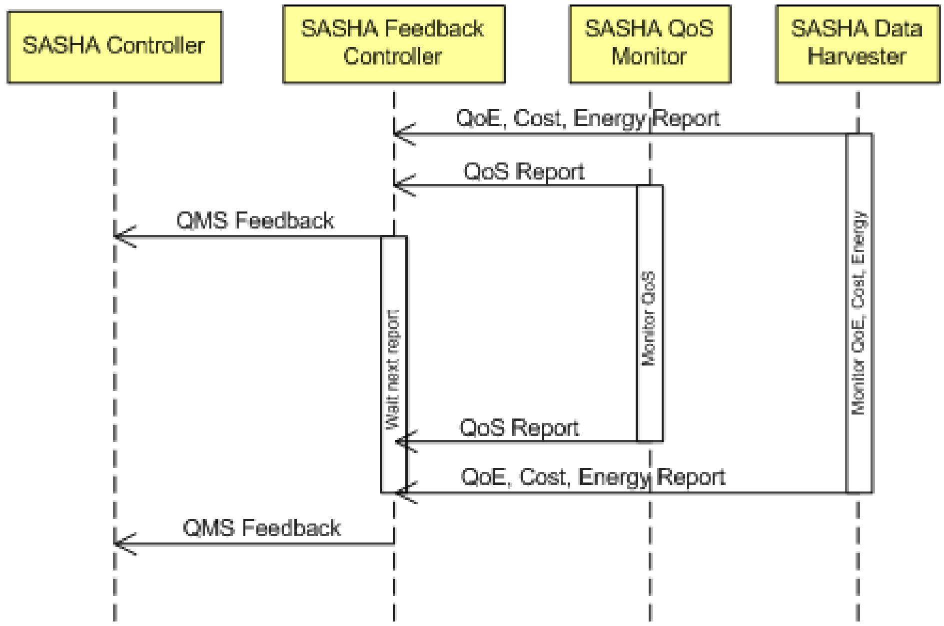

Figure 5.9: QMS feedback procedure

parameters and periodically report the average values to the SASHA Feedback Controller. When the report is received SASHA Feedback controller aggregates the data (QoS, QoE, Cost and Energy) in a data structure presented in Figure 5.7 and sends it to the server-side SASHA Controller.

5.5 SASHA Algorithm

As mentioned, SASHA simultaneously exploits several wireless links and performs handover by smoothly transferring the traffic load from one network to another. The traffic distribution decision is based on the QMS grades computed for each connection separately. The variation in QMS scores determine variations in the bitrates of the streams conveyed by each network. The QMS variation is determined by the variation in its parameters.

Link fading, network congestion, non efficient energy consumption, change in user preferences or monetary cost may trigger a handover by changing the QMS parameters which in turn determine a change in traffic distribution.

SASHA can be used for both horizontal and vertical handover, but for algorithm ex-

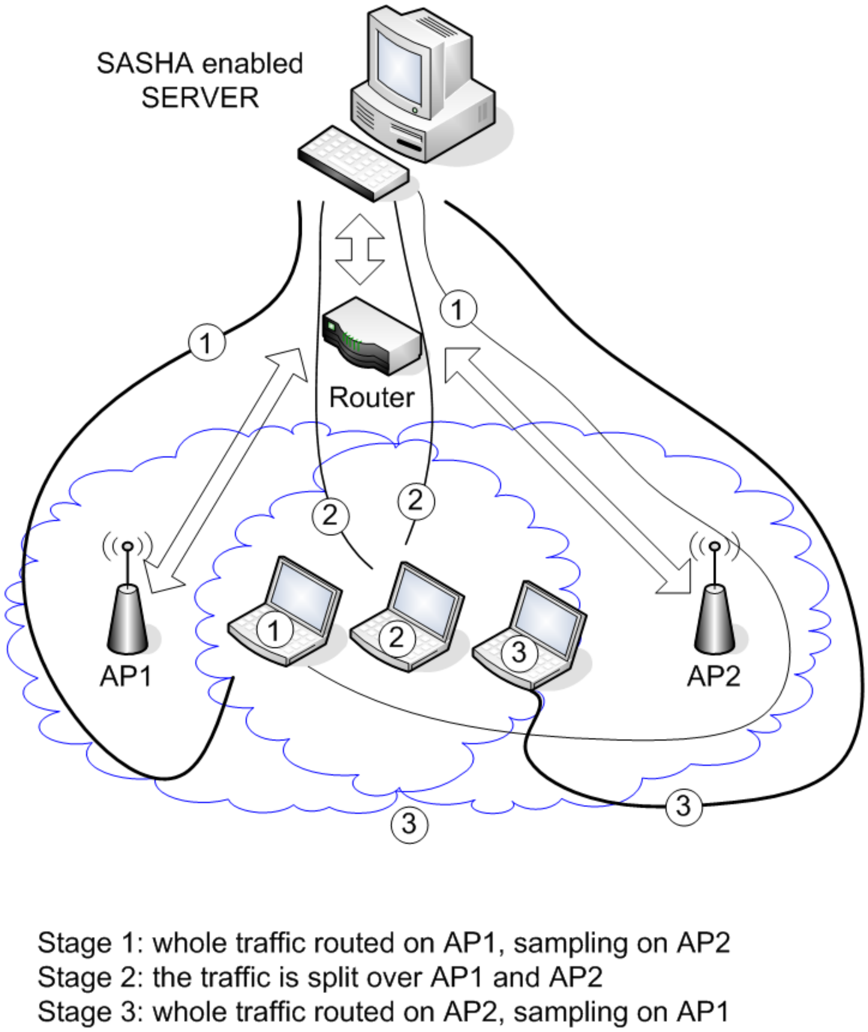

Figure 5.10: SASHA mechanism

emplification purposes the following scenario is considered, which includes a horizontal handover performed between two IEEE 802.11 networks. Moreover the same algorithm can be used when more than two networks are available. Such a generic algorithm is described later on in this chapter.

The scenario presented in Figure 5.10 involves two networks using infrastructure modes and having AR1 and AR2 as access routers and a mobile host (MH).

The mobile node is traveling from the coverage area of the network served by AR1 towards AR2's network coverage area, crossing the two networks overlapping area.

In Stage 1 the mobile node resides exclusively within AR1's coverage area. All the multimedia content is routed over the only available communication channel which is the one supported by AR1.

In Stage 2 MH enters the overlapping area and consequently the link via AR2 becomes available. MH sets a new communication channel to the server and the server sends the low bitrate sampling stream over the new channel to gather QoS information and to compute QMS for the new link. The QMS metric is now evaluated for the two communication channels resulting in values QMS1 and QMS2 respectively. Let us assume that due to the high distance to AR2, QMS2 is very much lower than QMS1. Consequently SASHA Controller decides to transfer all multimedia traffic via AP1.

When MH further moves towards AR2, AR1 signal strength starts fading, while AR2 signal strength increases and QMS scores computed for AR1's path decrease while QMS scores for AR2's link increase. Based on QMS scores, SASHA Controller starts to smoothly transfer the multimedia content from the AR1 communication channel to the AR2 communication link. This load transfer is gradual and based on a rate adaptive process which is performed based on the evolution of QMS values which are computed for each communication channel separately and updated periodically.

Finally, in Stage 3, when MH is about to enter exclusively in the AR2 coverage area the QMS values for the AR1 link decrease significantly, whereas the QMS value for the AR2 channel significantly increase. In these conditions QMS scores determine SASHA Controller to transfer all multimedia traffic over the AR2 network. While the AR1 link is still available, channel sampling is performed and QMS values are computed allowing the handover process to be reversed, if the MH moves back towards AR1's network.

In case the MH is roaming within the overlapping area of two (or more networks) the multimedia content will be continuously shared between the available communication links depending on the QMS scores. Consequently, SASHA's dynamic behavior accommodates any mobility pattern of the MH within the overlapping area of several networks. The ping-pong effect is avoided by using a QMS variation threshold. This is further detailed in algorithm pseudo code.

Algorithm 2 presents the pseudo-code of SASHA rate adaptation algorithm. The procedure is performed each time QMS related feedback is received from the client-side module. If the variation in QMS is significant according to the required algorithm sensitivity (a threshold value was introduced which is typically 10%), the rate adaptation procedure is triggered.

The first step consists of calculating the QMS scores for each communication channel separately. Based on the QMS values, p best channels are selected which have enough traffic capacity to deliver the multimedia content at the target bitrate and therefore at high quality. By choosing the channel with best scores and consequently higher data rates the number of concurrent channels is minimized. This is beneficial for traffic distribution, interface management and energy consumption. In order to avoid loading a communication channel to the limit or even overloading it, only 70% of the QMS score is accounted for in networks selection and rate allocation.

In the next step, the rate share is computed for each communication channel according to the QMS scores and application requirements. The QMS scores are expressed on a 100 point scale and represent the estimated share (expressed in percentage) of the total streaming rate that a certain connection can transport at high quality.

The rate $Rate_i$ associated with a connection i represents the amount of data from the total streaming rate which can be transported at high quality over the connection i and is calculated according to the connection's QMS score. The actual sending rate expressed in Mbps is computed from the application target bitrate and the previously computed QMS parameter as in the following equation. $Target\_Rate$ represents the application required bitrate.

$$Rate_i = Target\_Rate * QMS_i/100 \tag{5.1}$$

As input parameters, the algorithm uses, apart from the above mentioned $Target\_Rate$, $Loss\_Rate_i$ which represents the average loss expressed in Mbps, $Delay_i$ and $Jitter_i$ which are average values as well expressed in ms and $Throughput_i$ expressed in Mbps also as an average value computed during time interval t. $Monetary\_Cost_i$ and $Energy\_Consumption_i$

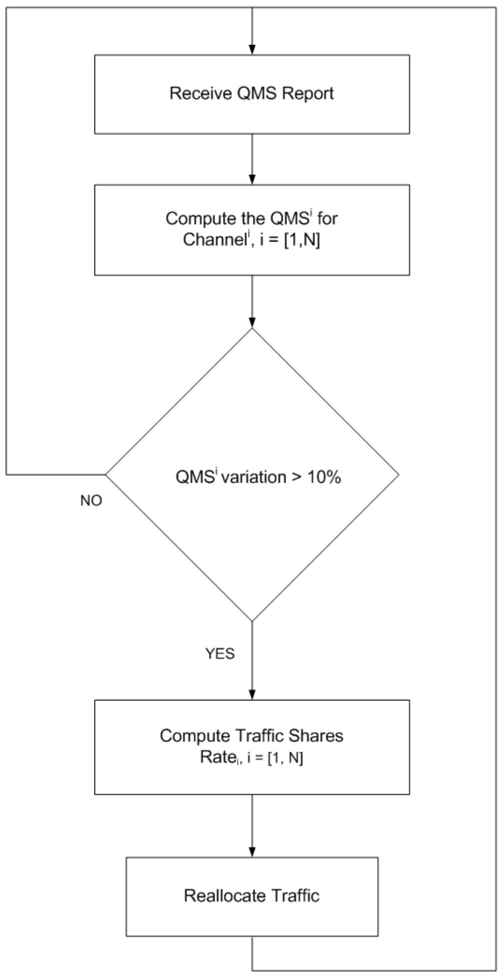

Figure 5.11: SASHA rate allocation algorithm

represent the monetary cost and energy consumption required for transporting the data over communication channel i and are expressed in Euro/Mb and J/Mb.

The last step distributes the traffic load according to the rates computed in the previous step. However in case the total capacity as estimated by QMS cannot cover the required target rate, the adaptation scheme is involved. A new target rate is computed to fit into the available capacity and rate allocation algorithm is performed again with the new value of the target rate.

Figure 5.11 presents the block diagram of the rate allocation process as described in Algorithm 2.

For increased performance the sensitivity and reaction speed of the algorithm has to be

Algorithm 2 SASHA Handover Algorithm

Input:

$Target\_Rate$;
$Loss\_Rate_i$;
$Delay_i$;
$Jitter_i$;
$Throughput_i$;
$Monetary\_Cost_i$;
$Energy\_Consumption_i$;

Output:

$Rate_i$;

Procedure:

```
START:
i ⇐ 0
for all i such that 0 ≤ i ≤ No_Networks do
   Compute QMS_i
end for
if QMSvariation > Threshold then
   Sort QMS rates ascending;
   Total_Ratio ⇐ 0
   for all i such that 0 ≤ k ≤ No_Networks do
      if Total_Ratio < 100 then
         if Total_Ratio + QMS_i < 100 then
            Rate_i = Target_Rate ∗ (100 − Total_Ratio)/100;
            Total_Ratio = 100;
         else
            Rate_i = Target_Rate ∗ QMS_i/100;
            Total_Ratio = Total_Ratio + QMS_i;
         end if
      else
         Rate_i = 1Kbps;
      end if
   end for
   if Total_Ratio < 100 then
      Target_Rate = Target_Rate ∗ Total_Ratio
      GOTO : START
   end if
end if
```

correlated with environmental factors like network dynamics, size of the networks, overlapping areas and also MH speed and trajectory.

The QMS scores are computed by the server side module while the QMS parameters are harvested by the client side module. Consequently the proposed solution involves a certain network overhead determined by QMS feedback sent by the client to the server. Although some QMS parameters (i.e. energy and monetary cost) change more seldom then the others, for consistency and robustness all parameters are updated with the same frequency.

The time interval at which QMS feedback is sent from the client-side module to the server-side module impacts the reaction time of SASHA handover algorithm it requires extensive evaluation and testing in real application environments. However typical values for the feedback time interval ($[t_n, t_{n+1}]$) are between 500ms and 1000ms, usually the value of t being a multiple of the average round-trip time (RTT) measured by the client through probing.

Solutions like MIP and Mobile DCCP present less network overhead as the decision is made by the mobile device (client) and only a location update is required. However if these mobility management solutions are used in conjunction with a feedback-based adaptive multimedia streaming scheme, when using SASHA there is a significant advantage of sending the feedback information only once and therefore reducing the overall overhead by approximately 50%.

Computing the QMS scores and making the decision at the server side has the benefit of a reducing energy consumption, processing power and delay at the client side. Many handover decision making algorithms suffer from the increased complexity which lead to high processing power requirements, and consequent delay and energy consumption. This is not the case for SASHA due to the server driven handover mechanism.

5.6 Traffic Splitting and Merging

Traffic splitting and merging is a very important process of the SASHA handover management algorithm. It has the role of distributing the main application traffic over the existing connections. SASHA adopts two methods for splitting the main multimedia stream.

5.6.1 Packet-based Traffic Splitting

The first method simply allocates the packets in the queues of each connection according to the data rates computed for each of them separately. The basic principle of this method is described in Figure 5.12.

The packets forming the main application data stream are organized in groups, each group containing a different number of packets. The number of packets in each group has to correspond with the rate allocate to the communication channel on which the packet group will be sent.

The main advantage of this solution is its computational simplicity which makes it suitable for any application which accepts a certain level of delay and packet reordering.

The merging process consists mainly in reordering the packets received over each communication channel into one single array.

The major disadvantage of this solution appears at the merging process and it is the before mentioned packet reordering which is determined by the variation in delay between the multiple connections used and which complicates the merging process. Packet reordering is time consuming and in case of high delay variations it may be unacceptable for time constrained multimedia applications leading to packet loss.

5.6.2 Stream-based Traffic Splitting

The second method is a stream-based traffic splitting and is applicable to multimedia content exclusively. The basic principle is presented in Figure 5.13.

The original video frames are split into multiple sub-frames using an interlaced-based method. This method allocates each frame row to a different description in a round-robin manner. The resulting sub-frames are then encoded with a standard MPEG2/MPEG4 encoder resulting in multiple independent video streams.

The encoder is equipped with a real-time adaptation module which allows for each

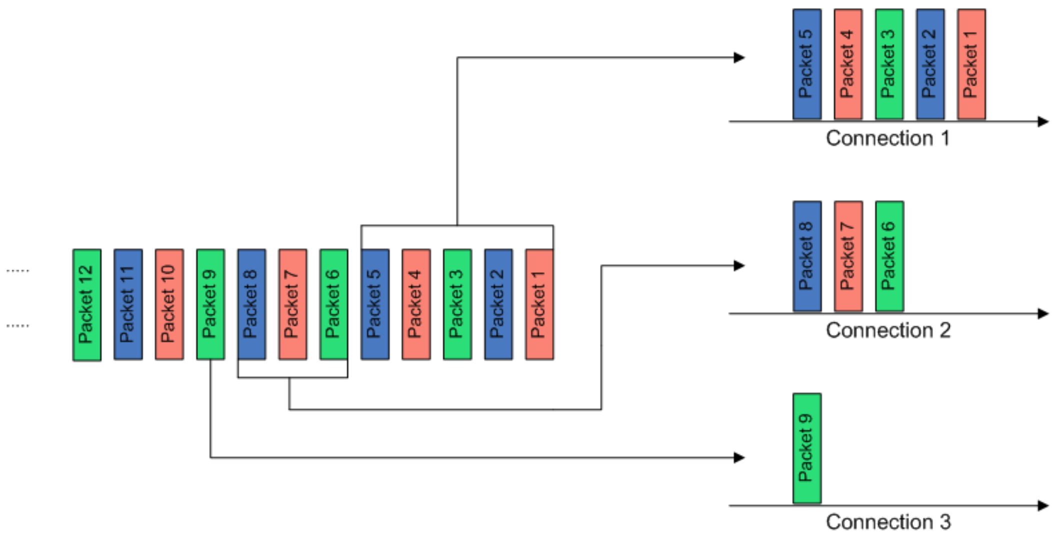

Figure 5.12: Packet-based traffic splitting

stream to be encoded at the bitrate determined by the SASHA algorithm for its corresponding connection. The adaptation is performed by varying the quantization parameter which in turn adjust the compression ratio and consequently the content's bitrate.

At the client side the separate streams are decoded and the resulting frames are merged in the same manner they were split.

The main advantages of this approach is its improved resilience to packet delay variations and increased robustness.

The disadvantage is the alteration of the resolution of individual descriptions which may be perceived by the user in case one or more descriptions are lost.

The advantages of this solution makes it the best candidate for integration with SASHA handover management algorithm.

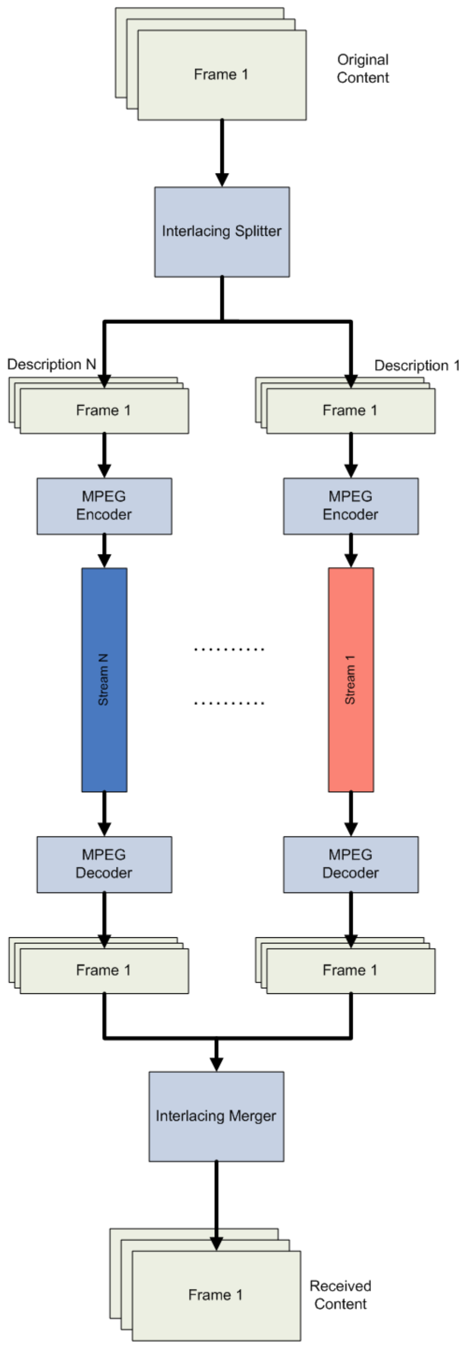

Figure 5.13: Stream-based traffic splitting

5.7 Performance Goals

The main impact of handover operation on active data sessions consist mainly in increased packet delay and packet loss which affect QoS and consequently QoE. In an ideal scenario, performing handover should not disrupt in any way the ongoing data sessions, preserving the required QoS levels and maintaining the user perceived QoE at high levels.

Consequently to evaluate the performance of the handover algorithm the evolution of the packet loss and packet delay should be monitored as well as the video quality of the received multimedia content as an estimation of user QoE. The target is to achieve best performance in terms of video quality with scores exceeding 50 dB for video quality assessed by PSNR and with no more than a 10% percent average loss rates in any networking scenarios considered.

Apart from the main goal which is to preserve QoS levels and maintain high levels of user QoE the handover solution should also be evaluated from the point of view of scalability, stability and robustness. All these aspects will be evaluated by simulations and experiments using an emulator prototype.

Another performance aspect of SASHA is the control overhead. The QMS information updates require on average a 128bytes message to be sent to the server-side SASHA Controller. Depending on the networks dynamic the frequency of these update message may vary, between 500 ms and 1000 ms. This generates an overhead of 128 bytes to 256 bytes per second which is an insignificant amount of data compared with the bitrate required by the multimedia content.

To exemplify a 10 minutes video clip encoded at 2 Mbps occupies 150 MBytes while the feedback overhead required is only between 0.07 Mbytes and 0.14 MBytes.

However, as mentioned before, when an adaptive multimedia streaming application is used the feedback required by the adaptive technique can be used by SASHA as well and vice-versa. Depending on the level of correlation between the feedback requirements of SASHA and the adaptive scheme and overhead reduction of around 50% can be obtained.

5.8 Chapter Summary

This chapter introduced the Smooth Adaptive Soft-Handover Algorithm (SASHA) a novel handover management solution for multimedia content delivery to mobile devices over heterogeneous wireless environment.

Heterogeneous wireless networks are discussed followed by the introduction of SASHA architecture which consists of a server-side module and a client-side component. Each of the two modules incorporate several sub-modules such as SASHA Controller, SASHA Connection Manager, SASHA Data Harvester, SASHA QoS monitor and SASHA Feedback Controller. Each of these sub-modules are detailed and discussed.

Based on a concrete example SASHA algorithm is introduced and discussed in details.

Last but not least methods for traffic splitting and merging, used by SASHA in the handover management process, were presented followed by the performance goals.

Chapter 6

Quality of Multimedia Streaming Metric Evaluation

6.1 Introduction

Quality of Multimedia Streaming Metric (QMS) is a novel capacity estimation mechanism for wired and wireless networks. QMS estimates how much traffic can be transported over a certain network in order to efficiently deliver high quality multimedia content to mobile devices.

In a heterogeneous wireless environment various wireless networks coexist and cover a certain geographical area in cellular manner with overlapping coverage. Each network has its characteristics in terms of coverage, available bandwidth, monetary cost, energy consumption of the corresponding network interface, QoS provisioning, etc.

Under these circumstances offering the user with a seamless and high quality data communication service which also accommodates mobility becomes a challenging task. User satisfaction is extremely important for the success of any service, consequently all the factors regarding data communication and service delivery to the mobile device which may affect user quality of experience have to be considered.

QMS is a comprehensive and highly flexible metric which incorporates several parameters including network QoS, user QoE, monetary cost, energy consumption and user preferences. All these parameters are incorporated in the QMS function in a flexible manner. Each parameter has a weight which can be adjusted to modify the effect of each parameter on the overall QMS score and is expressed on a 0 to 100 scale. In order to achieve normalization, the sum of the weights attributed to the QMS parameters has to be unitary.

As mentioned above the weight of the a parameter modifies its importance in the computation of the overall score. The importance of each parameter represents basically the importance of the network characteristic it measures. Consequently the variation of QMS parameter's weights distribution impacts the overall behavior and performance of the application which uses SASHA.

The QMS evaluation presented in this chapter is done in the context of SASHA as a handover management algorithm which uses QMS for rate allocation purposes. The focus of this analysis is on evaluating QMS parameters and not SASHA performance.

In the following subsections the simulation environment and scenarios are presented as well as testing results. Suggestions for setting the values for QMS parameters are made in the context of the SASHA handover algorithm.

As the QMS metric is highly flexible and complex here it is not tested in every possible scenario and application context instead a methodology for choosing the weights is proposed in order to optimize QMS usage in the context of a specific application. The application developer will be in charge of fine tuning the QMS parameters importance and setting the weights through field testing for any particular application. Machine learning techniques also can be used to perform in-service evaluation of weighting schemes and parameter tunning, however these aspects are not the subject of this work.

6.2 Simulation Environment and Models

The simulations were performed using the NS-2 Network Simulator (v2.29) [181] enhanced with the No Ad Hoc (NOAH) [182] wireless routing agent in order to allow only direct communication between the mobile hosts (MH) and the access points (AP) or base stations

(BS). To obtain a more realistic behavior of the wireless communication scenarios, the realistic radio patch developed by Marco Fiore [183] was also included in the simulation platform.

SASHA was deployed in NS-2 as an application which simulates a multimedia streaming application with a server-side and a client-side component. The application is capable of sending a constant bitrate multimedia content for handover management. SASHA uses QMS scores computed for each wireless link separately to determine the corresponding communication channel, which will be used to send each of the data packets.

As SASHA requires simultaneous communication over multiple connections the standard implementation of the NS-2 wireless node was altered. Each node is equipped with several wireless interfaces, each interface being able to communicate over a different wireless channel. The channel used by each interface is set when the node is created and the SASHA module at the application level is made aware of the existence of each new separate channel.

Figure 6.1 schematically presents the client-side module used in NS-2 for simulating the SASHA algorithm. Several channels are used to allow multiple simultaneous connection to be used. The network interface and the subsequent protocols have been replicated for each network (channel) separately. A distinct transport protocol agent, UDP in this case as multimedia content will be transported, for each network was attached to the mobile node. The transport agent is connected to its corresponding agent located at the server node. Consequently each network will have a distinct pair of connected agents, one on the mobile node and the other on the server node. This represents the connection pool which will be used to convey the multimedia content to the mobile device.

The client-side SASHA-enabled multimedia application uses the transport agents and their corresponding connections to receive the multimedia traffic from the server. This application module also monitors and reports to the server the QoS parameters (i.e. throughput, loss, delay, jitter) and the QoE estimated using a stream-based video quality metric, according to the description in Chapter 4.

The server node is presented schematically in Figure 6.2. The server node is a wired

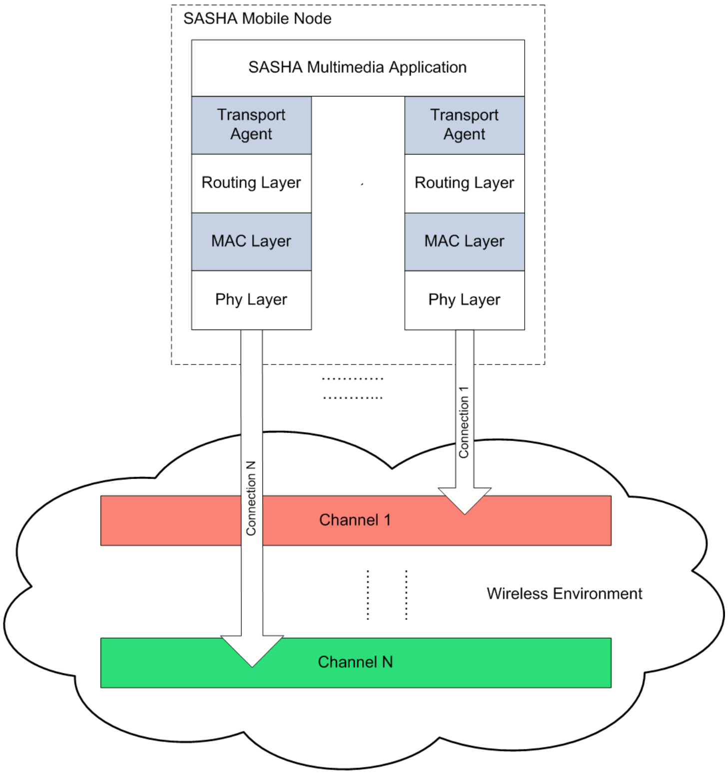

Figure 6.1: NS-2 client-side component

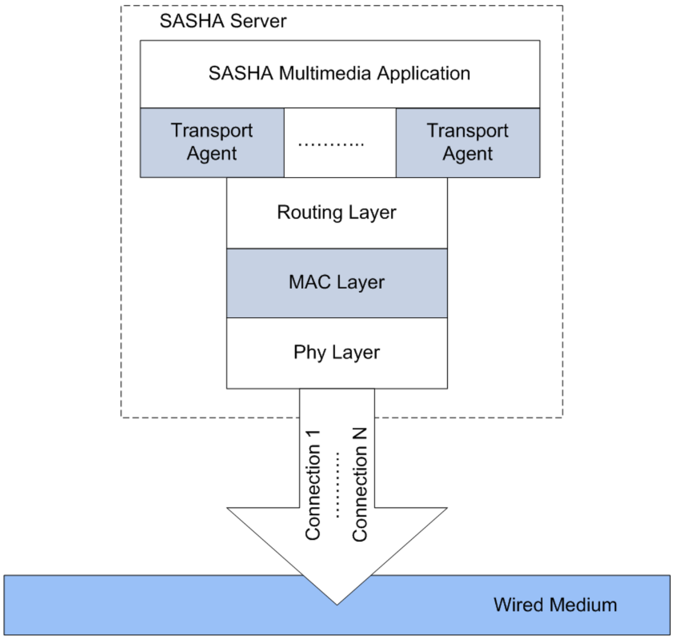

Figure 6.2: NS-2 SASHA server-side component

node and has a single high bandwidth wired connection. The server-side SASHA-enabled multimedia application computes the QMS scores and distributes the traffic over the existing connections.

The server-side application has the monetary cost and energy consumption preset during the simulation session. This is determined by the fact that these parameters are quasi-static and consequently the assumption is made that they are not changing during the active data transfer session.

6.3 Simulation Scenarios

The simulation topology used is presented in Figure 6.3. Four different wireless networks (IEEE 802.11b) are considered, with theoretical data rates of up to 11Mbps. The actual maximum data rate as measured during simulations was 5Mbps per access router.

The corresponding access routers (AR) are positioned at equal distance forming a 5m square. The transceiver energy level was set in such a way that a sufficient overlapping area was obtained. The overlapping area was tested by measuring the maximum throughput received in various locations.

The SASHA server-side and client-side modules were deployed as distinct applications and the corresponding UDP agents were attached and connected.

The SASHA enabled mobile terminal is positioned within the overlapping area of the four networks. Each network uses a separate non-interfering channel, allowing the multi-homed mobile terminal to communicate simultaneously with all four networks.

As presented in Figure 6.3, Network 1 and Network 3 as well as Network 2 and Network 4 are connected through wired links to a separate common router. The two common routers are connected through two separate wired links to a main router which is further connected to the SASHA enabled media server.

Mechanisms were deployed within the two SASHA-modules (client-side and server-side) to measure QoS parameters (i.e. throughput, loss, delay and jitter) and to compute the stream-based PSNR metric.

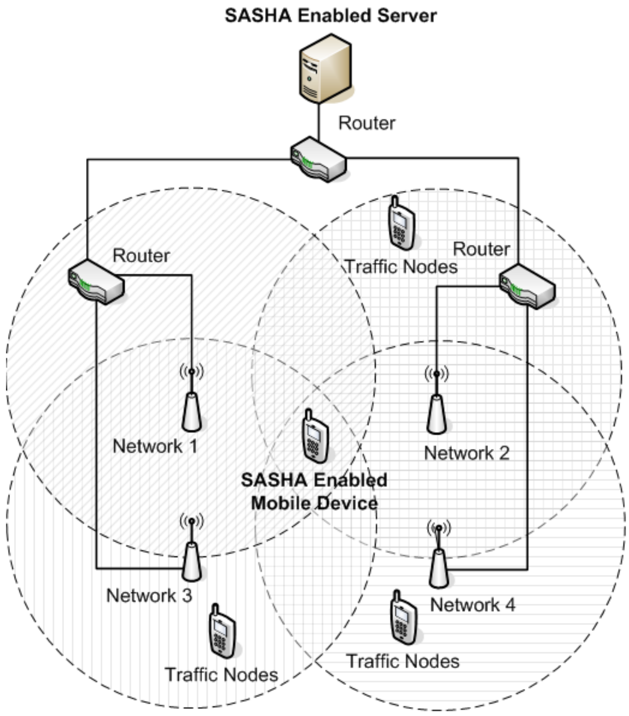

Figure 6.3: Simulation topology

The pricing information is measured in Euro/Mb and the energy requirements in Joule/Mb for each network separately. This information is assumed to be preloaded in the mobile device and is preset in the server side module at the start of each data session.

The application scenarios include streaming 2Mbps CBR video content to the mobile host through multiple simultaneous links. Traffic distribution is performed based on assessing the QMS metric for each the four networks separately.

Considering the fact that SASHA demonstrates its benefit based on QMS readings especially in situations when the available networks are highly loaded or radio links have poor quality these types of network conditions were simulated in the following scenarios.

SCENARIO 1: the wired links' bandwidth was limited, simulating congestion at the core network level. The bottleneck bandwidths are 0.2Mbps, 0.4Mbps, 0.6Mbps and 0.8Mbps respectively and are illustrated with various degrees of thickness in the links presented in Figure 6.4. All other wired links are such provisioned not to generate any bottleneck (100Mbps). It can be noted that none of the four networks can support the 2Mbps traffic

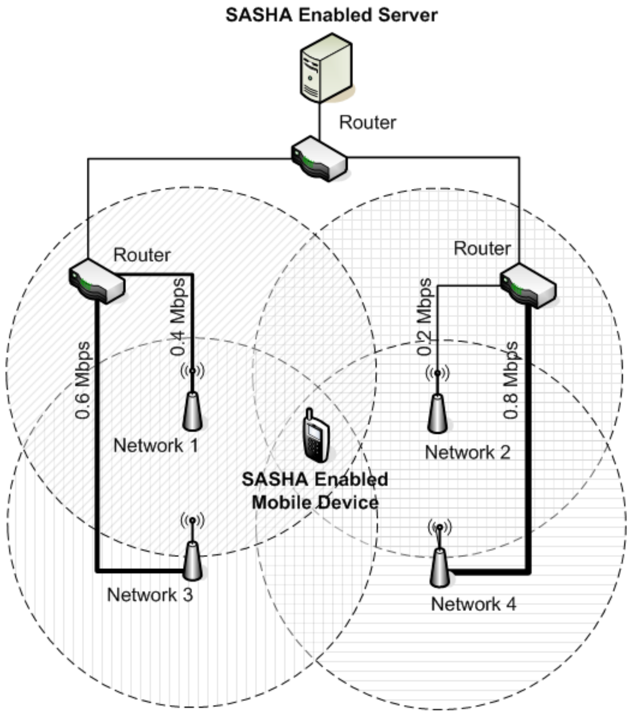

Figure 6.4: Simulation topology for the first scenario

alone, but their aggregated capacity reaches the required target.

SCENARIO 2: the core wired network does not present any congestion, all the wired links are provisioned (100Mbps) not to cause any negative influence on the data traffic. Traffic is generated by one separate wireless traffic node connected to each of the four wireless networks as presented in Figure 6.5. Each node generates heavy traffic (up to 4.5Mbps) to each wireless router determining high level of congestion. The background traffic pattern considered is constant bit rate (CBR) multimedia traffic.

SCENARIO 3: CBR traffic (up to 4.5Mbps) is generated at the wireless networks level by multiple wireless nodes (up to four) in each cell as it can be seen in Figure 6.6. This scenario was selected to evaluate the performance of the proposed metric in the context of higher level of collisions determined by the higher number of mobile nodes present in each cell.

SCENARIO 4: both wired and the wireless links present no additional traffic apart from that generated by SASHA server. However the wired links present different delays

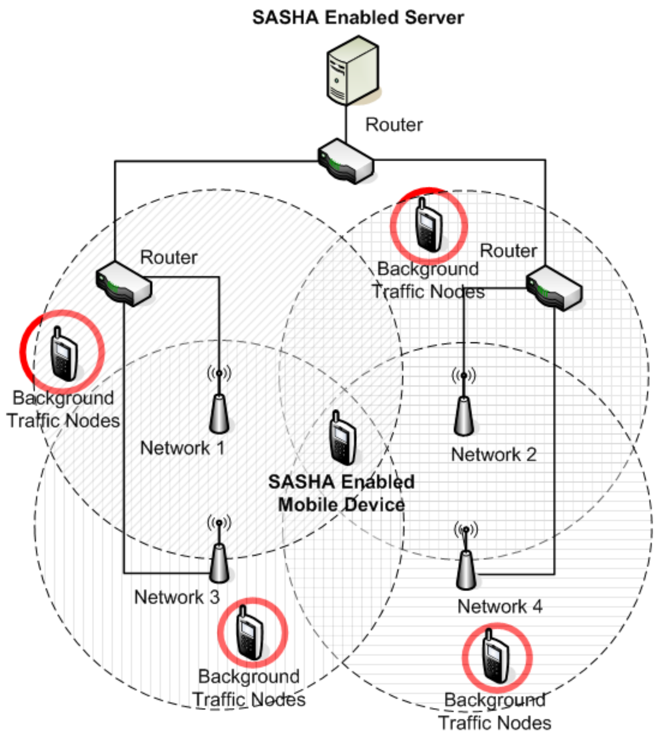

Figure 6.5: Simulation topology for the second and fifth scenario

determining each separate path to have different delay levels. The delays presented by the wired links through which Network 1, Network 2, Network 3 and Network 4 are connected are 60 ms, 80 ms, 100 ms and 120 ms respectively as illustrated in with dashed lines in Figure 6.7.

SCENARIO 5: the wired links are not loaded with any additional traffic presenting full bandwidth (100Mbps) while the wireless networks each have a mobile node acting as background traffic generator. The traffic nodes generate randomly exponential burst traffic in order to generate variable jitter over the multiple simultaneous connection used. The traffic was generated using the default exponential traffic generator implemented in NS-2. The traffic generation was triggered at random time intervals chosen for each background traffic node. The network topology used is the same with SCENARIO 2 and is illustrated in Figure 6.5.

Each simulation runs for 600s with the main data traffic being initially equally distributed over the four networks.

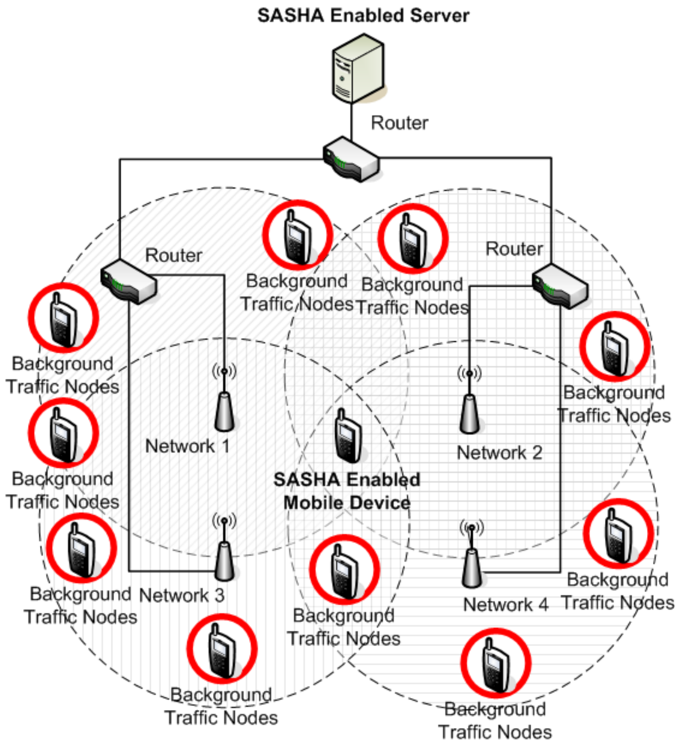

Figure 6.6: Simulation topology for the third scenario

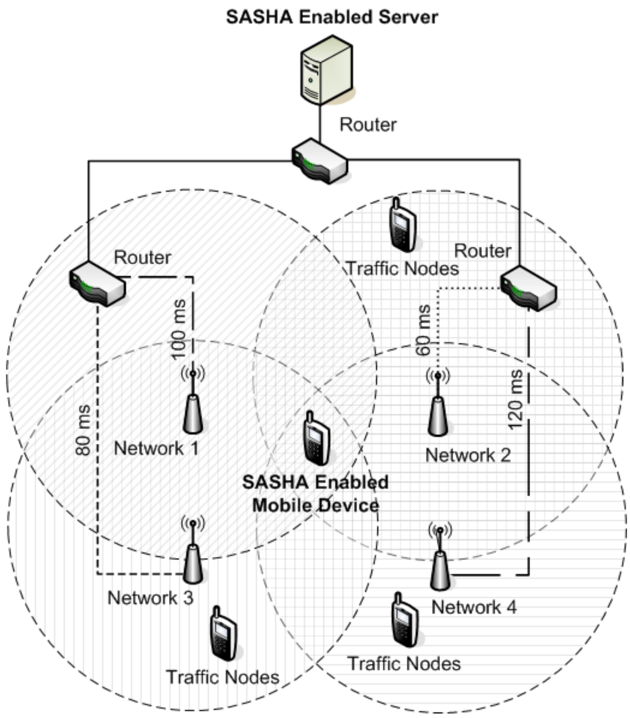

Figure 6.7: Simulation topology for the fourth scenario

6.4 Results and Result Analysis

As the QoS parameter is the most important [184], it is assessed first. All other parameters are evaluated later on with respect to QoS in terms of weight values.

6.4.1 QoS Component Assessment

Performance analysis is made in terms of estimated user QoE measured by a stream-based version of the Peak Signal-to-Noise Ratio (PSNR) metric [180]. PSNR is computed based on the throughput and streaming bandwidth [180].

6.4.1.1 Throughput and Loss

Table 6.1, Table 6.2 and Table 6.3 present the average PSNR, throughput and loss measured at the mobile device when various combinations of throughput and loss weights are used. W_T vary between 0.0 and 1.0 and W_L takes values between 1.0 and 0.0. The normalization condition is respected ($W_T + W_L = 1$).

QoS_{grade} assessment was performed in the context of congested wired links (Scenario 1), congestion determined by a small number of nodes generating high bit rate traffic (Scenario 2) and a larger number of traffic nodes (Scenario 3). The other two scenarios, Scenario 4 and Scenario 5 are only relevant for delay and jitter, consequently were not consider at this stage.

For the first scenario, relying only on throughput (W_T = 1 and W_L = 0) good quality is achieved (PSNR = 36.45dB) but including the loss component in the QMS evaluation (W_T = 0.9 and W_L = 0.1) the quality increases by 30% (PSNR = 46.24dB). Due to its optimistic behavior, by increasing the weight of the loss component of the QMS a serious decrease in user perceived quality is obtained as throughput is reduced ($W_T > 0.3, PSNR < 10dB$).

In the second scenario the behavior is similar. When using the throughput component in QMS, SASHA offers good quality in terms of average PSNR (20.64db). When including the loss component ($W_L > 0$), a slight increase in quality is achieved. Any further increase in the weight for the loss component determines quality degradation to unacceptable levels

Table 6.1: QoS evaluation for the congestion Scenario 1

W_T	W_L	PSNR (dB)	Throughput (Mbps)	Loss (Mbps)
0	1	0.88	0.0851	1.9149
0.1	0.9	1.17	0.1045	1.8955
0.2	0.8	1.19	0.1059	1.8941
0.3	0.7	2.96	0.2071	1.7929
0.4	0.6	3.21	0.2176	1.7824
0.5	0.5	3.63	0.2450	1.7550
0.6	0.4	9.20	0.7071	1.2929
0.7	0.3	24.78	1.4208	0.5792
0.8	0.2	24.78	1.4206	0.5794
0.9	0.1	46.24	1.8007	0.1993
1	0	36.45	1.6764	0.3236

(much bellow 20dB which is the lower acceptable limit). The quality drops as low as 4 dB when the QMS relies on the loss component only.

The third scenario also presents a better performance when both QMS throughput and loss components are considered. The maximum quality (PSNR = 34.74dB) is obtained when the throughput's component weight is set to 0.8 and the loss weight is set to 0.2.

Based on the testing results presented above the QMS metric has to be fine-tuned for best performance. However the results analysis shows that SASHA using QMS achieves good performance in terms of user perceived quality for wider range of weight values as it can also be observed in Figure 6.8.

QMS presents good performance for WT and WL ranging from 1 to 0.7 and 0 to 0.3 respectively. To determine which combination of weights offers the best performance, a 15 minutes video clip (a compilation of the ones used for SASHA performance evaluation in Chapter 7) was encoded at the stable bit rates (the average throughput presented in tables) achieved by SASHA for each of the three scenarios described above and all four weights

Table 6.2: QoS evaluation for the congestion scenario 2

$\mathbf{W_T}$	$\mathbf{W_L}$	**PSNR (dB)**	**Throughput (Mbps)**	**Loss (Mbps)**
0	1	4.00	0.1523	1.8477
0.1	0.9	4.06	0.1456	1.8544
0.2	0.8	4.60	0.1494	1.8506
0.3	0.7	5.36	0.1519	1.8481
0.4	0.6	4.14	0.1566	1.8434
0.5	0.5	3.99	0.1478	1.8522
0.6	0.4	12.78	0.7705	1.2295
0.7	0.3	12.14	0.7934	1.2066
0.8	0.2	22.60	1.2830	0.7170
0.9	0.1	19.70	1.1375	0.8625
1	0	20.64	0.4364	1.5636

combinations grayed in table were used. The video quality was objectively assessed based on the VQM metric using the MSU Video Quality Assessment Tool [185].

VQM assessment for the three scenarios is presented in Fig. 6.9. A t-test performed on the results involving W_T = 0.7 & W_L = 0.3 and W_T = 0.8 & W_L = 0.2 respectively has confirmed that there is a statistical difference between these results in favor of the later with a significance level of $\alpha = 0.1$ (99% accuracy).

A similar t-test between the results involving W_T = 0.9 & W_L = 0.1 and W_T = 1 & W_L = 0 respectively showed a statistical difference in favor of the first with a confidence level of $\alpha = 0.1$. As there was no significant statistical difference between the results obtained when W_T = 0.9 & W_L = 0.1 and W_T = 0.8 & W_L = 0.2 respectively it can be concluded that the point of quality maxima is obtained when W_T & W_L are chosen in between these levels. In consequence the following values were chosen as default for the two weight parameters, W_T = 0.85 & W_L = 0.15.

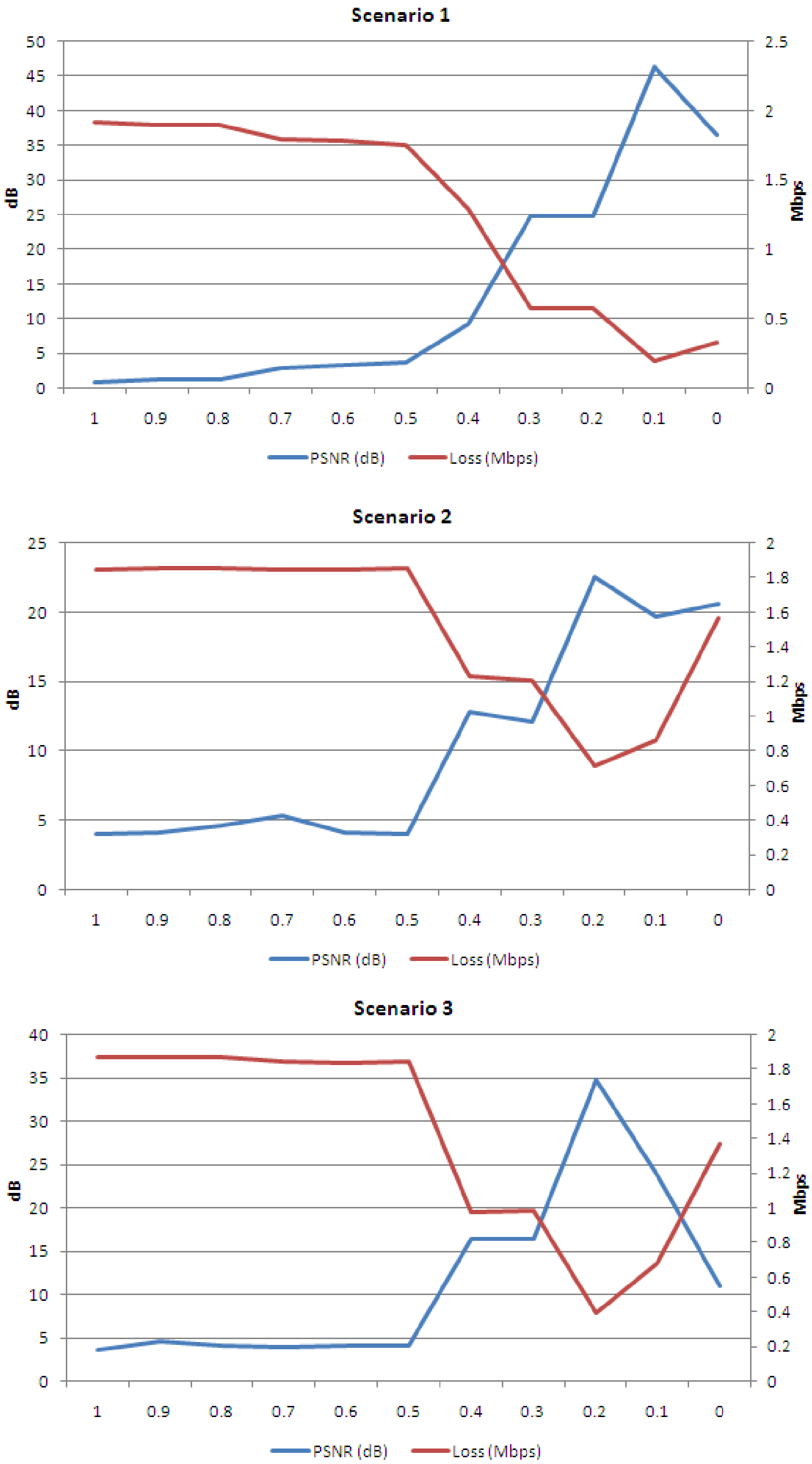

Figure 6.8: Loss vs. PSNR for the three congestion scenarios plotted against $\mathbf{W_L}$

Table 6.3: QoS evaluation for the congestion scenario 3

W_T	W_L	PSNR (dB)	Throughput (Mbps)	Loss (Mbps)
0	1	3.58	0.1284	1.8716
0.1	0.9	4.63	0.1295	1.8705
0.2	0.8	4.11	0.1295	1.8705
0.3	0.7	4.00	0.1536	1.8464
0.4	0.6	4.03	0.1603	1.8401
0.5	0.5	4.08	0.1572	1.8428
0.6	0.4	16.36	1.0208	0.9792
0.7	0.3	16.42	1.0201	0.9802
0.8	0.2	34.74	1.6053	0.3947
0.9	0.1	23.81	1.3219	0.6781
1	0	10.94	0.6287	1.3713

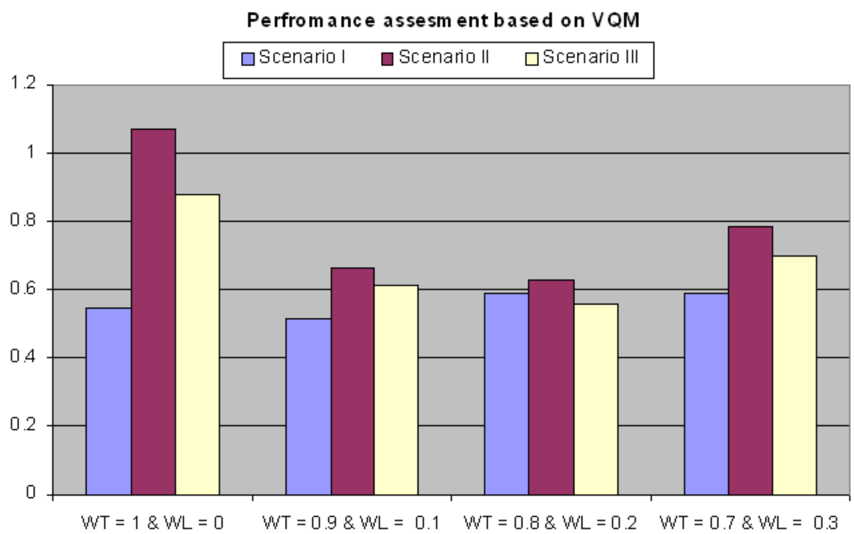

Figure 6.9: Video quality assessed using VQM

6.4.1.2 Delay Component

The delay component of the QMS metric was evaluated using the simulation Scenario 4 presented in Section 6.3. The results of the simulation are presented in Table 6.4. The weights

Table 6.4: Delay Evaluation

$\mathbf{W_{Thr.}}$	$\mathbf{W_{Loss}}$	$\mathbf{W_{Delay}}$	**PSNR (dB)**	**Throughput (Mbps)**	**Delay (ms)**
0.850	0.150	0.0	89	1.99	275
0.765	0.135	0.1	100	2.00	203
0.680	0.120	0.2	92	2.00	167
0.595	0.105	0.3	100	2.00	161
0.510	0.090	0.4	100	2.00	138
0.425	0.075	0.5	100	2.00	130
0.340	0.060	0.6	100	2.00	125
0.255	0.045	0.7	89	2.00	122
0.170	0.030	0.8	93	1.94	116
0.085	0.015	0.9	84	1.93	116
0.000	0.000	1.0	58	1.74	123

of the throughput and loss components are calculated so that the ratio as demonstrated in section 6.4.1.1 (i.e 0.85 and 0.15) is kept.

It can be observed in the Table 6.4 that the received throughput is reasonably constant for delay component weight taking values in the interval 0.0 and 0.9. For the case when the adaptation is performed based on the delay component only the throughput degrades significantly.

Considering the video quality estimated by PSNR and the measured delay, the default value for delay component weight was chosen to be 0.8. This value offers the best tradeoff between quality and delay.

Figure 6.10 presents the variation of PSNR and packet delay when $\mathbf{W_D}$ varies between 0 and 1 ($\mathbf{W_T}$ and $\mathbf{W_L}$ keep the proportions determined in the previous section). It can be observed that the minimum delay while keeping the PSNR at a sufficient high level is achieved for $\mathbf{W_D} = 0.8$.

Figure 6.11 presents the video quality evaluation using VQM as the quality metric. The

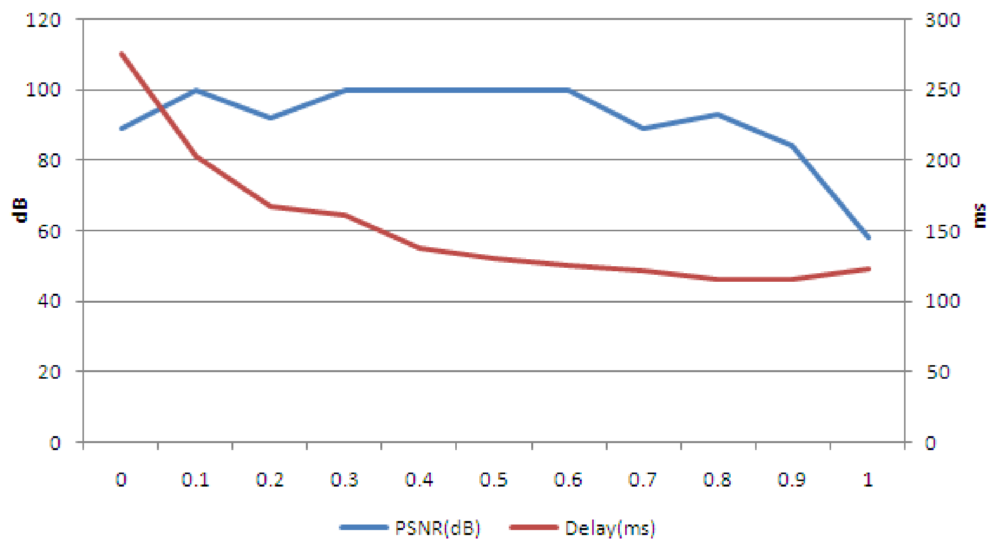

Figure 6.10: Delay vs. PSNR plotted against $\mathbf{W_D}$

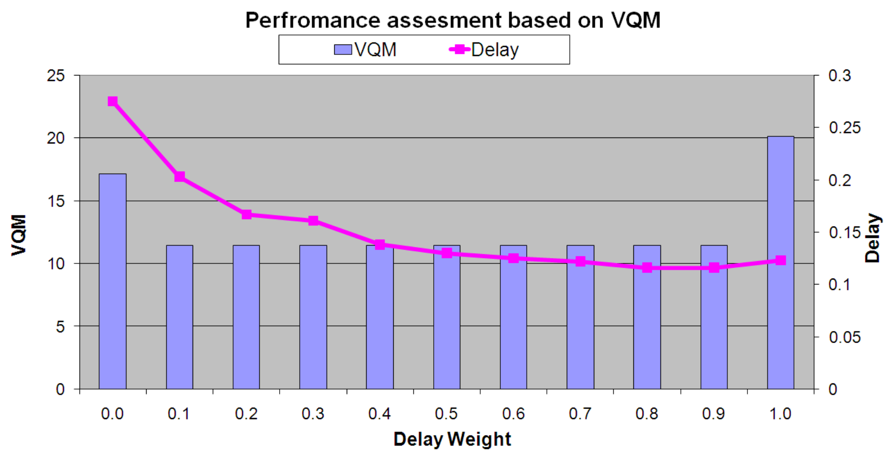

Figure 6.11: Video quality assessed using VQM

video quality is plotted against measured average delay for each weight value attributed to the delay component separately. The lower the VQM score the better the video quality. The result demonstrates how a weight of 0.8 determines the best results in terms of both quality and average delay.

Table 6.5: Jitter Evaluation

$\mathbf{W_{Thr.}}$	$\mathbf{W_{Loss}}$	$\mathbf{W_{Jitter}}$	**PSNR (dB)**	**Throughput (Mbps)**	**Jitter (ms)**
0.850	0.150	0.0	89	1.99	273
0.765	0.135	0.1	96	2.00	167
0.680	0.120	0.2	99	2.00	71
0.595	0.105	0.3	71	1.99	76
0.510	0.090	0.4	80	1.98	56
0.425	0.075	0.5	81	1.97	62
0.340	0.060	0.6	71	1.92	62
0.255	0.045	0.7	70	1.86	59
0.170	0.030	0.8	62	1.79	58
0.085	0.015	0.9	55	1.73	62
0.000	0.000	1.0	49	1.67	64

6.4.1.3 Jitter Component

The jitter component of the QMS metric was evaluated using the simulation Scenario 5 presented in Section 6.3. The results of the simulation are presented in Table 6.5. The weights of the throughput and loss components are set to the values determined in Section 6.4.1.1 (i.e 0.85 and 0.15 respectively).

It can be observed in the table that the received throughput is reasonably constant for the jitter component weight taking values in the interval 0.0 and 0.6. For the rest of the weight values the throughput starts to degrade. Considering the video quality estimated by PSNR and the measured jitter, the default value for jitter component weight was chosen to be 0.4. This value offers the best tradeoff between quality and jitter.

Figure 6.12 presents the variation of PSNR and packet delay jitter when $\mathbf{W_J}$ varies between 0 and 1 ($\mathbf{W_T}$ and $\mathbf{W_L}$ keep the proportions determined in the previous section). It can be observed that the minimum jitter while keeping the PSNR at a sufficient high level is achieved for $\mathbf{W_J} = 0.4$.

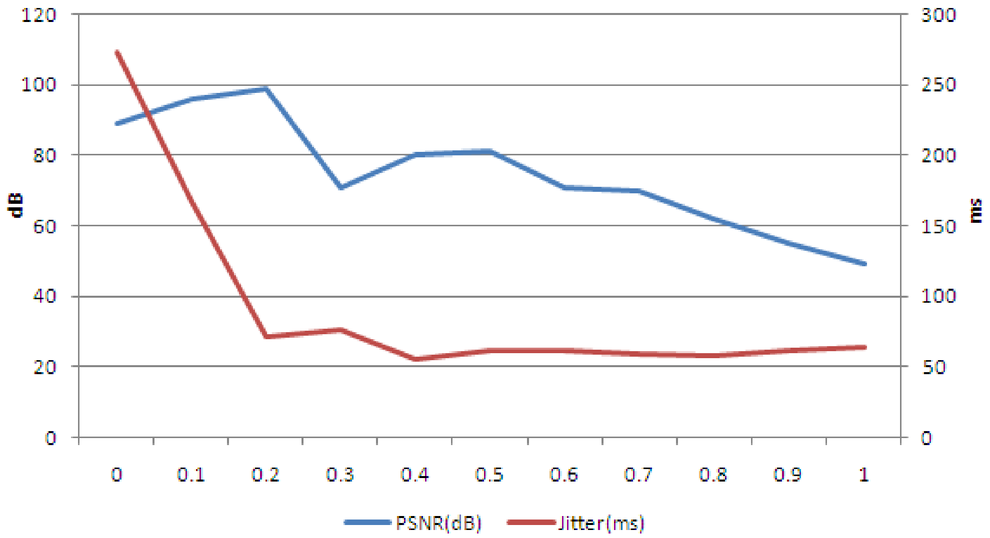

Figure 6.12: Jitter vs. PSNR plotted against $\mathbf{W_J}$

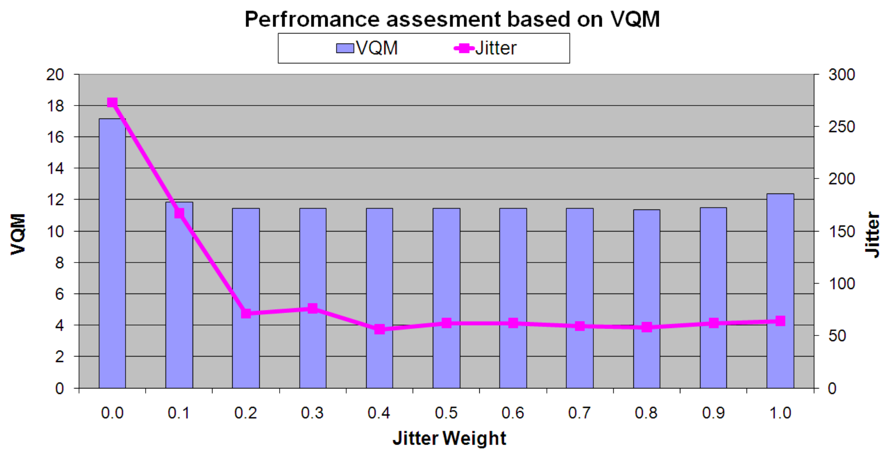

Figure 6.13: Video quality assessed using VQM

Figure 6.13 presents the video quality evaluation using VQM as the quality metric. The video quality is plotted against measured average jitter for each jitter weight value separately. For VQM the lower the score the better the quality. The results demonstrates how jitter weight 0.4 determines the best result in terms of quality throughput and jitter.

To conclude the default values for the QoS components' weights are calculated based on the results presented above as follows: $W_T = 0.15$, $W_L = 0.03$, $W_D = 0.7$ and

$W_J = 0.12$. The weight ratios were kept ti match the ones used during tests and the unitary sum requirement was respected.

6.4.2 Cost Component Assessment

Monetary cost component was evaluated using the same simulation environment and models which were previously used to assess the QoS component. Scenario 1 was considered for simulations only as the monetary cost does not vary across scenarios. Table 6.6 presents the evolutions of PSNR, throughput and monetary cost when various combinations of weights were used for the QoS_{grade} and $Cost_{grade}$.

As it can be seen in Table 6.6, there is a direct proportional relationship between the monetary cost, the received throughput and the consequent video quality (as estimated by PSNR). The relationship between the two quality parameters and the weight given to $Cost_{grade}$ is inverse proportional. The higher the weight given to $Cost_{grade}$ the lower the cost and the received quality. Table 6.6 also presents the video quality assessed using VQM. The relationship between the monetary cost and video quality (VQM) is graphically presented in Fig. 6.14. It can be observed that the cost increases as higher video quality is delivered denoted by lower VQM scores [76]. The variation in cost for various weight configurations is relatively small, however if a large number of streams and longer play time is considered these will determine a significant cost benefit. The weights interval to be considered for QMS fine tuning in terms of monetary cost is outlined in Table 6.6. The weight for monetary cost component is set based on the user profiles and devices classes as described in Chapter 4. The above mentioned interval was chosen based on the values of PSNR which must not go bellow 20dB.

6.4.3 Energy Component Assessment

Energy component evaluation employs the same simulation platform and models and Scenario 1 is considered as well. This is because the energy consumption does not vary between scenarios.

Table 6.6: Monetary Cost Evaluation

$\mathbf{W_{QoS}}$	$\mathbf{W_C}$	**PSNR (dB)**	**Throughput (Mbps)**	**Cost (Euro)**	**VQM**
1.0	0.0	47	1.80	0.37	0.77
0.9	0.1	38	1.70	0.35	0.78
0.8	0.2	29	1.54	0.30	0.81
0.7	0.3	25	1.43	0.29	0.83
0.6	0.4	24	1.39	0.28	0.84
0.5	0.5	21	1.28	0.24	0.86
0.4	0.6	18	1.17	0.20	0.89
0.3	0.7	15	1.02	0.17	0.93
0.2	0.8	11	0.83	0.14	1.00
0.1	0.9	9	0.71	0.13	1.07
0.0	1.0	5	0.34	0.05	1.45

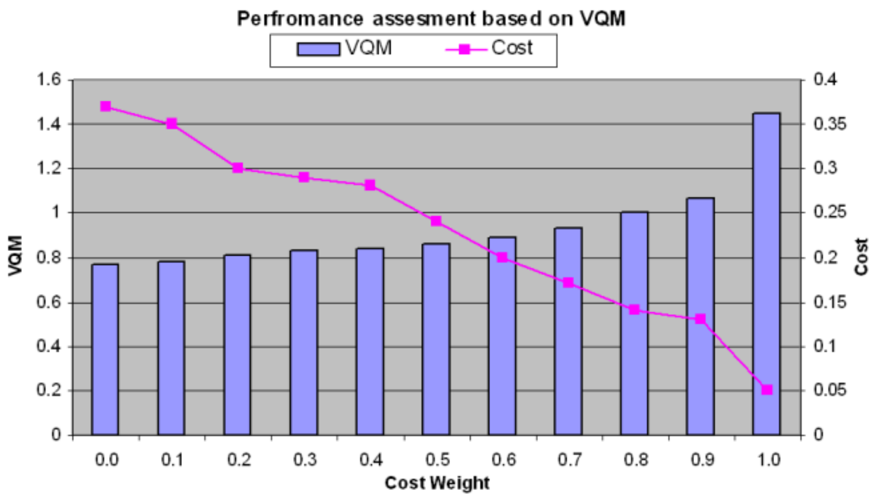

Figure 6.14: Cost vs. Quality assessment

The simulation experiments performed aim at determining the relationship between energy component weight and the achieved video quality and energy saving.

Table 6.7 presents PSNR, throughput and energy consumption when various combinations of weights were used for the QoS_{grade} and $PEff_{grade}$. An inverse proportional

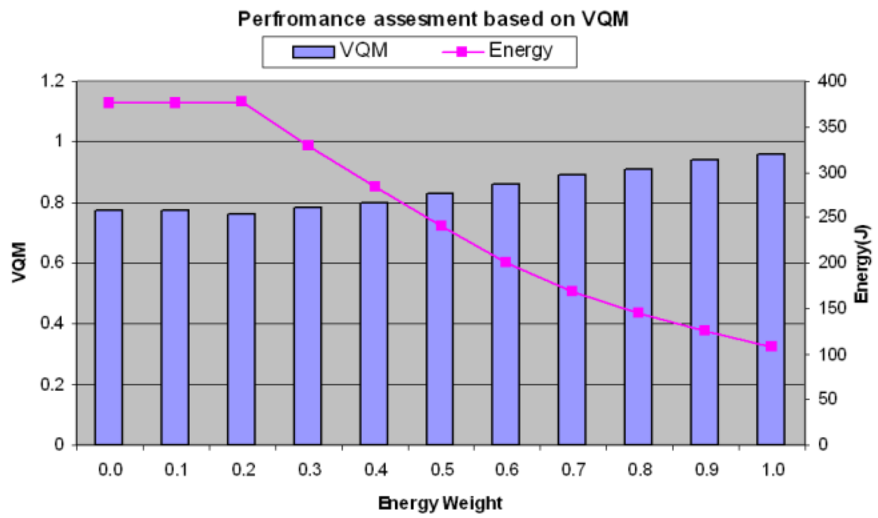

Figure 6.15: Energy vs. Quality assessment

relationship between energy consumption and received throughput as well as PSNR-based estimated video quality can be observed. However a slight degradation in video quality can be observed while a high improvement in terms of energy saving is achieved. More exactly a 71% improvement in energy consumption is achieved with only 24% decrease in video quality as estimated by VQM.

The weights interval to be considered for QMS fine tuning in terms of energy is outlined in Table 6.7. The weight for energy component is set based on the user profiles and devices classes as described in Chapter 4. The above mentioned interval was chosen based on the values of PSNR which must not go bellow 20dB.

6.4.4 QoE Component Assessment

Quality of Experience or QoE denotes the user level of satisfaction when using a certain service, in this case a multimedia streaming service. User QoE in general includes various parameters and aspects, such as playback delay, video playback continuity, audio-video synchronization etc., among which one of the most important is video quality.

Although video quality is influenced by the network QoS parameters such as throughput, packet loss, packet delay and delay jitter, these QoS parameters cannot be used to straightforward estimate video quality as they poorly correlate with the human perception

Table 6.7: Energy Consumption Evaluation

$\mathbf{W_{QoS}}$	$\mathbf{W_{P}}$	**PSNR (dB)**	**Throughput (Mbps)**	**Energy (J)**	**VQM**
0.0	1.0	12	0.93	108	0.96
0.1	0.9	13	0.99	125	0.94
0.2	0.8	15	1.07	145	0.91
0.3	0.7	17	1.16	169	0.89
0.4	0.6	20	1.28	200	0.86
0.5	0.5	25	1.43	240	0.83
0.6	0.4	32	1.59	284	0.80
0.7	0.3	41	1.74	328	0.78
0.8	0.2	52	1.85	378	0.76
0.9	0.1	46	1.80	375	0.77
1.0	0.0	47	1.80	375	0.77

characteristics [69].

Various models have been proposed [186] for estimating user QoE based on the QoS parameters. In case of QMS, the QoE parameter is calculated based on the estimated video quality. For video quality estimation the stream-based PSNR metric proposed in [180] is used.

Using Scenario 5, where various traffic nodes generate and exponential on-off traffic pattern the correlation between QoS parameters and QoE is evaluated. The results are presented in Table 6.8 and represent the average throughput, average loss ratio, average delay, average jitter as well as QoE estimated by both PSNR and VQM.

It can be observed in the table that even for an average throughput of 2 Mbps which would suggest the maximum quality possible in this scenario (the streaming rate is 2 Mbps) the QoE evaluated with PSNR and VQM show some variations in quality assessment. In a similar manner variations in average throughput determines variation in the stream-based PSNR scores, however the QoE estimated by VQM remains fairly constant.

Table 6.8: QoE Evaluation

Throughput (Mbps)	Loss (%)	Delay (ms)	Jitter (ms)	PSNR (dB)	VQM
1.99	0.5	274	273	89	17.15
2.00	0.0	288	167	96	11.82
2.00	0.0	139	71	99	11.46
1.99	0.5	150	76	71	11.46
1.98	0.1	123	56	80	11.46
1.97	1.5	138	62	81	11.42
1.92	4.0	144	62	71	11.42
1.86	7.0	137	59	70	11.42
1.79	10.5	140	58	62	11.34
1.73	13.5	147	62	55	12.48
1.67	16.5	149	64	49	20.42

Consequently it can be concluded that the QoE parameter is necessary for providing the best performance and flexibility. However the impact of its weight value in relationship with other QMS parameters depends very much on the accuracy of the video quality metrics used and especially the stream-based PSNR.

6.5 Chapter Summary

This chapter evaluated QMS and its parameters in various networking scenarios in order to determine the impact of each QMS component on the overall behavior of QMS and the application which uses it. A test methodology for QMS fine tuning in the deployment stage has been presented. This methodology will help application developers to set the values of QMS component's weight parameters for best performance in the context of a particular application scenario. The evaluation has been made and the methodology proposed considering SASHA as the application which uses QMS for handover management and traffic distribution in this particular case.

Chapter 7

SASHA Testing Results and Result Analysis

7.1 Introduction

In order to test SASHA and to validate its performance as a good handover management solution, various experiments have been conducted. SASHA was thoroughly tested through both simulations and experiments performed using a prototype emulator system.

To demonstrate the performance benefits, the testing results were compared with the ones achieved when other similar solutions were employed in the same application context.

Various performance metrics were considered for performance assessment such as throughput, loss, delay, and user perceived quality. Among these one of the most important performance metric is video quality of the delivered content. Although audio content has a great impact on user perceived quality, audio requires far less bandwidth than video; consequently this performance analysis will focus on video quality more than audio and speech quality.

Video quality was measured, assessed or estimated using there different methods. First the stream-based version of PSNR [180] was used mainly in the first stages of validating the SASHA handover concept. Objective video quality metrics, including VQM and SSIM, were used to assess the quality of the delivered video stream in the later stages of simulations and prototyping. Subjective video quality assessment was also performed to verify the results obtained by objective video quality metrics.

7.2 Simulation Environment, Models and Prototype

7.2.1 Simulation Setup

NS-2 Network Simulator (v2.29) [181] was used for modeling and as simulation environment. It was enhanced with the No Ad Hoc (NOAH)[182] wireless routing agent in order to allow only direct communication between the mobile hosts (MH) and the access points (AR) or base stations (BS) only. For a more realistic behavior of the wireless communication channels the radio patch developed by Marco Fiore [183] was also included in the simulation platform.

Each node was enhanced to support several wireless interfaces, each interface being able to communicate over a different wireless channel. This was done to allow multiple simultaneous connections to be used for multimedia content delivery.

SASHA was deployed in NS-2 as an application which simulates a multimedia streaming application with a server-side and a client-side component. The application is capable of sending a constant bitrate multimedia content to mobile devices. SASHA uses QMS scores computed for each wireless link separately to determine the corresponding communication channel, which will be used to send each of the data packets.

Several simulation purpose handover models were considered to provide a comparison basis for the performance evaluation of SASHA.

For **Mobile IP** [121], the implementation distributed with NS-2 was used. The implementation and operation scenario include a Home-Agent (HA) and a Foreign-Agent (FA) which are access router(AR) nodes capable of communicating over wired and wireless links. The Mobile-Host (MH) is represented by a mobile node with only wireless communication capabilities. The MH has its home address set to the AR's address which is considered to correspond to its home network. When the MH achieves a new care-of-address it sends binding updates to it's HA to inform it about its new location. The tunneling-based algorithm is used for handover according to the Mobile IPv6 specification.

The **Mobile DCCP** [148] simulation model was developed by the author based on the existing DCCP implementation under NS-2. Each MH is capable of communications with two BSs and a simplified concept of generalized connection was implemented. The MH has several connections open however only one is active and transports data traffic. The connections are activated and deactivated according to the handover decision. The handover decision for Mobile DCCP is based on evaluating the throughput received from each BS. The optimal moment for switching the data traffic to the new network was determined through experiments.

The simulation model for **Mobile SIP** [131] was developed by implementing the RE-INVITE message, using an older version of SIP patch developed by NIST and ported to NS 2.27 [187]. The mobility support was added by the author by allowing the mobile client to send a RE-INVITE message to inform the server about the new address of the mobile host. The precise timing of the handover was determined similar to Mobile DCCP based on received throughput.

7.2.2 Prototype Emulator System

For video quality assessment a **SASHA emulator prototype** has been developed. The emulator is composed of two main modules. One module is the network component implemented on NS-2 and the other module is composed of encoding, traffic splitting and rate adaptation components implemented in C++.

The architecture of this prototype is presented in Figure 7.1. The network module is implemented in NS-2 and uses it as the platform for simulating networking scenarios. NS-2 offers the possibility to simulate network topologies with various transport protocols allowing data exchange between fixed and mobile nodes. The SASHA handover mechanism is implemented in NS in the same manner it was implemented for simulation purposes.

Above the network simulation module a special **MPEG-2 encoder** was implemented. This is a multiple description encoder. It allows a video clip to be encoded in multiple

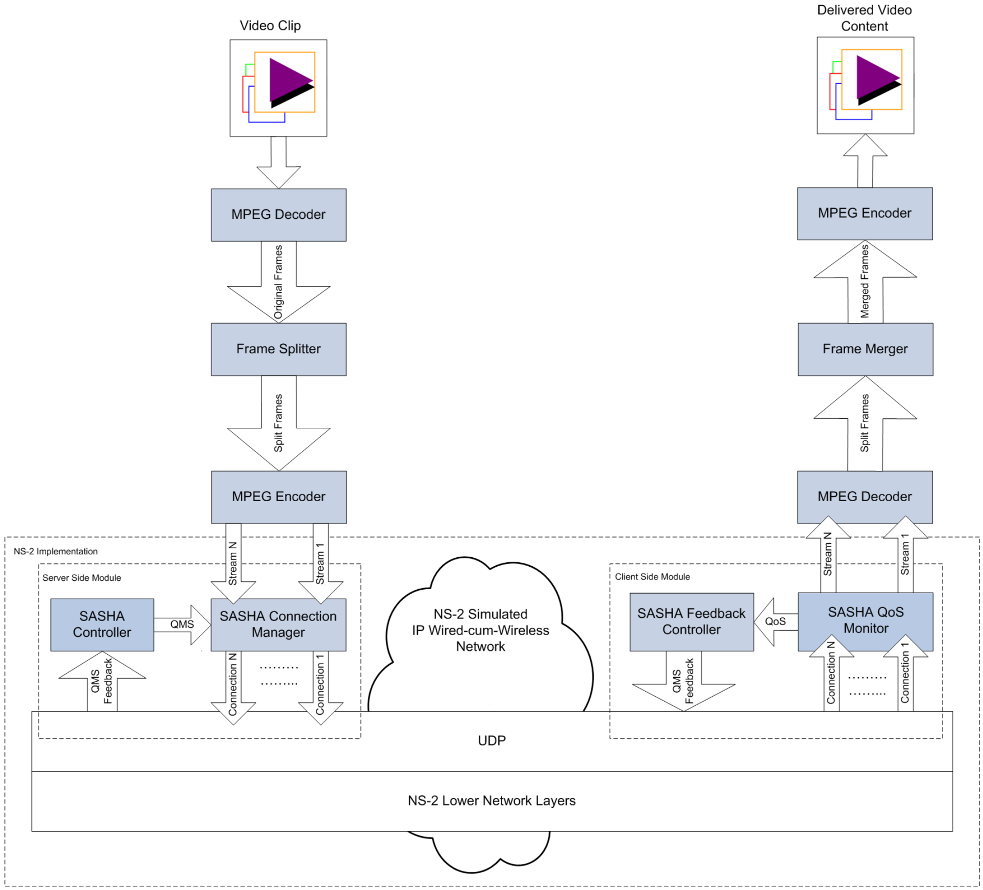

Figure 7.1: SASHA prototype emulator architecture.

descriptions or streams which can be independently sent over multiple connections.

The encoding-delivery-decoding process works as follows. The original video clip is first decoded into frames in the PPM format. These frames are red by the splitter module which splits each frame into sub-frames using and interlaced method. The interlaced method consist in distributing each line in the main frame to a separate dub-frame in a round-robin manner.

The split frames are then separately fed into the MPEG encoder producing independent sub-streams (descriptions). Each stream is encoded at the bitrate dictated by the SASHA controller according to the networking scenario. The independent descriptions are then processed in order to emulate the effect of transmission over the wireless channels. The video stream is packetized and according to the simulation results the data corresponding to the received data is copied in the sub-stream while the data corresponding to the lost packets is discarded.

The delivered sub-streams are then decoded into independent frames which are then merged in the inverse manner they were split. The merged frames are then re-encoded into the final delivered content. The resulting clips are then used for the video quality assessment.

7.3 SASHA Handover Mechanism Validation

SASHA performs handover by smoothly distributing the traffic load between existing wireless networks. This is a novel concept in handover management and its capability to offer high quality mobile data services has to be validated.

To validate SASHA's handover mechanism, a simple scenario was used as illustrated in Figure 7.2. Two networks using the same wireless technology (IEEE 802.11b) were positioned (50m between ARs) in such a way that a sufficient overlapping area was achieved.

A multimedia streaming server was used to stream a 1.5Mbps, constant bitrate (CBR), video content MH.

The MH is crossing the coverage area of the two networks including their overlapping area at a constant speed of about 5 km/h which is the average speed of a walking pedestrian. When passing through the overlapping area, the MH performs handover. One of the following solutions is used: SASHA, Mobile DCCP and Mobile IP (MIP).

The throughput received by the mobile device when handover is performed when employing one of the three solutions is presented in Figure 7.3. It can be observed in the figure that for one mobile host performing handover SASHA performs similar with Mobile DCCP and both perform much better then MIP.

The stream-based PSNR is used to estimate the video quality as received by the mobile user. The results are graphically presented in Figure 7.4. It can be seen in the figure that SASHA performs much better then MIP and slightly better then Mobile DCCP. SASHA does not experience drops in PSNR lower than 40 dB, while Mobile DCCP presents PSNR

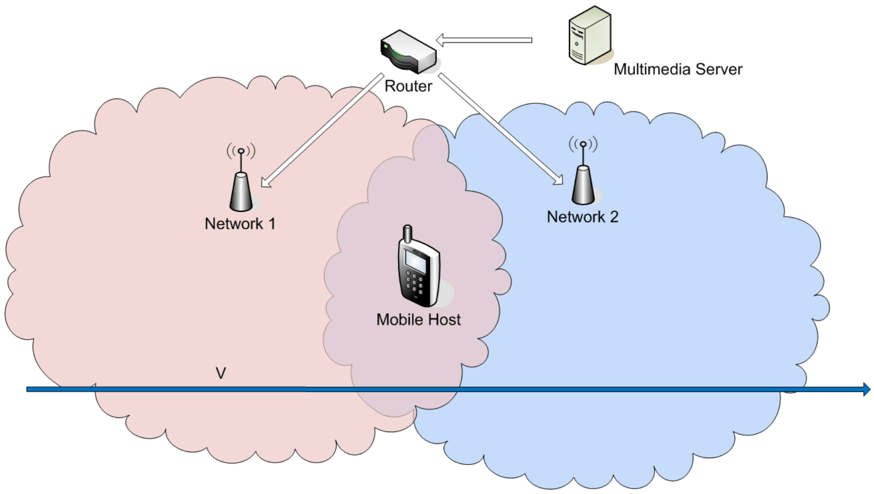

Figure 7.2: SASHA handover validation

Table 7.1: Average Throughput and PSNR for the validation scenario

Scheme	**Throughput (Mbps)**	**PSNR (dB)**
SASHA	1.50	61
Mobile DCCP	1.49	59
Mobile IP	1.27	46

dropping as low as 30dB.

The average values for throughput and PSNR measured during the handover process (between 150s and 350s) are presented in Table 7.1 It can be seen that SASHA achieves maximum throughput (1.5 Mbps) and high PSNR scores (61dB) while Mobile DCCP only achieves a throughput of 1.49Mbps and average PSNR of 59dB. Mobile IP performs much worst reaching just 1.27Mbps in terms of throughput and 46dB in PSNR scores.

In conclusion SASHA performs well in novel situations which involves one pedestrian user heading over between two networks at pedestrian walking speed.

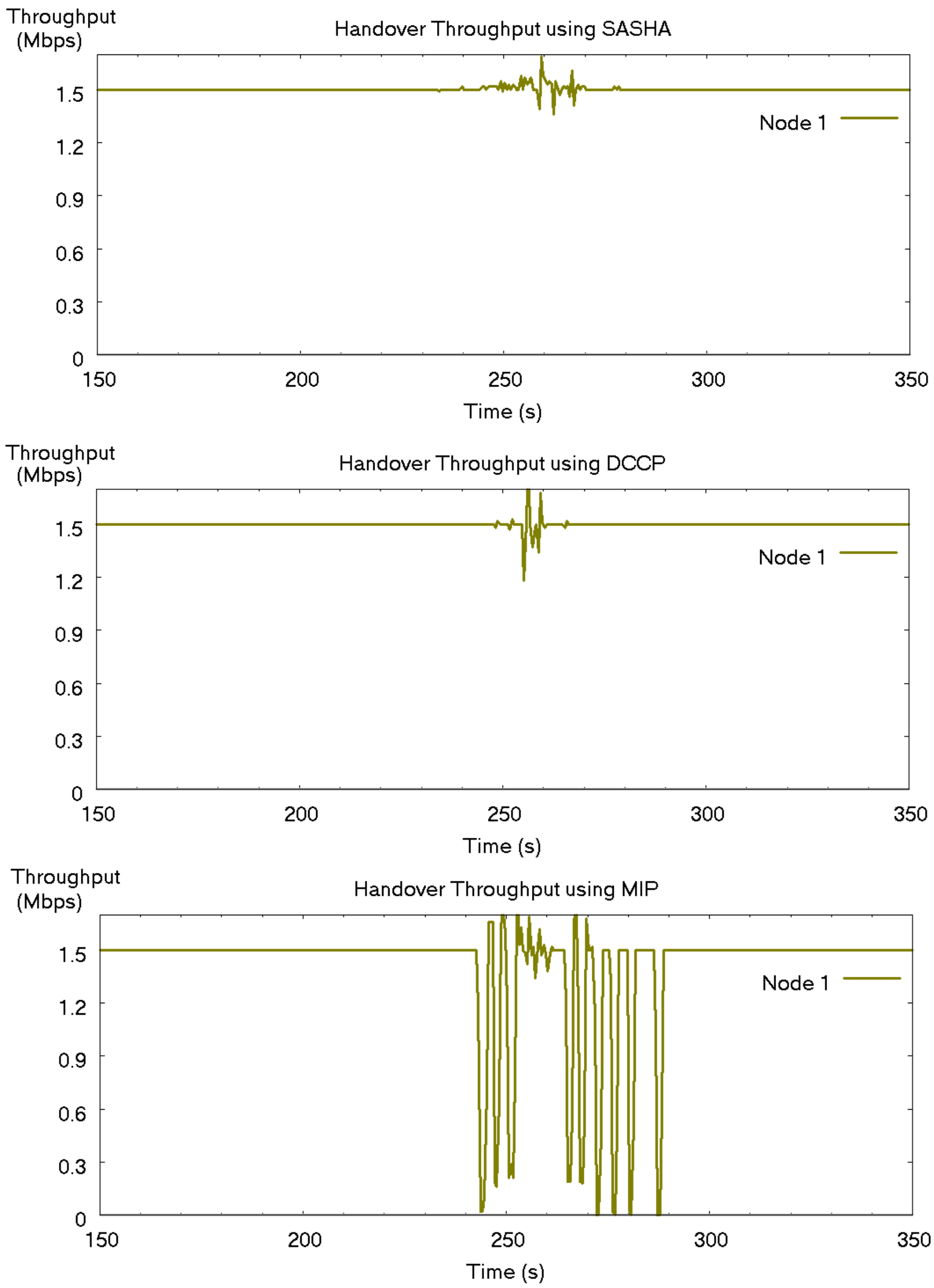

Figure 7.3: Throughput received by the mobile host in the SASHA validation simulation scenario

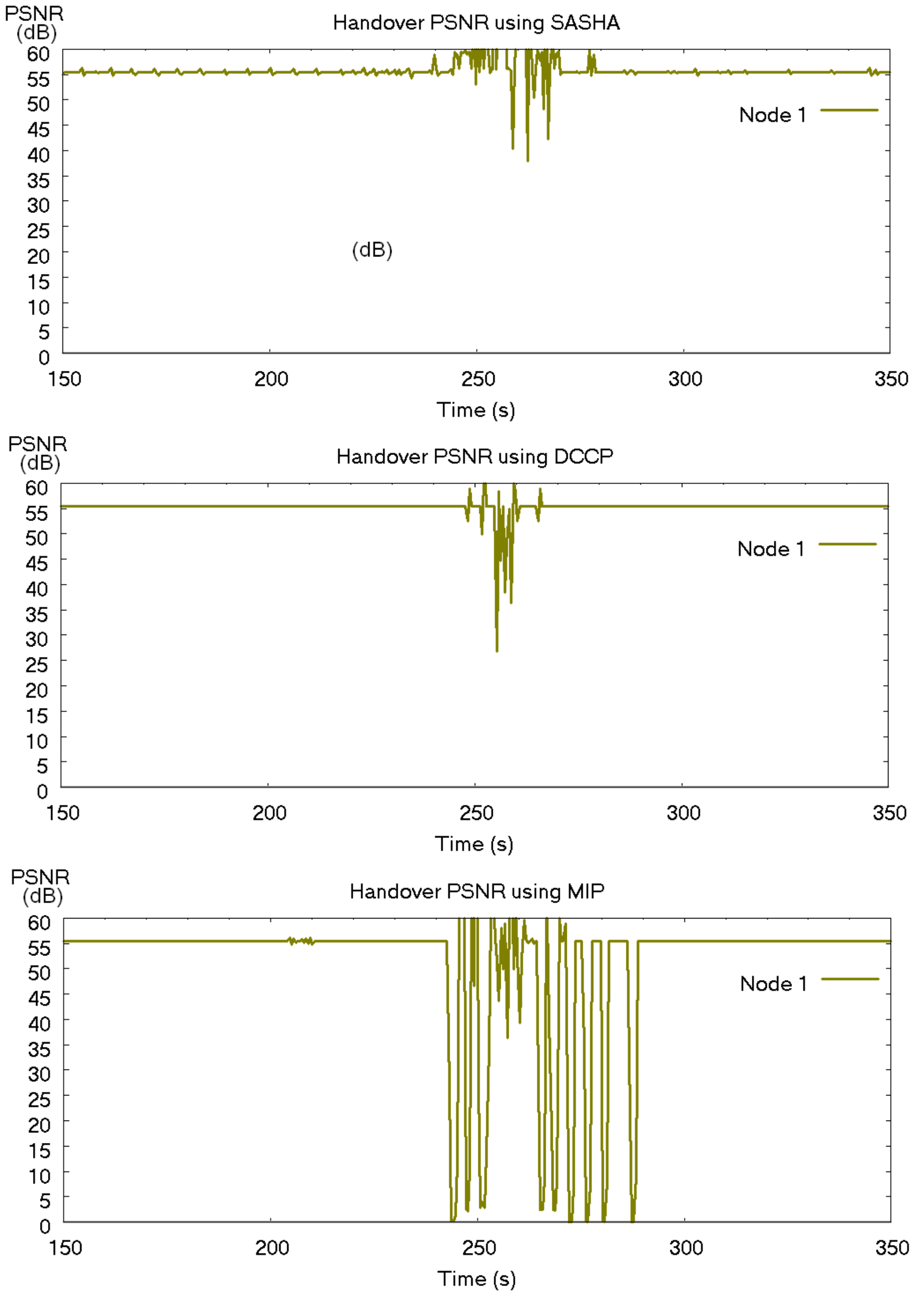

Figure 7.4: PSNR-based video quality assessment for SASHA validation

7.4 Study of SASHA in a Wireless Heterogeneous Network Environment

SASHA and the QMS metric were designed for both homogeneous and heterogeneous wireless environments converging towards an all-IP networking infrastructure. In order to demonstrate SASHA's functionality in a heterogeneous wireless environment the following testing scenario was setup.

NS-2 was used as network simulator and platform for modeling network services and applications. The network topology considered is presented in Figure 7.5. Three distinct IEEE 802.11-based wireless networks are considered, each covering a certain area.

The IEEE 802.11 MAC layer is configured as follows: 11Mbps bitrate, short retry limit is 7 and long retry limit is 4, preamble length 144 bits, slot time 20us and SIFS 10us.

The coverage areas of the three networks are overlapping. The SASHA-enabled mobile host resides within the overlapping area of these three networks.

A forth wireless network is considered which uses IEEE 802.16 (WiMAX) as the wireless communication technology. This network uses $OFDM - 64QAM3/4$ for modulation and 1/8 cyclic prefix. This network has a wider coverage area and incorporates the coverage areas of the other three networks similar to that of an umbrella.

Each network is highly congested by background mobile nodes generating video-like CBR traffic. The traffic bitrate was set according to the maximum available bandwidth measured for each network separately (5Mbps for WiFi and 8Mbps for WiMAX), when no background traffic is applied. The available bandwidth after the background traffic is applied are as follows. The three WiFi networks have available 100 kbps, 200kbps and 300kbps respectively. The WiMAX network is less congested providing 1.5Mbps available bandwidth.

The SASHA enabled multimedia server streams the video content at a constant bitrate of 2Mbps. Under the above mentioned load conditions none of the available wireless networks could provide the required 2Mbps throughput. However the aggregated bandwidth of the four networks could support the required throughput.

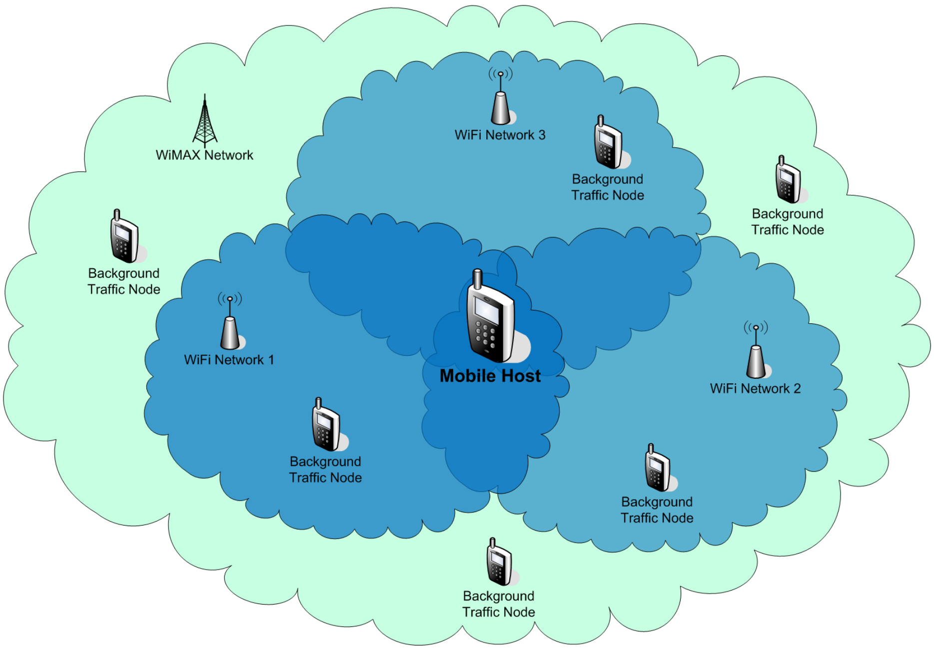

Figure 7.5: Testing scenario involving a heterogeneous networking environment

In Figure 7.6 the traffic distribution performed by SASHA in the highly loaded heterogeneous network environment described above is presented. It can be observed that SASHA by using QMS efficiently estimates the available bandwidth and distributes the traffic as follows: WiMAX conveys 1.5Mbps of the total amount of 2Mbps while WiFi 3 is allocated 0.3Mbps and WiFi 2 0.15Mbps. WiFi 1 is not used.

The efficiency of the traffic distribution is proved by the high level of aggregated throughput and PSNR scores achieved. The evolution of the throughput and PSNR scores during the 200s simulation can be seen in Figure 7.7 and Figure 7.8.

The average throughput achieved is about 1.95Mbps which is 98% of the original streaming rate of 2Mbps. The average PSNR scores reach the value of 80dB which suggest a very good quality of the delivered multimedia content.

As it can be seen in Figure 7.7 and Figure 7.8 there is a startup period when the received throughput is quite low (0.6Mbps) and increases to the maximum (2Mbps) over a period of approximately 40s. This startup period is determined by several factors: initial rate distribution (in this scenario each network was first allocated an equal share of the target throughput), the reaction speed of SASHA determined by the frequency of the feed-

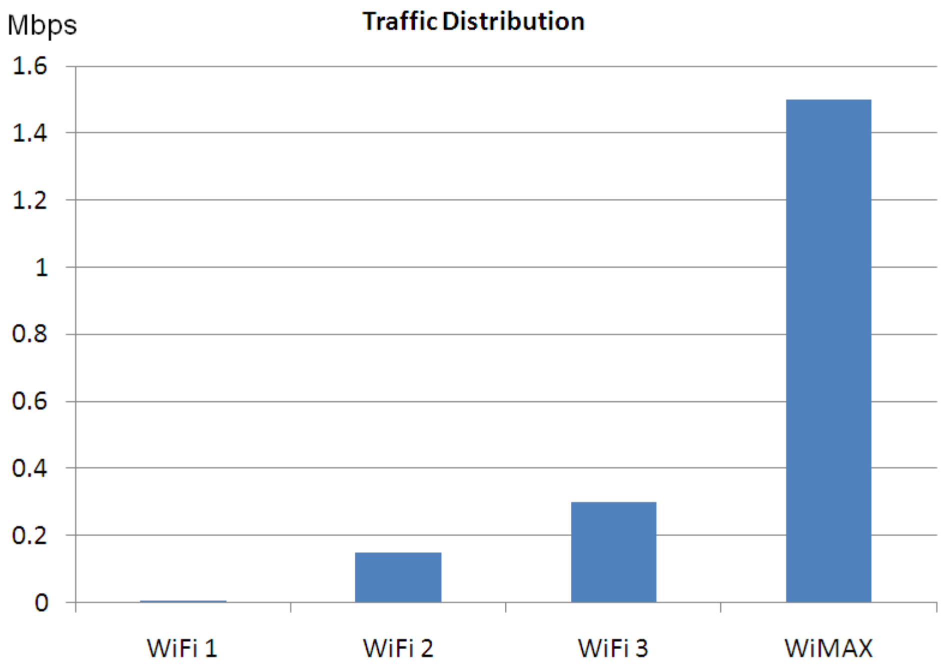

Figure 7.6: Throughput distribution performed by SASHA in the heterogeneous wireless environment.

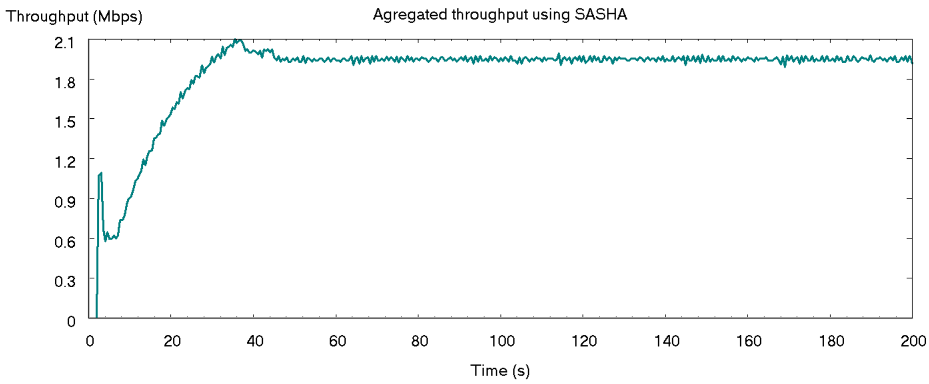

Figure 7.7: Aggregated throughput achieved by SASHA in the heterogeneous wireless environment.

back packets and the capacity of each particular network to adapt and provide the required throughput.

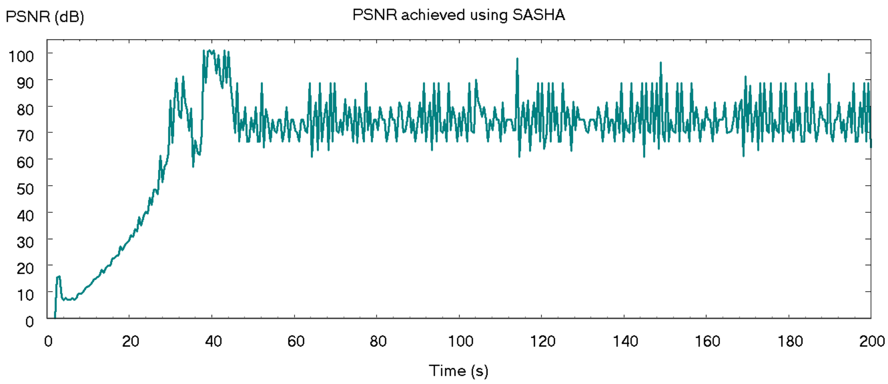

Figure 7.8: PSNR scores achieved by SASHA in the heterogeneous wireless environment.

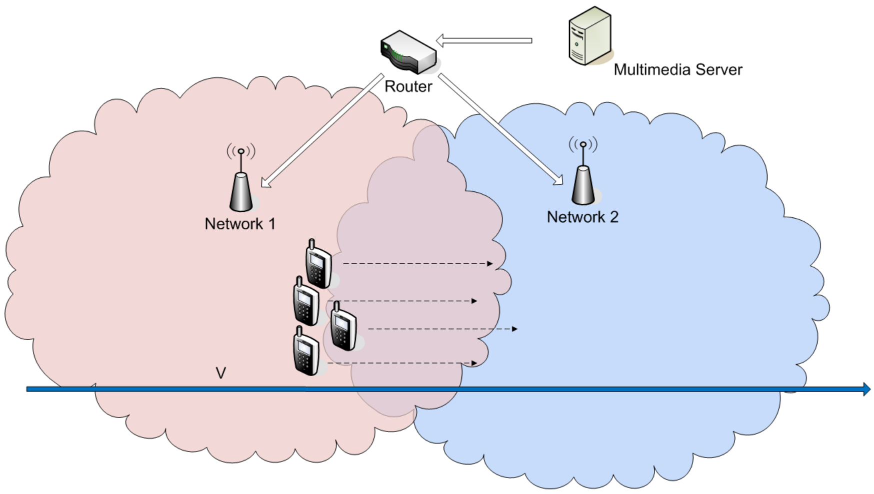

Figure 7.9: Multiple mobile nodes performing handover simultaneously

7.5 Study When Multiple Nodes Perform Handover Simultaneously

Multiple mobile nodes performing handover simultaneously determine a rapid loading of the new network which leads to significant quality degradation both in terms of QoS and user perceived quality.

Figure 7.9 schematically presents the network topology as well as the mobility scenario of the multiple mobile hosts. Two IEEE 802.11b wireless access networks are connected to an intermediate router which is further connected to a multimedia streaming server. The two networks are positioned close enough to each other (170m) to obtain a sufficient overlapping of the coverage area (approximately 20m) which was tested by sampling the received throughput in various locations.

For simplicity both wireless access technologies are based on the IEEE 802.11 family. The IEEE 802.11 MAC layer is configured as follows: 11Mbps bitrate, short retry limit is 7 and long retry limit is 4, preamble length 144 bits, slot time 20us and SIFS 10us.

At the multimedia application level two situations were considered. Scenario 1 which involves a reduced number of mobile nodes performing handover (up to 3 mobile nodes) the QOAS-based [184] multimedia application deployed on the server is not adapting and consequently it is streaming video content at a constant bitrate of 1.5Mbps. The bitrate of the video stream was chosen in such a way that the wireless available bandwidth is not exceeded.

In Scenario 2 the number of nodes was increased to 10 nodes crossing the overlapping area simultaneously. To avoid network overloading, the QOAS adaptation scheme is active and it varies the streaming bitrate according to network conditions. The speed of the mobile hosts was considered to be the human walking velocity of 5 kmph.

The performance of SASHA is compared with that of Mobile IP, Mobile DCCP and Mobile SIP with respect to packet loss and stream-based PSNR video quality metric. Figure 7.10 presents the packet loss rates experienced by one mobile device performing handover, expressed in Mbps. Figure 7.11 presents the PSNR scores achieved in the same scenario. It can be observed that SASHA, Mobile DCCP and Mobile SIP perform better then MIP. Mobile DCCP presents a slightly better performance then the other three approaches for one mobile node crossing the overlapping area of the two networks.

Figure 7.12 and Figure 7.13 present the loss rates encountered when the number of nodes performing handover simultaneously is increased to two and three, respectively.

Figure 7.14 and Figure 7.15 present the PSNR scores achieved when the number of nodes performing handover simultaneously is increased to two and three, respectively.

Table 7.2 present the average PSNR, Throughput and Loss for SASHA, Mobile SIP, Mobile DCCP and Mobile IP performing in simulation Scenario 1 when up to there mobile nodes perform handover simultaneously.

Table 7.2: Average PSNR, Throughput and Loss when streaming a maximum of 1.5Mbps video and performing handover using SASHA

Scheme	**Nodes**	**PSNR**		**Throughput**		**Loss**	
		(dB)	**%**	**(Mbps)**	**%**	**(Mbps)**	**%**
SASHA	1	60.90	-	1.50	100	0.0180	1.20
	2	55.20	-	1.50	100	0.0270	1.80
	3	44.90	-	1.36	90	0.1790	11.9
Mobile SIP	1	56.50	-	1.49	99	0.0038	0.25
	2	48.64	-	1.39	92	0.1116	7.44
	3	38.18	-	1.21	81	0.2838	18.9
Mobile DCCP	1	54.00	-	1.49	99	0.0000	0.00
	2	47.20	-	1.38	92	0.0800	5.30
	3	32.90	-	1.16	77	0.2910	19.4
Mobile IP	1	45.60	-	1.27	84	0.2269	15.1
	2	43.59	-	1.21	81	0.2824	18.8
	3	38.48	-	1.18	79	0.3100	20.6

In the above introduced figures and Table 7.2, it can be observed that Mobile SIP and Mobile DCCP performs very well for one and two nodes, presenting relatively low loss rates (less than 0.11Mbps), but encounter peak loss rates of around 0.4 - 0.5 Mbps (26% - 33%) for almost 5 seconds in case of Mobile DCCP and almost 10 seconds in case of Mobile SIP when three nodes are performing handover simultaneously. Mobile IP experiences frequent loss rates as high as 1.5 Mbps (100%) for short periods of time (1 - 2 seconds). The time intervals with high loss rates increase when the three nodes mobility scenario is employed.

Although Mobile SIP and Mobile DCCP outperform both Mobile IP and SASHA for mobility scenarios involving only one or two nodes, SASHA scales better outperforming both Mobile SIP and Mobile DCCP for the scenario involving three mobile nodes. As it can be seen in the figures SASHA encounters loss rates around 0.3 Mbps (20%) for periods of time no longer then 1 second.

The average PSNR score usually achieved by all mobility solutions outside the overlap-

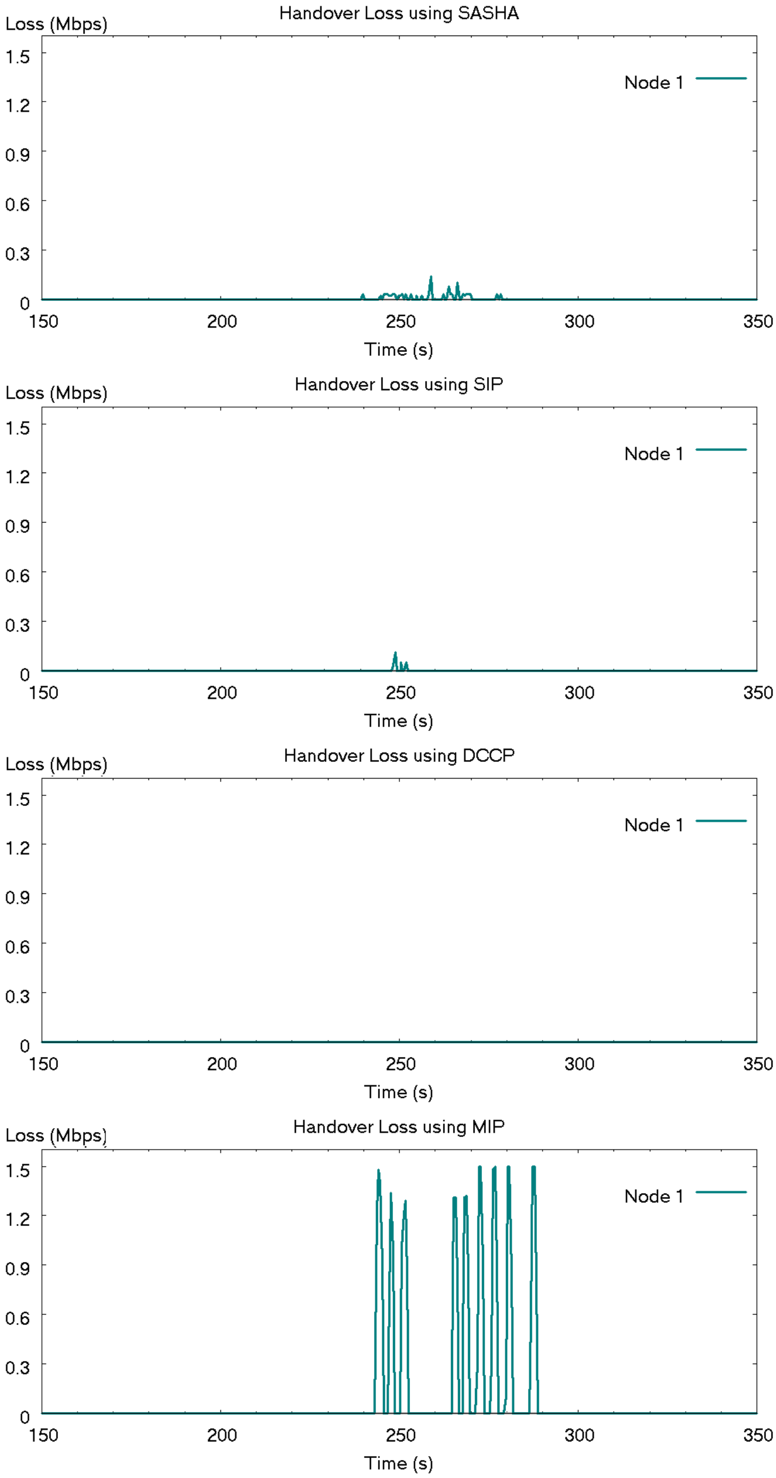

Figure 7.10: Loss rates for 1 mobile node performing handover

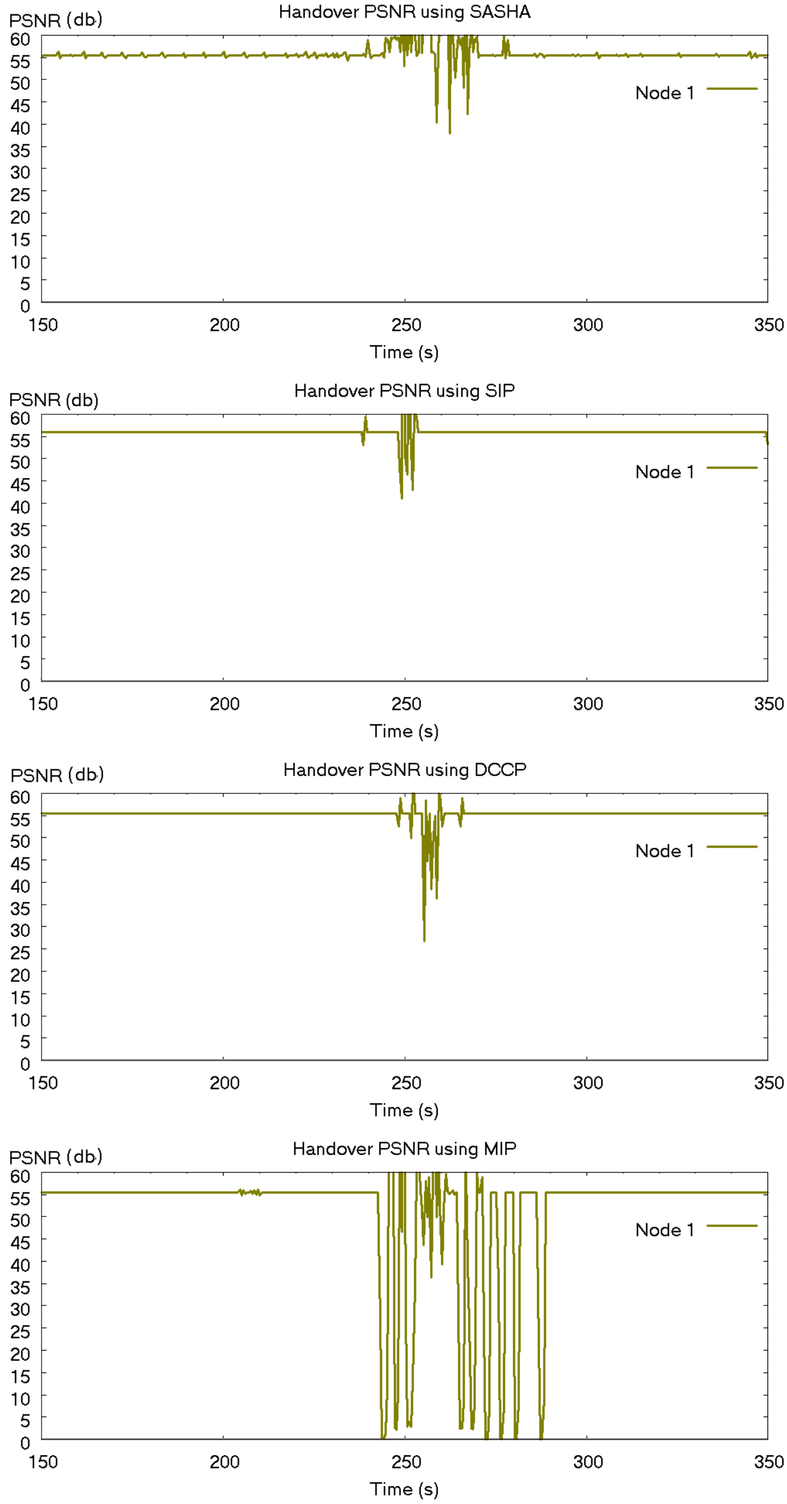

Figure 7.11: PSNR scores for 1 mobile node performing handover

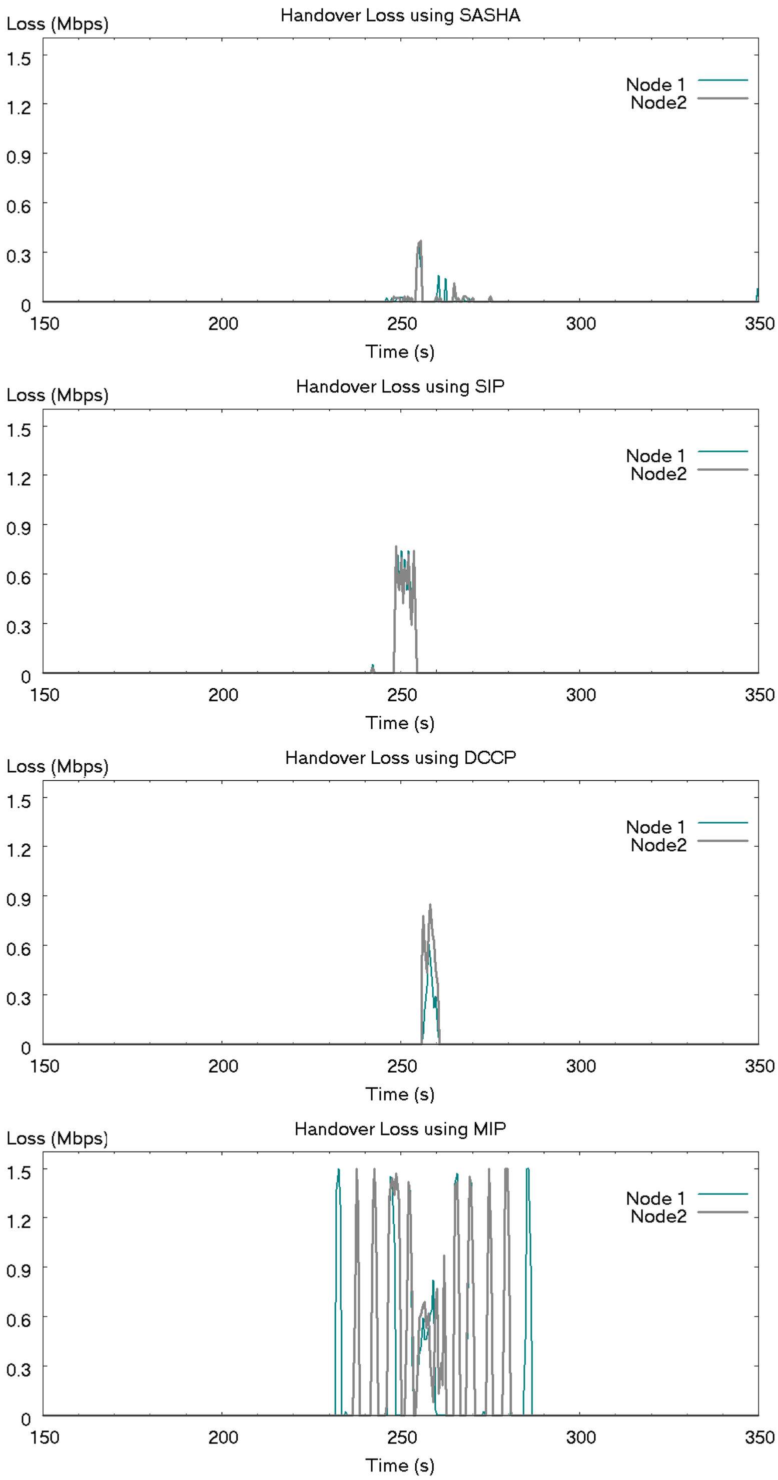

Figure 7.12: Loss rates for 2 mobile nodes performing handover simultaneously

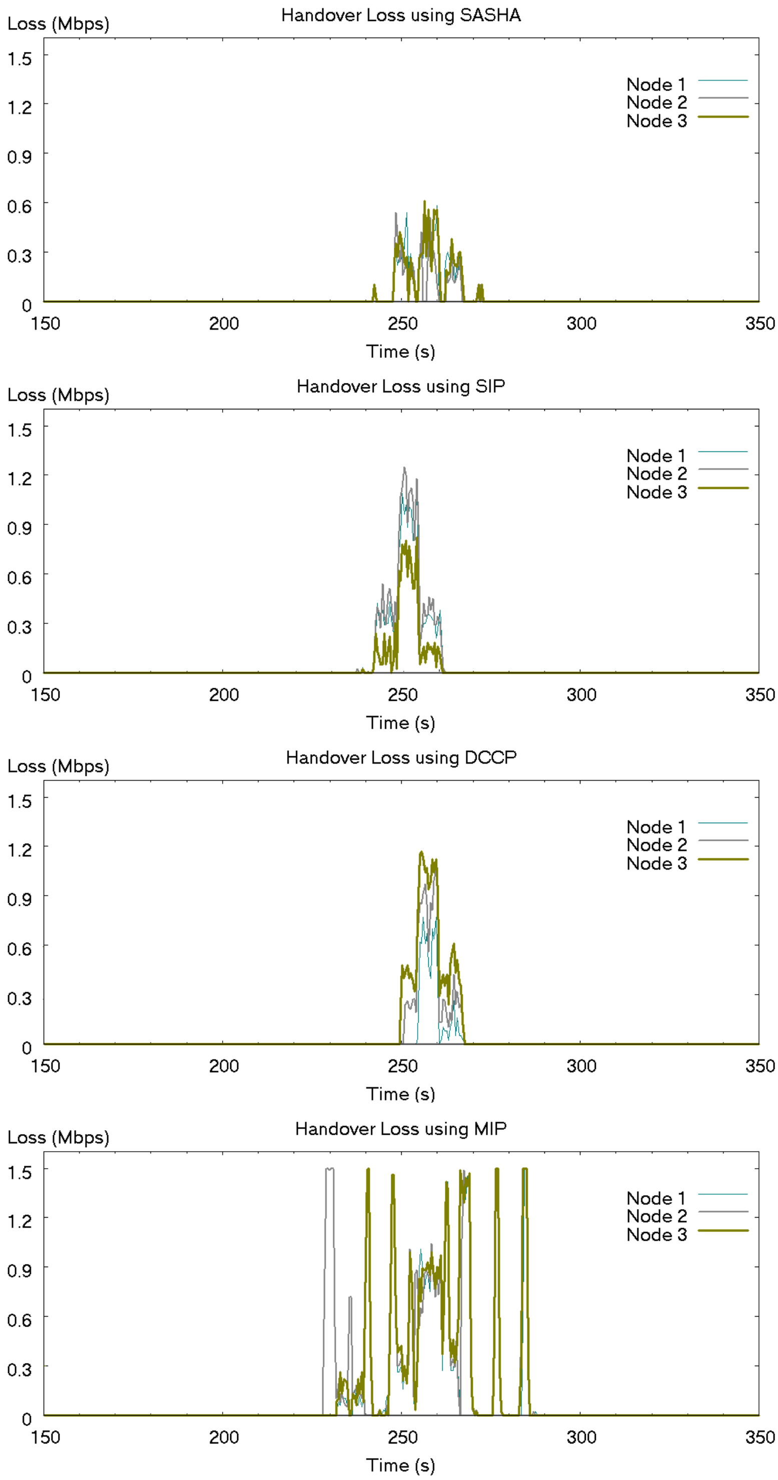

Figure 7.13: Loss rates for 3 mobile nodes performing handover simultaneously

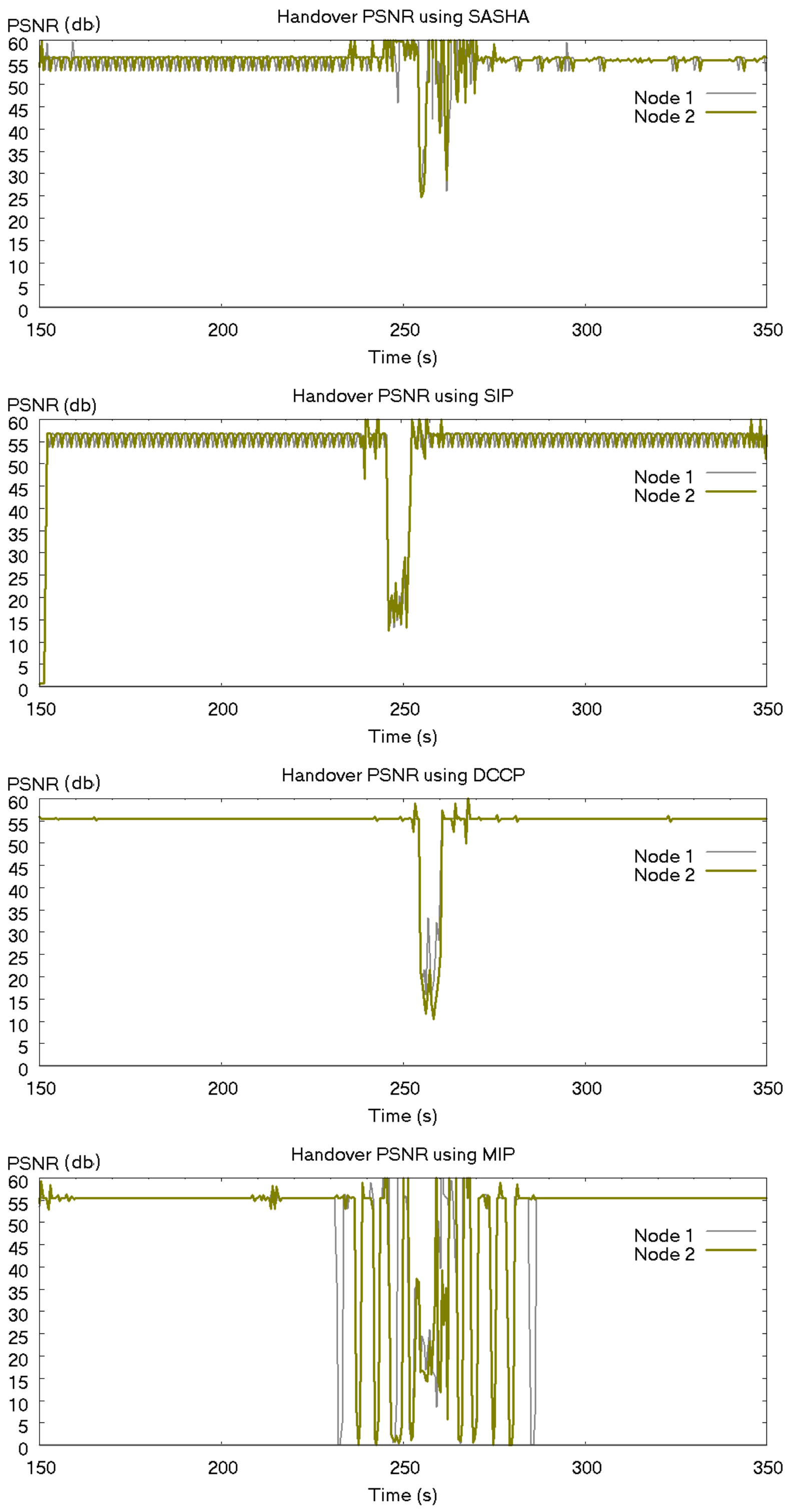

Figure 7.14: PSNR scores for 2 mobile nodes performing handover simultaneously

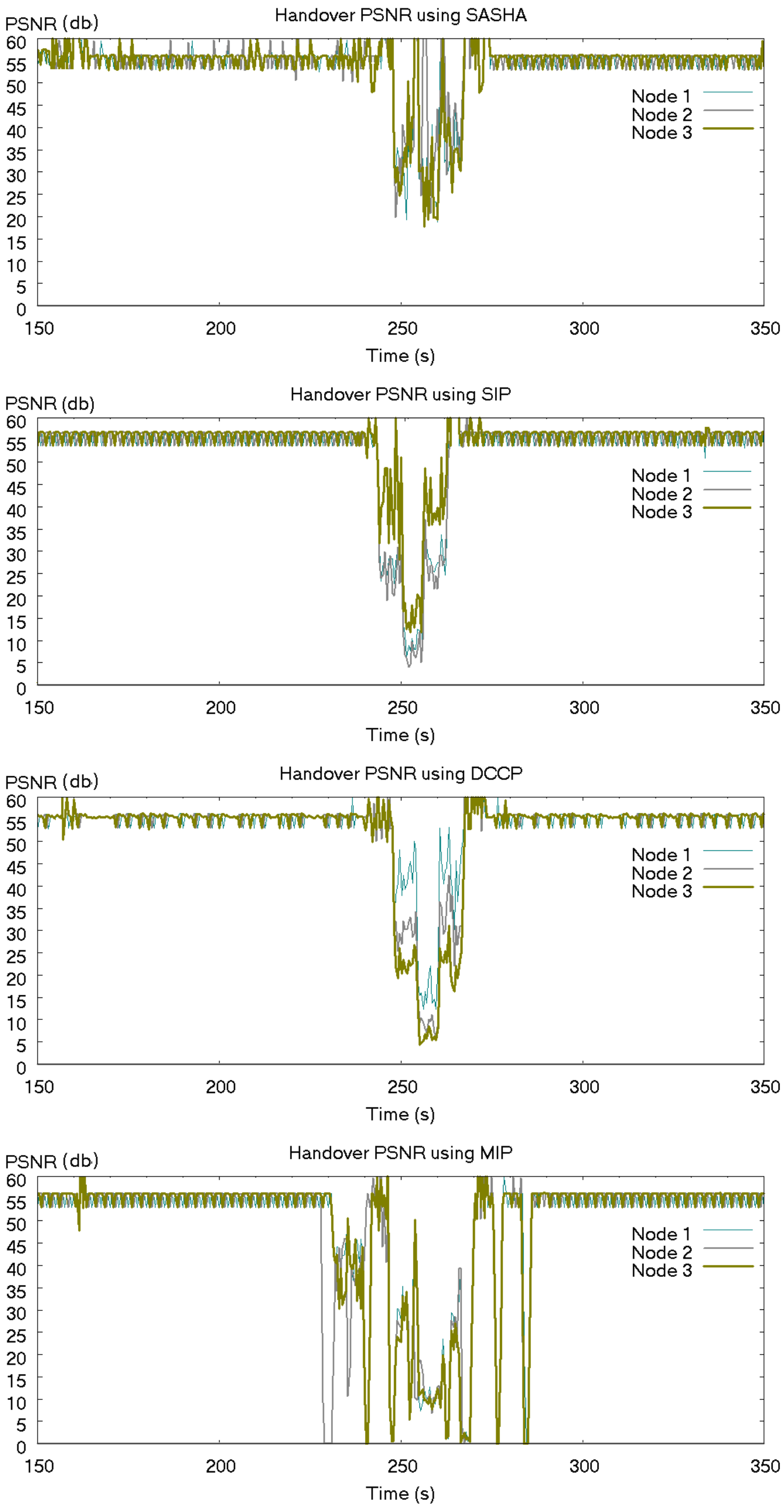

Figure 7.15: PSNR scores for 3 mobile nodes performing handover simultaneously

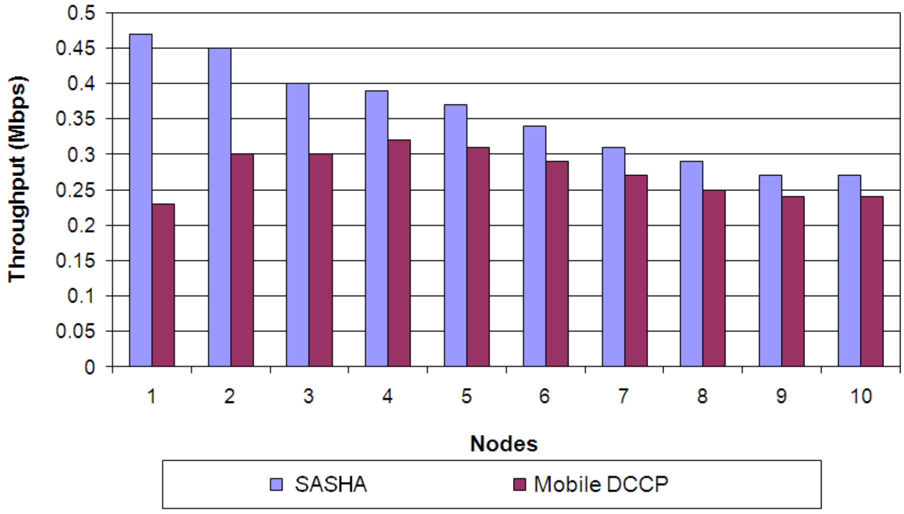

Figure 7.16: Throughputs for SASHA and Mobile DCCP

ping area is around 55dB. For example, as seen in Figure 7.14, in the two node handover scenario Mobile DCCP and Mobile SIP present PSNR scores as low as 20dB for almost 10 seconds while SASHA has its lowest score of 35dB for no more then 3 - 4 seconds.

In the case of the three node scenario as presented in Figure 7.15, Mobile DCCP and Mobile SIP experience a drop in PSNR to 10dB for almost 5 seconds and a period of 14 seconds with PSNR of around 30dB. In the same scenario SASHA presents very high PSNR scores, which drop to 30 dB for shorter periods of time only. Mobile SIP performs similar to Mobile DCCP with a slightly better scaling with the number of nodes as it can be seen in Figure 7.15.

Regarding Scenario 2, Figure 7.16 and Figure 7.17 present the performance results in terms of throughput and PSNR when the number of nodes is increased up to 10 and QOAS adaptation scheme is active.

Because Mobile DCCP, during the experiments performed employing the first scenario, proved to be the best competitor for SASHA for simplicity and more clear analysis SASHA's performance will be compared only against Mobile DCCP.

Figure 7.16 presents the average throughput received by the mobile devices while performing handover using SASHA and Mobile DCCP. As adaptive multimedia streaming

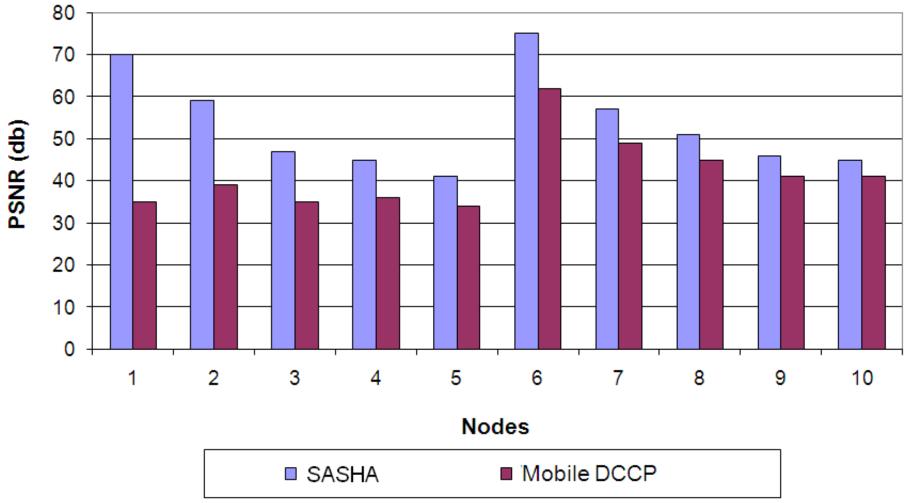

Figure 7.17: PSNR for SASHA and Mobile DCCP

application was used, the data traffic bitrate sent by the media server to each mobile node decreased as the number of mobile nodes increases.

As it can be seen in Figure 7.16, SASHA presents a very good scalability with the increase in number of nodes obtaining a 12.5% improvement for 10 mobile nodes as compared to Mobile DCCP. The throughput improvement is 19% for 5 mobile nodes and 33% in case of 3 mobile nodes simultaneously performing handover. The same performance increase can be observed in Figure 7.17 where the user quality of experience is presented (measured by PSNR) in relation with the number of mobile nodes.

The PSNR improvement is 37% for 3 mobile nodes, 20% for 5 mobile nodes and 10% for 10 mobile nodes.

Table 7.3 and Table 7.4 present the average throughput, loss and PSNR for handover performed with SASHA and Mobile DCCP respectively for one to ten mobile nodes passing simultaneously from one network to another. The packet loss percentage shows a constant evolution especially when more then six nodes perform handover which is determined by the use of QOAS multimedia adaptation scheme. The loss rates presented in Table 7.3 and Table 7.4 demonstrates SASHA's better scalability with the number of mobile nodes which can also be observed by analyzing throughput and PSNR.

Table 7.3: Average Throughput, loss and PSNR for SASHA

No. of Nodes	**Throughput (Mbps)**	**Loss (%)**	**PSNR (dB)**
1	0.47	2.0	70
2	0.45	2.1	59
3	0.40	4.7	47
4	0.39	4.8	45
5	0.37	5.1	41
6	0.34	5.5	75
7	0.31	6.0	57
8	0.29	6.4	51
9	0.27	6.8	46
10	0.27	6.8	45

Table 7.4: Average Throughput, loss and PSNR for Mobile DCCP

No. of Nodes	**Throughput (Mbps)**	**Loss (%)**	**PSNR (dB)**
1	0.23	52.0	35
2	0.30	34.7	39
3	0.30	28.5	35
4	0.32	21.9	36
5	0.31	20.5	34
6	0.29	19.4	62
7	0.27	18.1	49
8	0.25	19.3	45
9	0.24	17.2	41
10	0.24	17.2	41

7.6 Study of Variable Network Overlapping Area Size

In this section the impact of the access router position and consequently the size of the overlapping area of the networks is evaluated by measuring received throughput and estimating used QoE in networking scenarios where the overlapping area of two networks is varied.

The overlapping of the coverage areas of two or several networks depends on the range of the radio transceiver and the position of the access routers (AR). Although the position of the transceiver as well as its transmitting power can be easily controlled, the actual range and consequently the size of the coverage area very much depends on the environment and are highly dynamic. This also determines a dynamic QoS level received by the user especially in terms of throughput.

In order to determine the resilience of the proposed solution to variable network coverage area, SASHA was tested in various scenarios where the position of the ARs was changed to simulate a variable network coverage overlapping area.

Figure 7.18 presents the simulation network setup for this scenario. The ARs of the two networks are positioned at different distances, three situations being considered: 150m, 160m and 170m respectively. By analyzing the throughput received from the ARs by the mobile node in the three situations it can be concluded that the throughput is maintained almost constant at 1.5Mbps when the inter-AR distance is 150m. A small throughput drop appears when the distance is increased to 160m, with a significant throughput drop when the distance is further increased to 170m. The graphical representation of the throughput as described above can be found in Figure 7.3.

SASHA's resilience to different network overlapping areas can be observed by analyzing the evolution of loss rates and PSNR scores illustrated in Figure 7.20, Figure 7.21, Figure 7.22 and Figure 7.23, Figure 7.24, Figure 7.25 respectively.

The mobility scenario chosen involves three mobile nodes performing handover simultaneous. This number of mobile nodes was chosen in order to have enough mobile stations to determine a certain level of collisions while avoiding a highly overloaded network situations, which may not be relevant for normal network operation scenarios.

The same three solutions explored before were used for performance comparison: Mobile DCCP, Mobile SIP and Mobile IP.

Mobile IP presents the same frequent, short term (1 - 2 seconds), very high loss rates (up to 100%), with a longer period (5 seconds) of high loss (55%) when the distance between ARs is increased to 170 m. The trend of the dependency between overlapping area size and Mobile IP performance in terms of loss and PSNR cannot be clearly stated. As it can be seen in Figure 7.20, Figure 7.21 and Figure 7.22, the loss rate when there are 150m

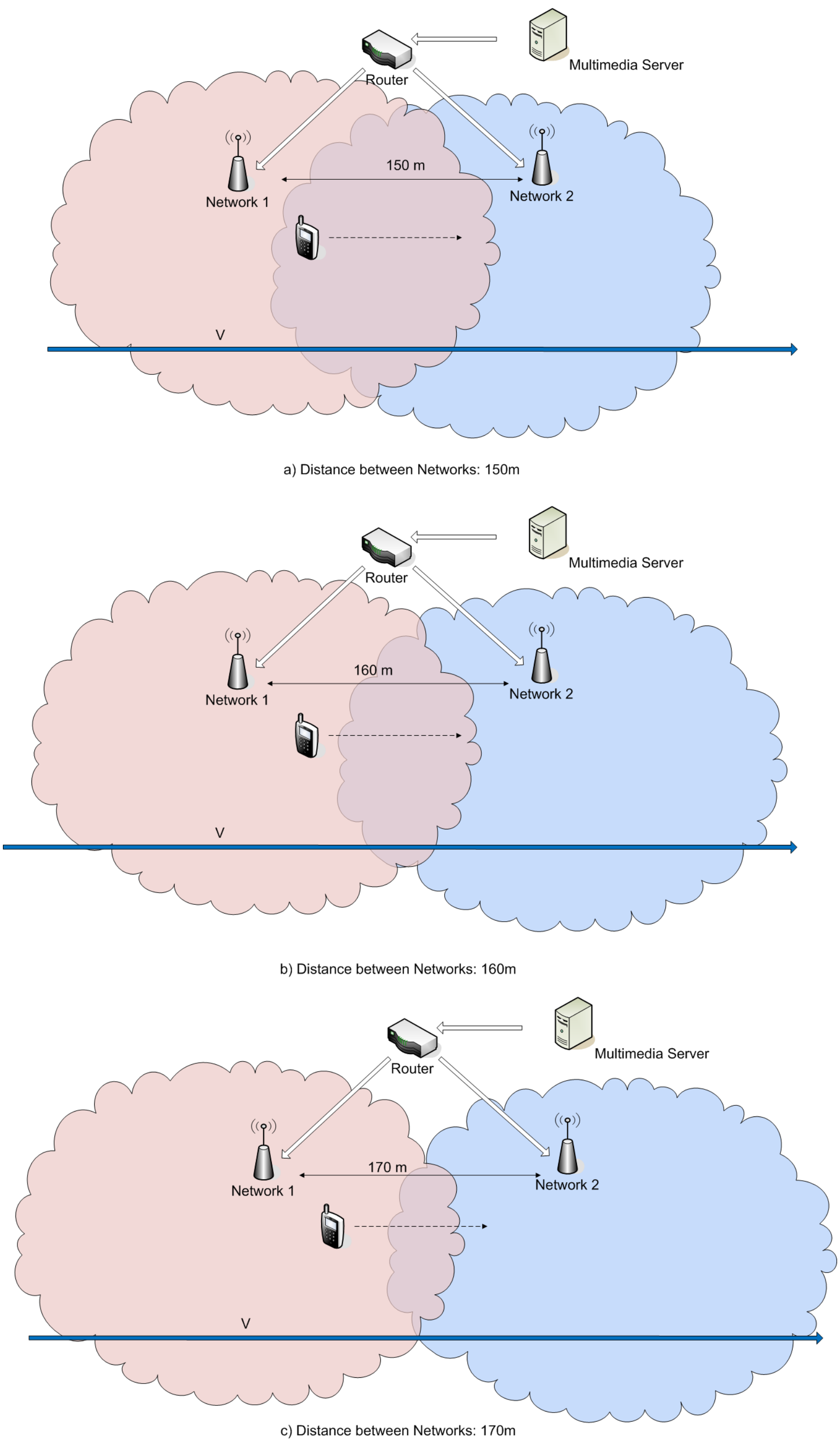

Figure 7.18: Variable network overlapping area sizes: a) 150m, b) 160m, c)170m

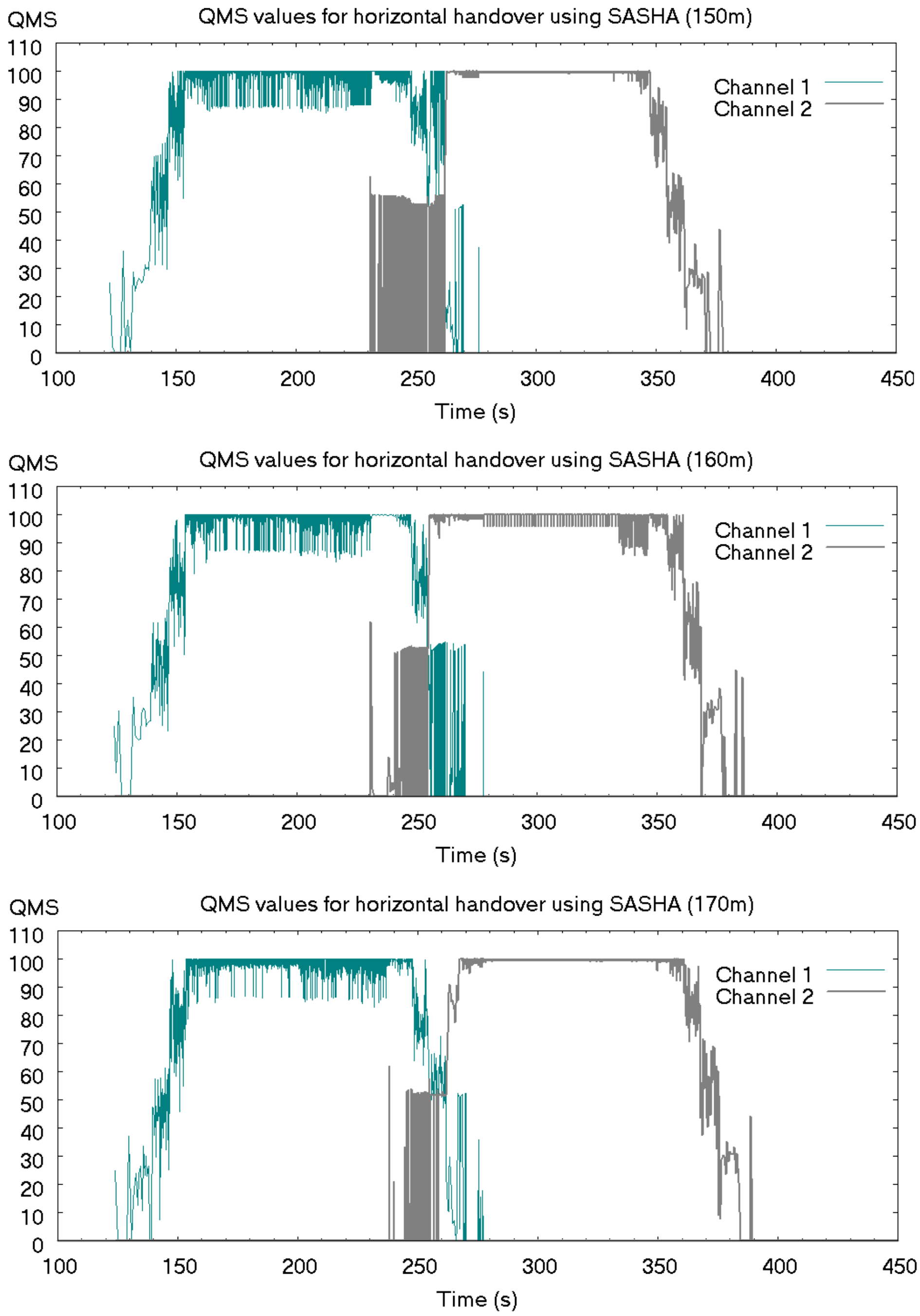

Figure 7.19: Throughput received by the MN via AR1 and AR2 with a distance between ARs of 150m, 160m and 170m, respectively

between the ARs is higher than the loss rate encountered when there are 170m between the ARs. This can be due to the random fluctuations which are meant to appear in wireless

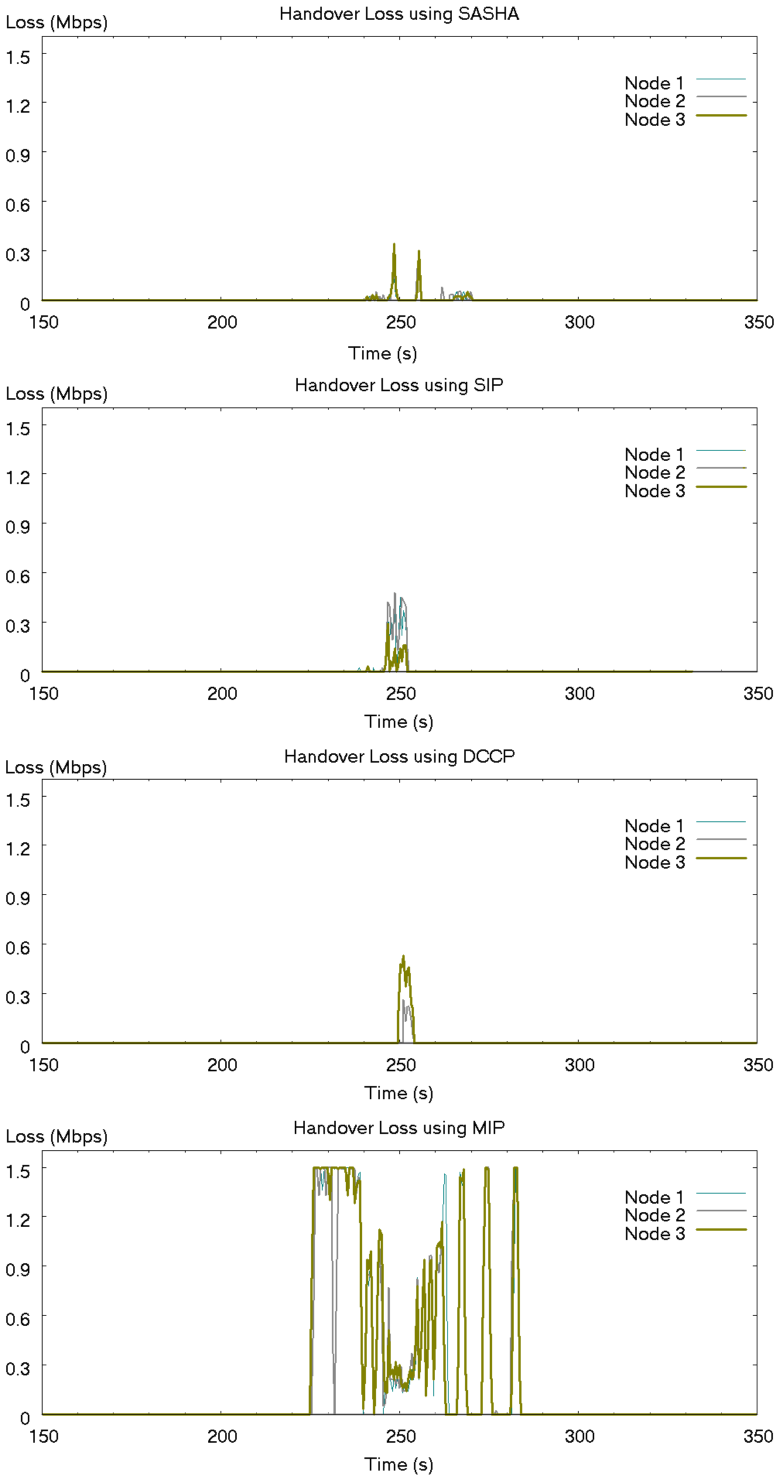

Figure 7.20: Loss rate reached when three MN perform handover from AR1 to AR2 with a distance between ARs of 150m.

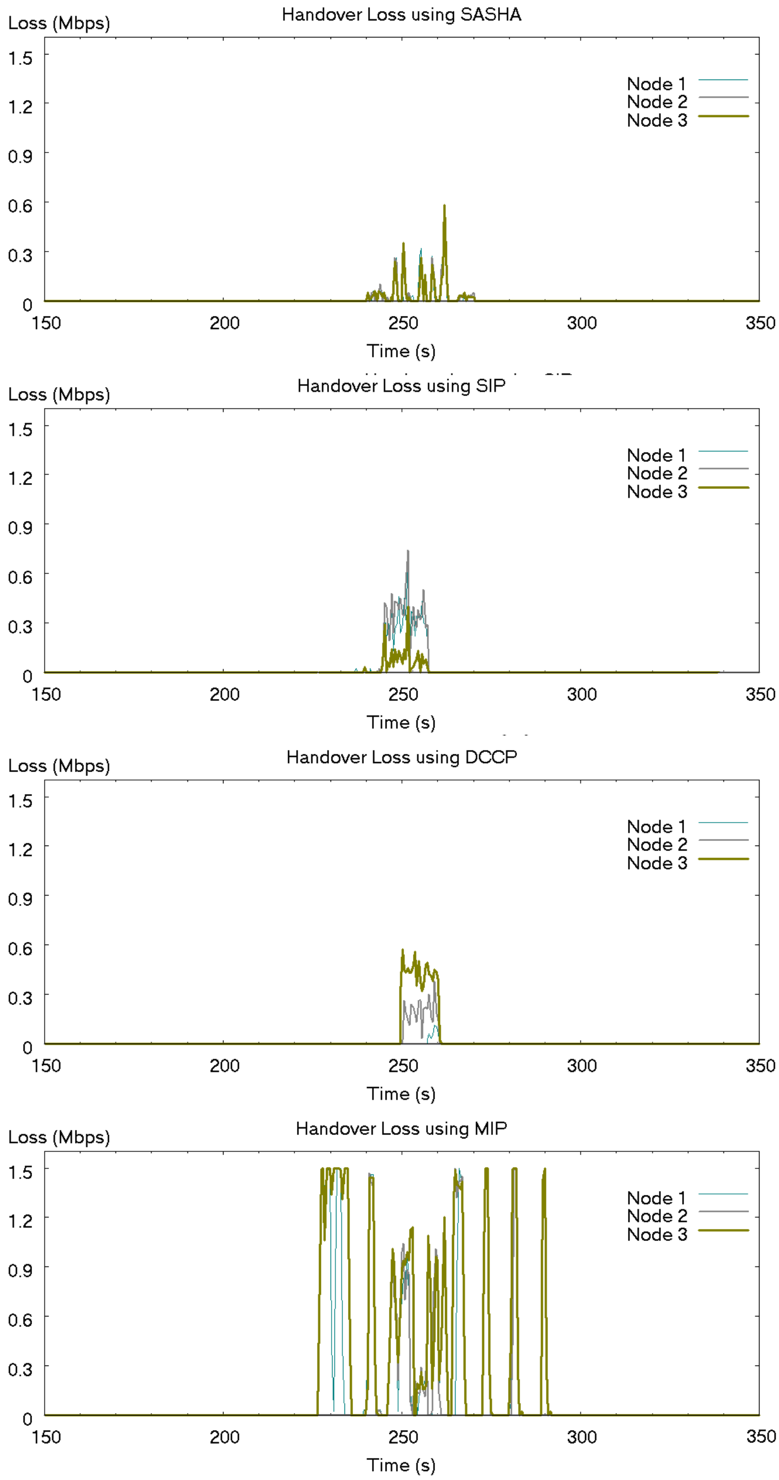

Figure 7.21: Loss rate reached when three MN perform handover from AR1 to AR2 with a distance between ARs of 160m.

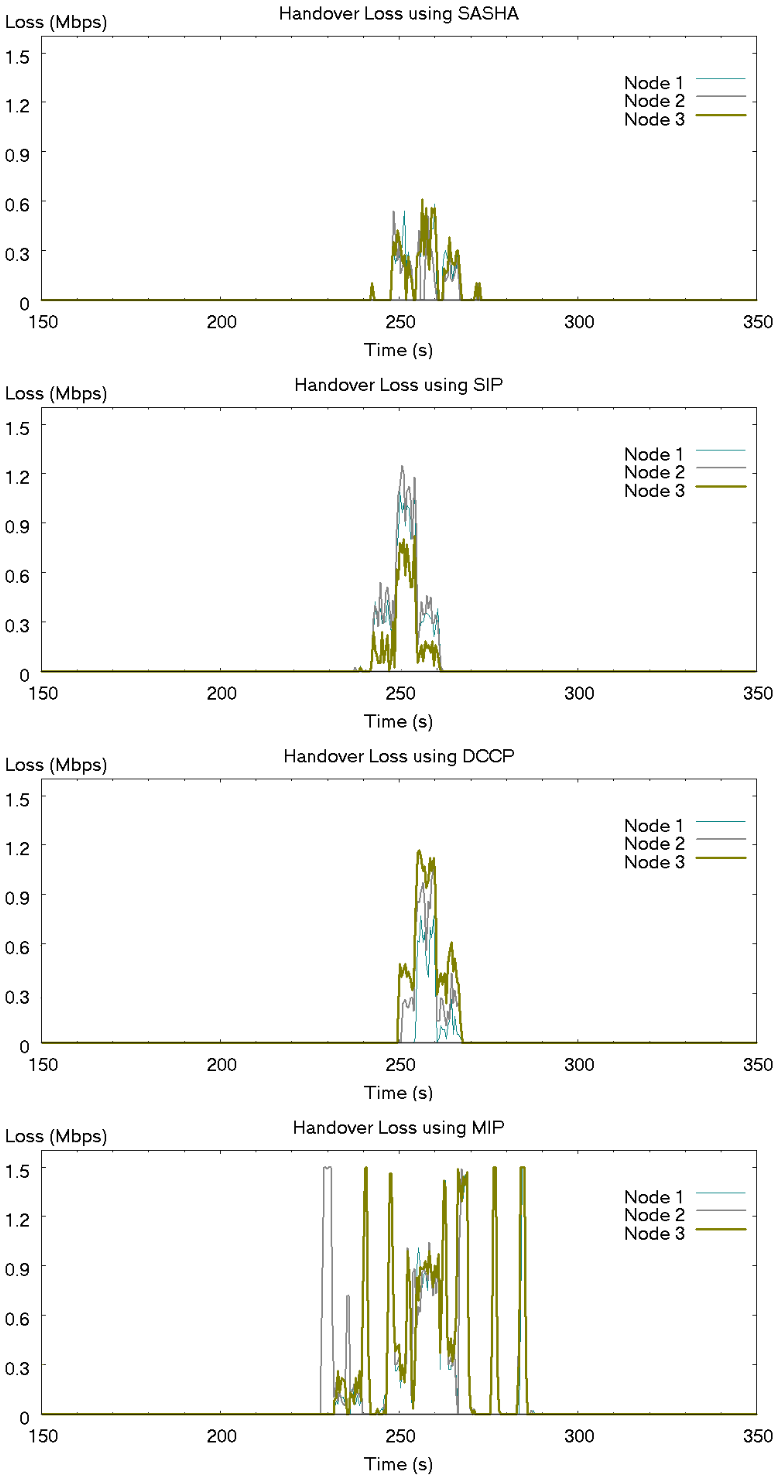

Figure 7.22: Loss rate reached when three MN perform handover from AR1 to AR2 with a distance between ARs of 170m.

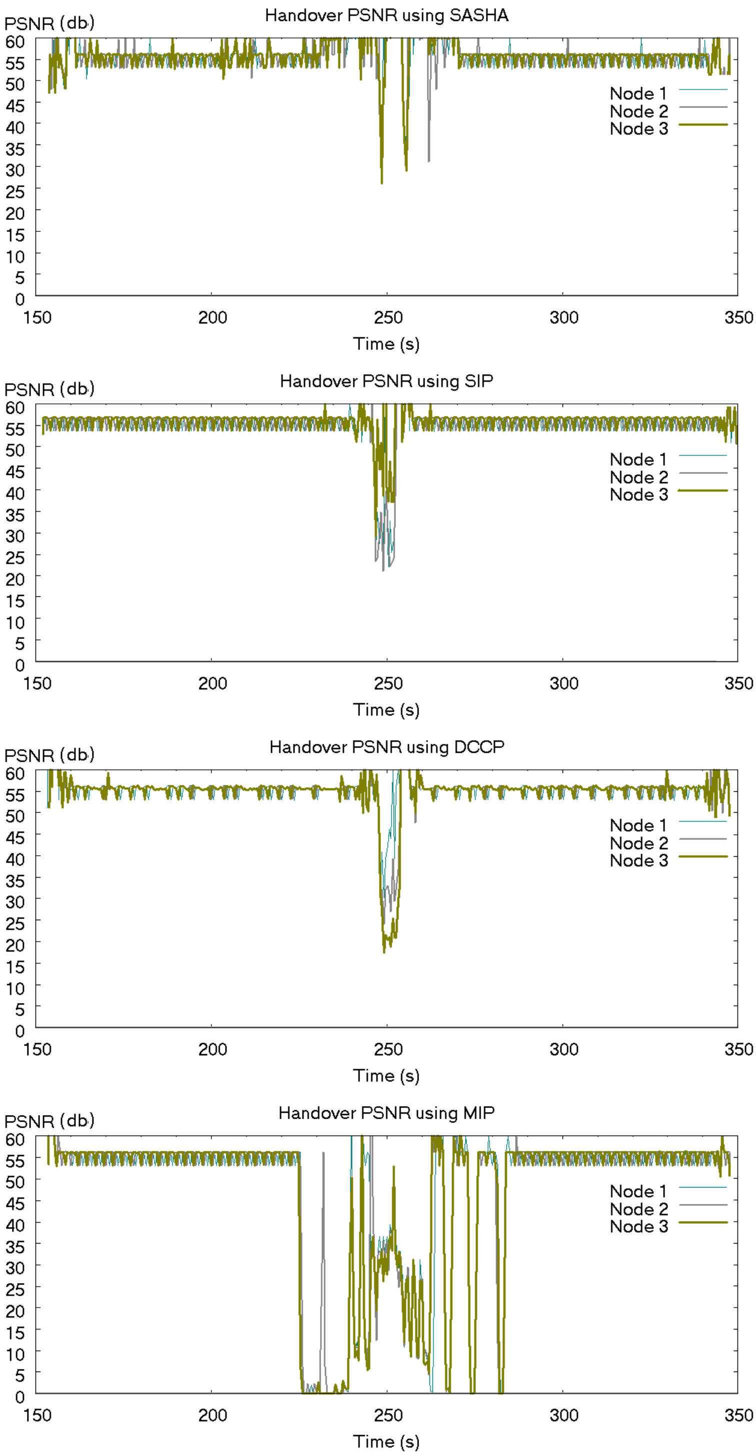

Figure 7.23: PSNR scores achieved when three MN perform handover from AR1 to AR2 with a distance between ARs of 150m.

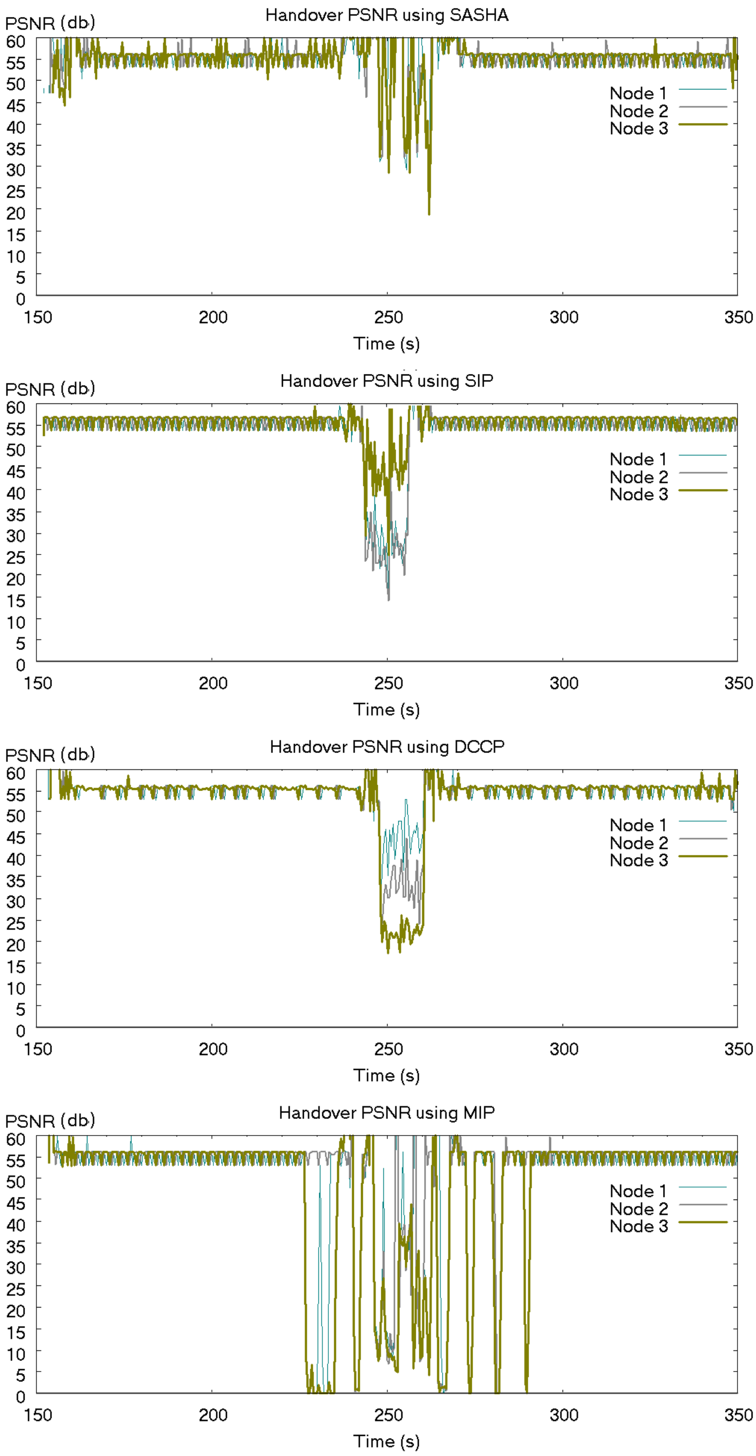

Figure 7.24: PSNR scores achieved when three MN perform handover from AR1 to AR2 with a distance between ARs of 160m.

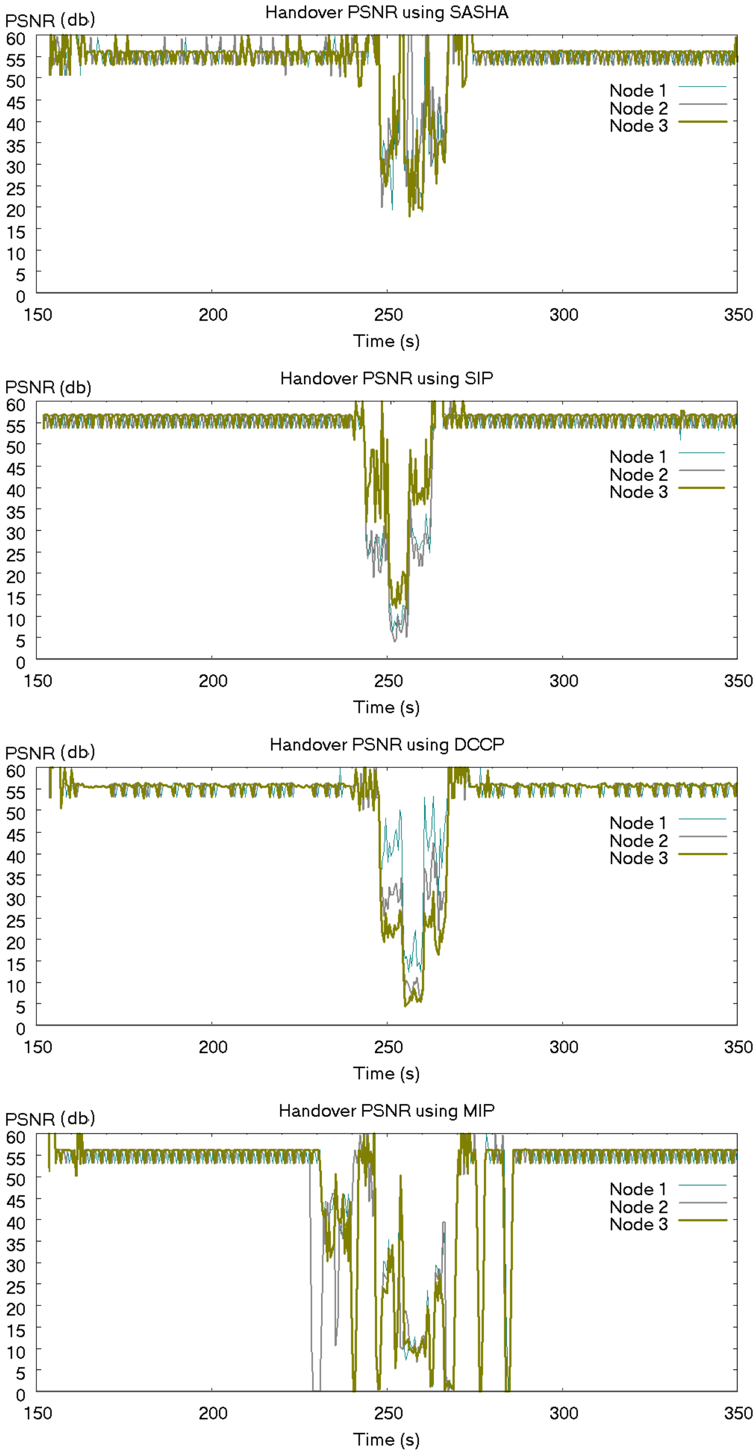

Figure 7.25: PSNR scores achieved when three MN perform handover from AR1 to AR2 with a distance between ARs of 170m.

communications and are introduced by the simulation tool.

When Mobile DCCP is employed, the loss rate is about 0.5 Mbps (33%) for around 5 seconds when the distance between ARs is 150m. When the overlapping area is decreased by increasing the distance between ARs to 160m the loss rate is still around 0.5 Mbps (33%), but the period of time this is encountered for increases to 10 seconds. By further decreasing the overlapping area by increasing the distance between ARs to 170m, a loss rate of about 33% is recorded for 12 seconds with a peak of 73% encountered for almost 5 seconds, significantly affecting user perceived quality.

The performance of Mobile SIP is similar to Mobile DCCP in terms of resilience to variable network overlapping area size. The loss rate is approximately constant when the distance between ARs is increased from 150m to 160m but the duration of the period when loss occurs is increasing. When the distance is further increased to 170m both the duration and the loss rate increase.

SASHA is more resilient to the decrease in the overlapping area determined by the increase in the distance between the two ARs. For the highest overlapping area size (150 m between the AR) the loss rate reaches 20% for 1 - 2 seconds only. When the overlapping area is decreased (160 m between the ARs) some very short term (1-2 seconds) 20% loss rates appear with a peak of 33% for about 1 second. For the smallest overlapping area (170 m between the ARs) a 13% loss rate is encountered for around 10 seconds with a peak of 26% for 4 seconds only.

In conclusion Mobile DCCP and Mobile SIP perform very well for large network overlapping areas. Mobile IP encounters short term PSNR drops with longer periods of low scores when the overlapping area is decreased.

Although a certain trend cannot be determined for Mobile IP, all three schemes affect significantly their users' perceived quality when the overlapping area decreases. In these conditions, SASHA outperforms both Mobile IP Mobile DCCP and Mobile SIP recording lower loss and higher user perceived quality.

Table 7.5: Average PSNR, Throughput and Loss when streaming a maximum of 1.5Mbps video and performing handover between networks whose ARs are 150m apart

Scheme	Nodes	PSNR		Throughput		Loss	
		(dB)	%	(Mbps)	%	(Mbps)	%
SASHA	1	64.52	-	1.50	100	0.0080	0.53
	2	56.60	-	1.49	99	0.0194	1.29
	3	62.45	-	1.49	99	0.0195	1.30
Mobile SIP	1	56.11	-	1.50	100	0.0000	0.00
	2	56.23	-	1.50	100	0.0005	0.03
	3	53.65	-	1.46	97	0.0451	3.00
Mobile DCCP	1	55.19	-	1.50	100	0.0000	0.00
	2	55.47	-	1.50	100	0.0000	0.00
	3	52.34	-	1.45	96	0.0235	1.56
Mobile IP	1	40.50	-	1.14	76	0.3499	23.3
	2	39.89	-	1.15	77	0.3470	23.1
	3	33.87	-	1.02	68	0.4674	31.1

Table 7.6: Average PSNR, Throughput and Loss when streaming a maximum of 1.5Mbps video and performing handover between networks whose ARs are 160m apart

Scheme	Nodes	PSNR		Throughput		Loss	
		(dB)	%	(Mbps)	%	(Mbps)	%
SASHA	1	61.05	-	1.50	100	0.0201	1.34
	2	61.44	-	1.50	100	0.0270	1.80
	3	58.29	-	1.49	99	0.0516	3.44
Mobile SIP	1	56.10	-	1.50	100	0.0000	0.00
	2	55.49	-	1.49	99	0.0075	0.50
	3	47.90	-	1.39	93	0.1053	7.02
Mobile DCCP	1	55.21	-	1.50	100	0.0000	0.00
	2	55.60	-	1.50	100	0.0000	0.00
	3	47.36	-	1.39	92	0.0757	5.04
Mobile IP	1	48.46	-	1.33	89	0.1611	10.7
	2	40.59	-	1.15	76	0.3451	23.0
	3	40.75	-	1.16	77	0.3274	21.8

Table 7.7: Average PSNR, Throughput and Loss when streaming a maximum of 1.5Mbps video and performing handover between networks whose ARs are 170m apart

Scheme	Nodes	PSNR		Throughput		Loss	
		(dB)	%	(Mbps)	%	(Mbps)	%
SASHA	1	60.90	-	1.50	100	0.0180	1.20
	2	55.20	-	1.50	100	0.0270	1.80
	3	44.90	-	1.36	90	0.1790	11.9
Mobile SIP	1	56.50	-	1.49	99	0.0038	0.25
	2	48.64	-	1.39	92	0.1116	7.44
	3	38.18	-	1.21	81	0.2838	18.9
Mobile DCCP	1	54.00	-	1.49	99	0.0000	0.00
	2	47.20	-	1.38	92	0.0800	5.30
	3	32.90	-	1.16	77	0.2910	19.4
Mobile IP	1	45.60	-	1.27	84	0.2269	15.1
	2	43.59	-	1.21	81	0.2824	18.8
	3	38.48	-	1.18	79	0.3100	20.6

Table 7.5, Table 7.6 and Table 7.7 present average PSNR, throughput and loss for the four mobility solutions, SASHA, Mobile SIP, Mobile DCCP and Mobile IP when used to perform handover between networks whose ARs are located at distances of 150 m, 160 m and 170 m respectively.

For completeness the impact of the number of nodes was also presented by including the results when one and two mobile nodes are performing handover simultaneously. The impact of the overlapping area size on multimedia streaming performance can be clearly observed. For example, Mobile DCCP presents 0% average loss rate for the one and two MN scenario with a 1.5% average loss rate in case of three MN's for the biggest overlapping area (Table 7.5). The average loss rate increases to 3.94% for three nodes with the decrease of overlapping area (Table 7.6). For the smallest overlapping area considered, the average loss rate increases to 19.5% for three node mobility scenario (Table 7.7).

A similar trend can be observed for Mobile SIP, the average loss rate being 3% for the largest overlapping area and 18.9% for the smallest one, when the three node mobility scenario is employed.

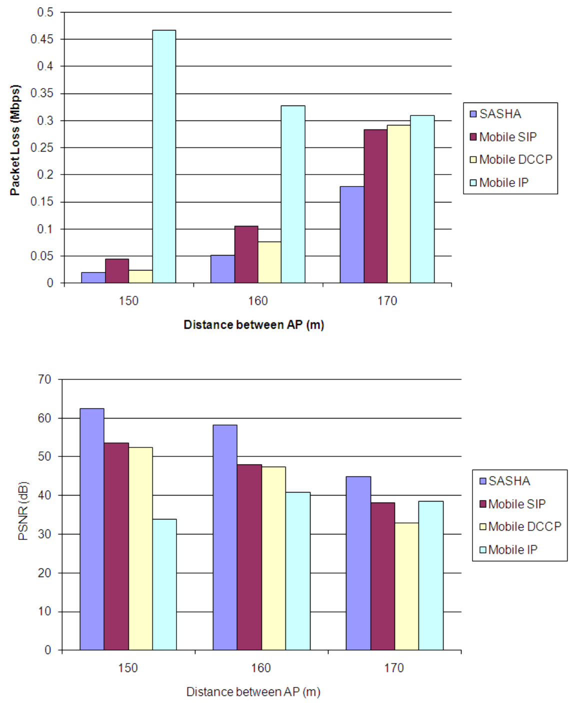

Figure 7.26: Resilience to variable overlapping area sizes evaluated in terms of PSNR and packet loss.

The performance of SASHA in terms of scalability and resilience to overlapping area variations is clearly depicted by the loss rates. SASHA presents a 1.3% average loss rate only for three mobile nodes with the largest overlapping area (Table 7.5), increasing no further than 11.9% when the overlapping area is minimal (Table 7.7).

As the same trend can be observed by analyzing the throughput and the PSNR scores presented, it can be concluded that SASHA maintains a high user perceived QoE during handover even when load increases and network overlapping area is minimal. The better performance of SASHA when the overlapping area size is variable can also be observed in Figure 7.26 when the comparison with Mobile SIP, Mobile DCCP and Mobile IP is made in terms of PSNR and packet loss.

7.7 Study the Impact of Background Traffic on Handover Performance

In a wireless environment, apart from the interferences and signal fading due to distance or environmental factors, network load is another major factor which impacts on QoS and consequently user perceived quality. The background traffic generated by other mobile stations attached to the same AR or BS reduce the amount of available bandwidth and increases the collision rates at MAC layer.

Under these circumstances, load variation determined by the dynamics of the background traffic affects the network condition and consequently may trigger a handover even for a static wireless station. In this scenario the role of handover management overlaps with the load balancing service required for maintaining an optimum distribution of wireless stations over the available access points or base stations.

The simulation scenario used to test load balancing consists of four mobile nodes deployed within the coverage area of each network. Each mobile node has a fixed position and communicates with its corresponding network generating constant bit rate (CBR) background traffic which will determine a certain level of network congestion. The background traffic is generated in stages, each stage corresponds to a particular bitrate. The traffic is constant during the period of each stage. More details about the background traffic will be provided later on in this section.

The network topology is presented in Figure 7.27. Two networks with overlapping coverage areas are loaded by several wireless stations by generating video-like CBR traffic.

The mobile host is positioned within the coverage area of the two networks and is receiving a constant 1.5Mbps video stream from the media server. Three distinct situations are considered. First SASHA is used for handover and load balancing, the second uses Mobile DCCP as a re-association handover mechanism, and the third does not have any load balancing or mobility mechanism in place. The mobile host has a fixed position in these testing scenarios as the main target is to determine the performance in the context of background traffic increase and not in the context of user mobility.

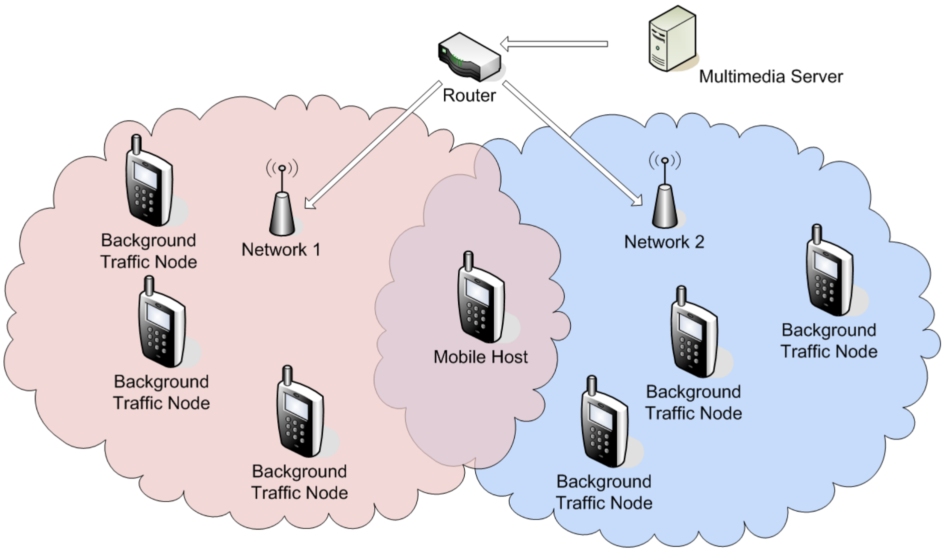

Figure 7.27: Network topology with multiple background traffic nodes

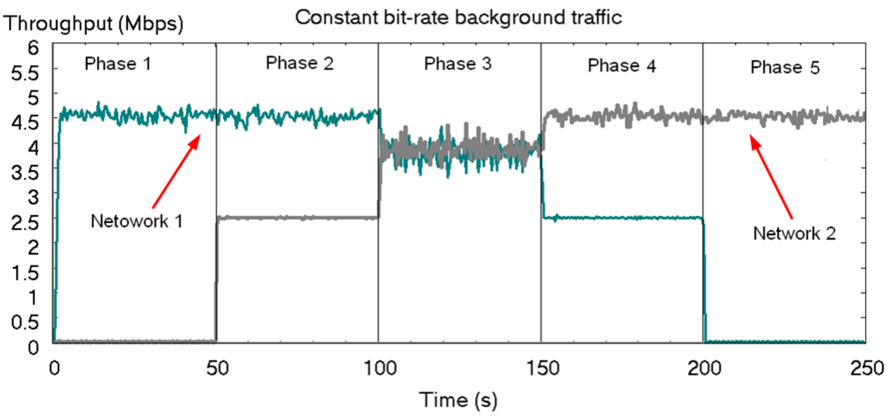

Figure 7.28: Constant bit-rate background traffic

Figure 7.28 presents the aggregated traffic generated by the background traffic nodes for each network separately. As it can be observed in the figure the traffic was generated in such a manner that the networks are alternatively increasingly loaded with a period when both networks encounter high load. The pattern of the background traffic aims at determining the capacity of SASHA and QMS, on which it relies on, to detect network load and efficiently distribute the traffic in order to maximize the overall QoS level and user perceived quality.

The performance of the load balancing technique based on SASHA was evaluated using throughput and loss as QoS metrics and PSNR as a user perceived quality metric. The simulation-based testing results for SASHA load balancing (SASHA), re-association load balancing (Mobile DCCP) and the situation when no load balancing is used (NO LB) were considered.

Each AR uses IEEE 802.11b as the wireless technology with the data rate set to 11Mbps. However due to distance and environmental factors a realistic total available bandwidth is approximately 5Mbps. Consequently the background traffic was set accordingly.

As presented in Figure 7.28 background traffic was generated in five different stages over a total period of 250 seconds.

In the first stage (0s to 50s), Network 1 is overloaded (4.5 Mbps background traffic) while Network 2 is totally unloaded (0 Mbps background traffic). In the second stage (50s to 100s) Network 1 is still overloaded (4.5 Mbps background traffic) while Network 2 lightly loaded (2.5 Mbps background traffic). In stage three (100s to 150s) both networks are fully loaded (3.5 Mbps background traffic). Stage four and five lasting from 150s to 200s and 200s to 250s respectively present the same pattern of background traffic, but in this case Network 1 becomes overloaded while Network 2 is first lightly loaded and then totally unloaded.

The testing results are presented in Figure 7.29, Figure 7.30 and Figure 7.31.

As illustrated in Figures 7.29 and 7.30, SASHA performs well in all these situations presenting insignificant loss rates excepting the period when both APs are fully loaded (100s - 150s) when a very low 0.08 Mbps loss rate is encountered (5.3%). From the point of view of PSNR, Figure 7.31 the average values are constantly above 51 dB with a maximum of 70 dB which is excellent.

The re-association-based load balancing (Mobile DCCP) performs well when at least one network is lightly loaded or has no load. As it can be seen in Figure 7.30, the throughput is maximum (1.5Mbps) for all network load situations except the one when both networks are fully loaded (100s - 150s) when the throughput drops to 1 Mbps. The same behavior can be observed by analyzing loss and PSNR in Figure 7.31.

When the no load balancing technique is used, QoS drops dramatically when the network is overloaded as the mobile device remains associated with same AR. When the back-

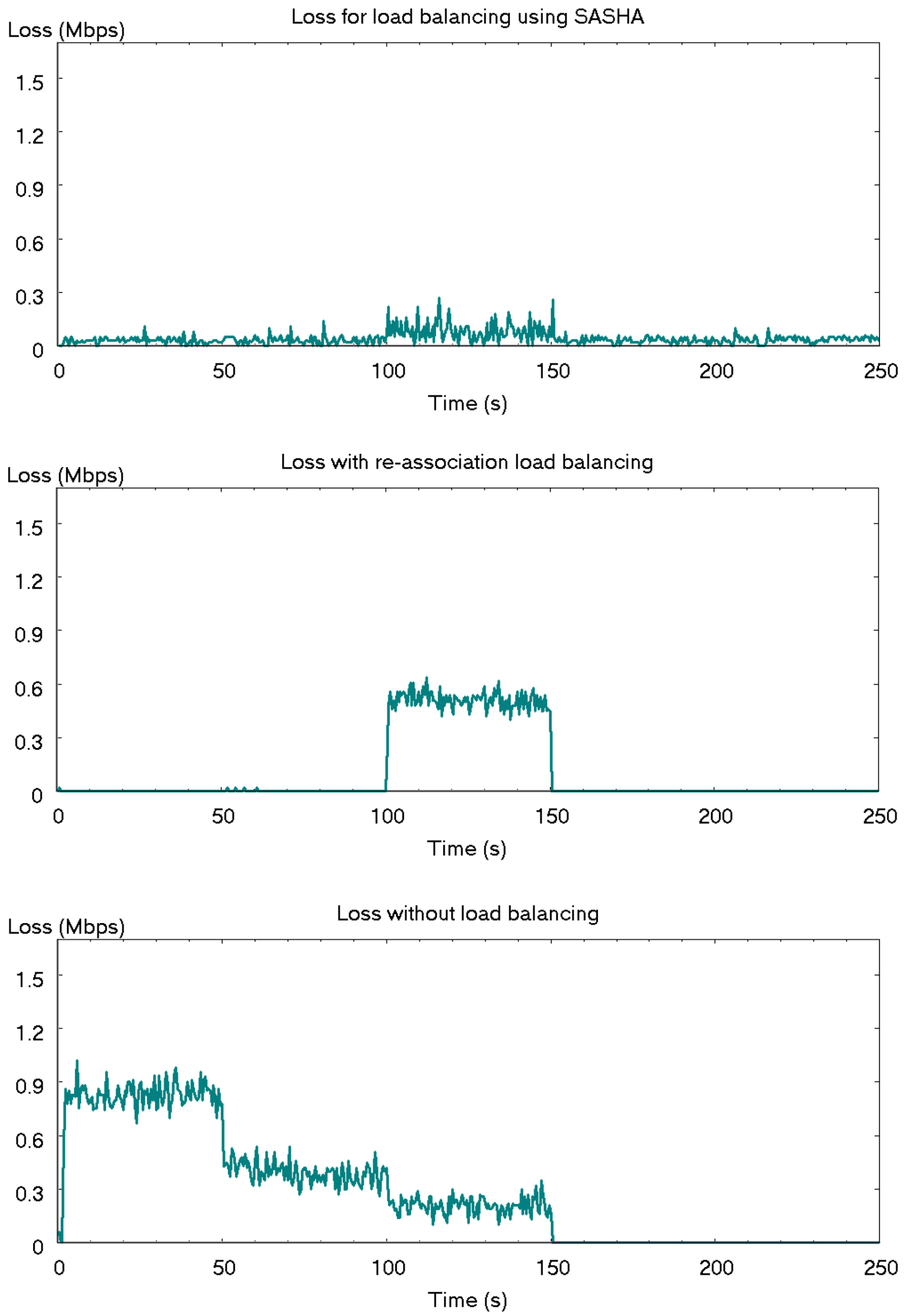

Figure 7.29: Loss for load balancing using SASHA, Mobile DCCP and NO LB.

ground traffic decreases, the network becomes lightly loaded or has no load at all determining a QoS increase to the maximum as it can be seen in Figure 7.29 and Figure7.30 over the period (150s - 250s).

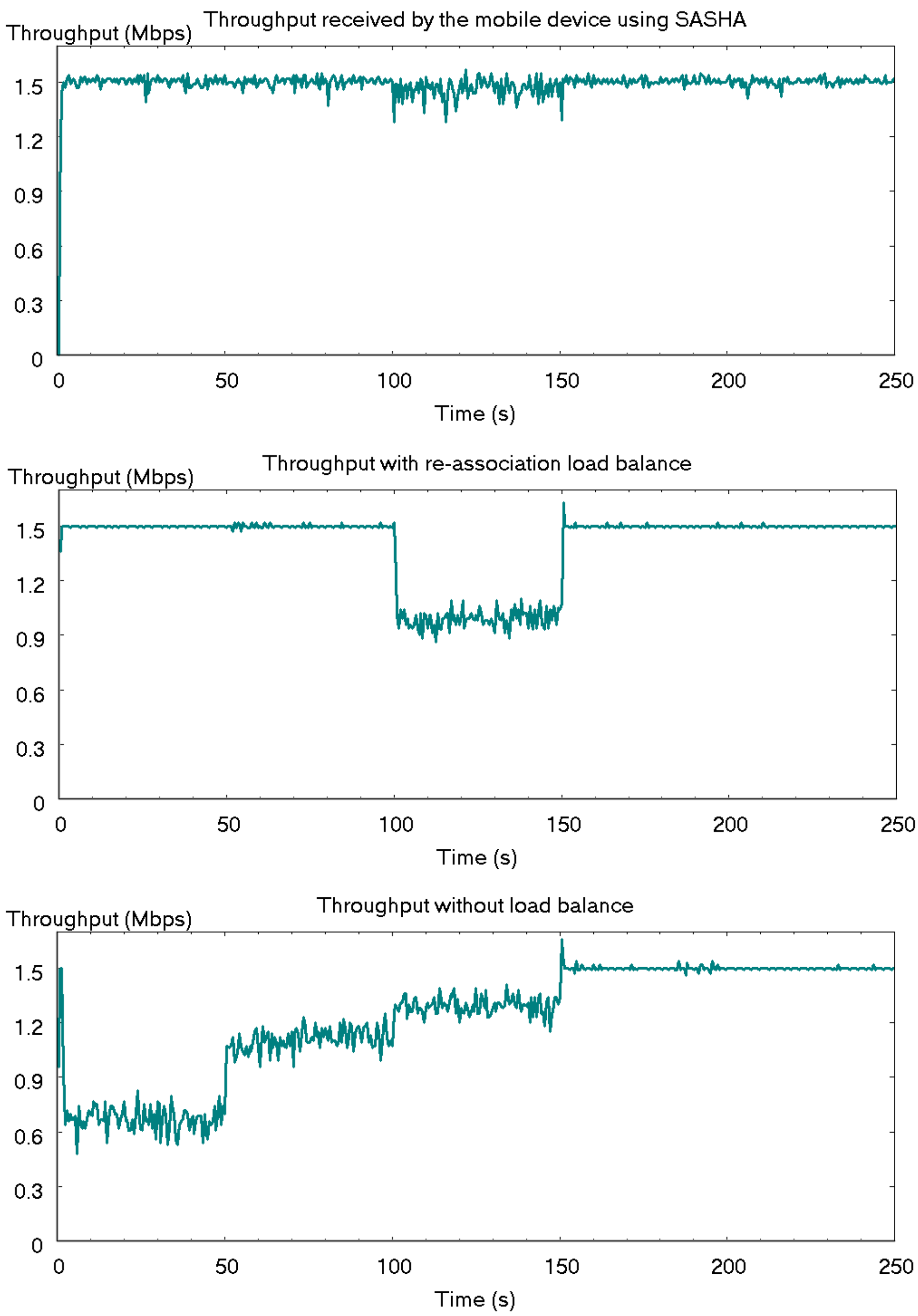

Figure 7.30: Throughput for load balancing using SASHA, Mobile DCCP and NO LB.

In Table 7.8 the average values of throughput, loss and PSNR are presented for the first three individual load situations (phases). The last two phases (Phase 3 and Phase 5) are similar with Phase 1 and Phase 2.

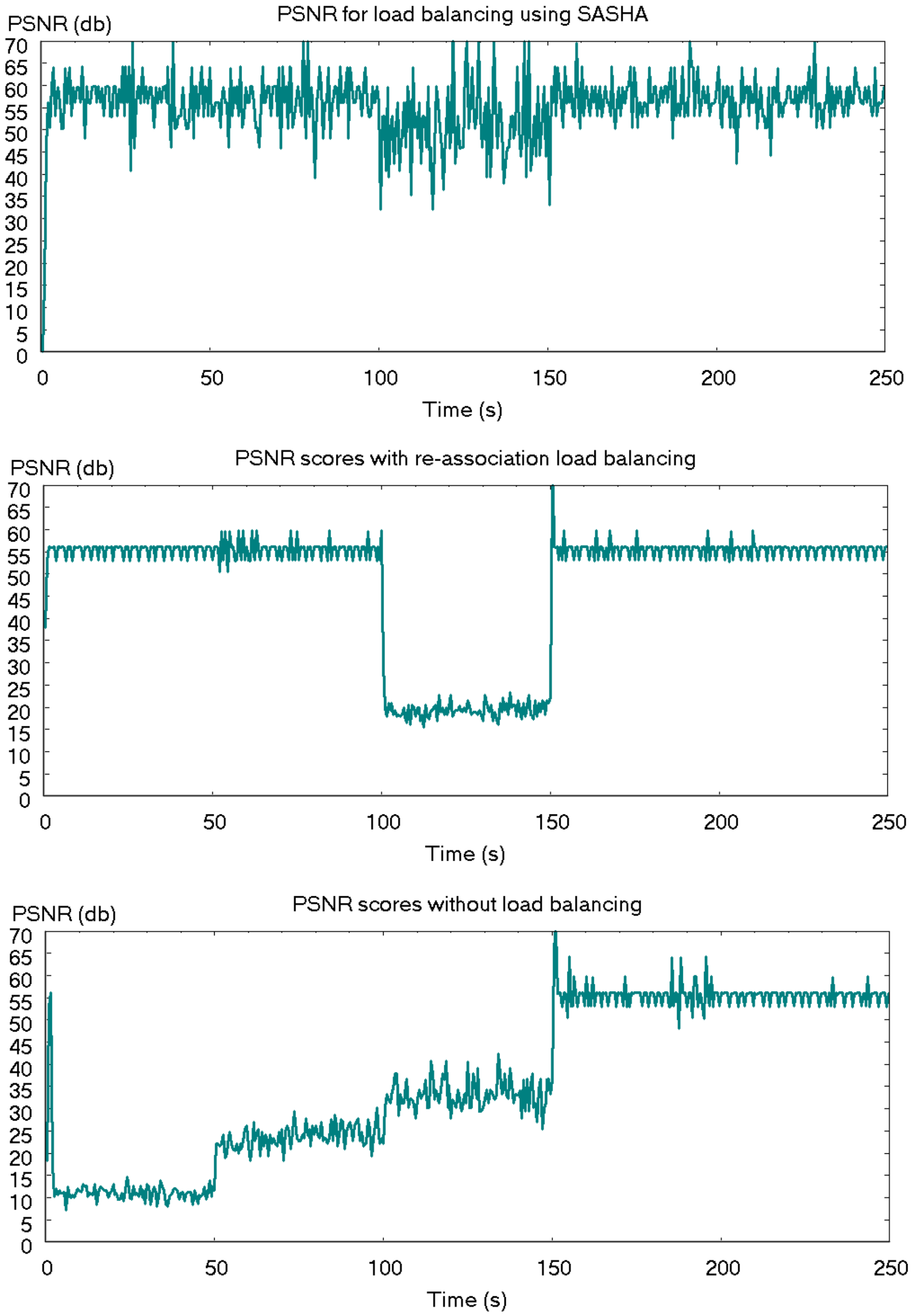

Figure 7.31: PSNR for load balancing using SASHA, Mobile DCCP and NO LB.

In Phase 1 Network 1 has 4.5 Mbps background traffic (overloaded) and Network 2 has 0 Mbps background traffic (no load). In Phase 2, Network 1 has 4.5 Mbps background traffic (overloaded) and Network 2 gets 2.5 Mbps background traffic (lightly). In Phase 3

Table 7.8: Average PSNR, Throughput and Loss when streaming a maximum of 1.5Mbps video in the context of three load phases

Scheme	Phase	PSNR		Throughput		Loss	
		(dB)	%	(Mbps)	%	(Mbps)	%
SASHA	1	57.6	-	1.50	100	0.00	0.0
	2	56.8	-	1.50	100	0.00	0.0
	3	51.0	-	1.46	97	0.08	5.3
Mobile DCCP	1	55.3	-	1.50	100	0.00	0.0
	2	55.5	-	1.50	100	0.00	0.0
	3	19.8	-	0.99	66	0.51	34
NO LB	1	11.90	-	0.69	46	0.80	53
	2	23.55	-	1.10	73	0.39	26
	3	33.16	-	1.29	86	0.20	13

both Networks have fully loading background traffic (3.5 Mbps).

As it can be seen in the table the performance of SASHA load balancing and re-association (Mobile DCCP) load balancing are similar when at least one network has capacity and is capable of providing the transport service to support the required QoS. The performance of SASHA load balancing is 51% better than Mobile DCCP in terms of throughput and 157% higher than Mobile DCCP in terms of PSNR when both Networks are fully loaded.

7.8 Video Quality Assessment

The main target of any multimedia application is to provide the user with the highest quality content possible. Consequently the evaluation of SASHA will be performed in the context of user perceived quality as well. This is done by estimating the video quality of the content delivered to the mobile user when SASHA is used for handover management.

Video quality assessment can be done in two manners. One direction involves **objective assessment** where complex metrics are used to estimate the video quality as it would be perceived by the end-user. There are various types of metrics from simple ones to very complex models which accurately include the characteristics of the human visual system.

The second direction for video quality assessment consists of **subjective testing**. Subjective video quality assessment is performed using human subjects who are watching a certain set of video sequences and rate their perceived video quality on a predefined scale.

The video quality assessment is performed using a set of video clips which are affected by the transport errors. For this purpose an emulator prototype was built to emulate the transport of video content from the media server to the mobile client over a wireless environment. It performs handover and load balancing using SASHA and Mobile DCCP respectively.

7.8.1 Test Video Sequences and Networking Scenarios

For video quality assessment four distinct video clips were chosen. Each represents a movie trailer with an average amount of spatial and temporal motion. The average length of each clip is 2 minutes. Clips are encoded using the MPEG-2 standard and have a resolution of 800x480, and a frame-rate of 25fps, typical values for video content manipulated on portable devices.

The reason for choosing MPEG-2 is its maturity and ease of access to open source encoders for prototype development. However the frame splitting used in the experiments is independent of the encoder used. Consequently any standard encoder can be used with good performance, including an MPEG4 one, for example.

The proposed handover solution targets any type of portable device, ranging from smartphone to notebook PC. Consequently the clips resolution has been such chosen to match a mid range graphical screen resolution which range from 480 x 360 in the case of Smartphones up to 1920x1200 in case of some notebook PCs.

SASHA's performance was evaluated in comparison with that of Mobile DCCP. The previous tests demonstrated that Mobile DCCP is the most versatile among the handover schemes considered. Choosing a single solution enabled to lower the number of test sequences to which viewers are exposed to, and limit the duration of the perceptual tests.

Two distinct scenarios are considered which use background traffic and the consequent network load to trigger handover. In the first scenario each network has a traffic node which

Figure 7.32: Frame from ”A-Team” clip

Figure 7.33: Frame from ”Nine” clip

generates video-like CBR background traffic of the pattern presented in Figure 7.36. In the second scenario there are four traffic nodes for each network, the cumulated CBR traffic presenting the same pattern as in Figure 7.36.

The traffic pattern is similar in both scenarios and is presented in Figure 7.36. However

Figure 7.34: Frame from ”Robin-Hood” clip

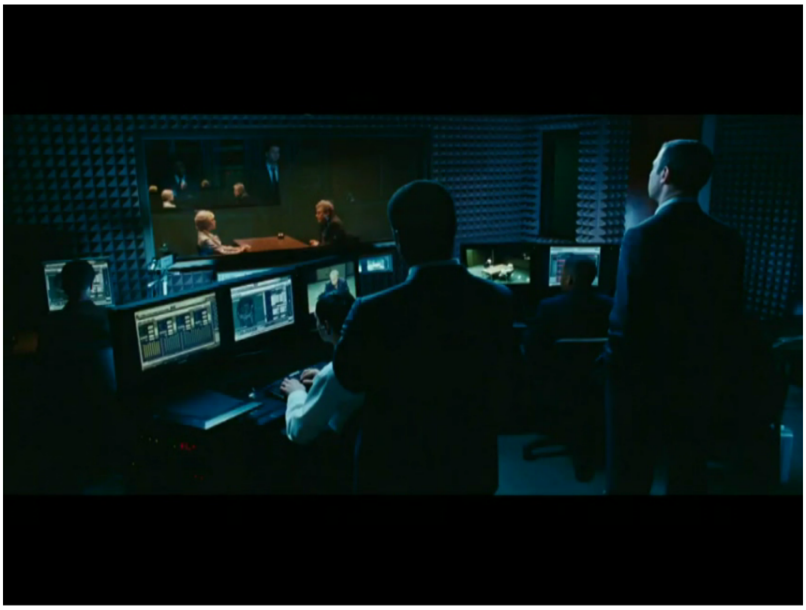

Figure 7.35: Frame from ”Salt” clip

multiple nodes are used in the second scenario to determine higher number of collisions.

The traffic pattern is similar with the one used in the simulation tests in order to allow for comparisons of results (see Figure 7.28.

The handover is performed in both scenarios by using SASHA and Mobile DCCP in

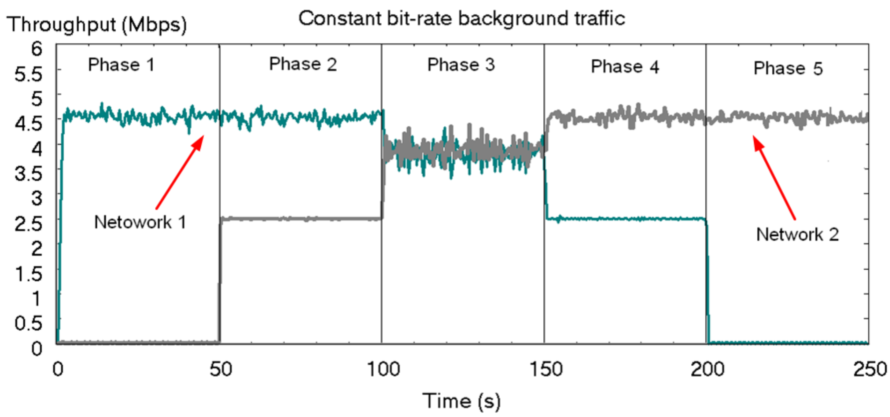

Figure 7.36: Constant bit-rate background traffic

turn. Using the prototype the delivery of each media clip to the mobile device using each of the two handover solutions is emulated. The resulting video clips are then used for video quality assessment purposes. Both objective metrics and subjective techniques were used on these video clips to assess the performance of SASHA and Mobile DCCP in the previously presented scenarios.

7.8.2 Objective Video Quality Assessment

Objective video quality assessment uses complex algorithms or models to evaluate the quality of the video content as close as possible to the way human visual system perceives it. There is no objective video quality metric generally accepted to be accurate enough in measuring user perceived quality.

Consequently, various metrics are used both in academic and industry research to assess video quality. To assess SASHA's quality performance three distinct full reference metrics have been used. Among these are the full-reference **Pick Signal-to-Noise Ratio (PSNR)** [68], **Video Quality Measurement (VQM)** [76], and **Structural Similarity (SSIM)** [75].

PSNR is easy to use, has low computational complexity, but was criticized for poor correlation with human perceived quality [188]. **SSIM** index is a full reference metric designed to assess the similarities between two images. SSIM aims at being more consistent with the perception of the human eye. **VQM** measures the perceptual effects of different

Table 7.9: Objective video quality assessment, average values.

Scheme	Scenario	PSNR (dB)	VQM	SSIM
SASHA	Scenario 1	20.89	4.39	0.85828
SASHA	Scenario 2	21.69	4.36	0.85966
MDCCP	Scenario 1	15.35	8.76	0.69093
MDCCP	Scenario 2	18.38	3.59	0.88242

kind of video impairments such as blurring, jerky motion, blockiness, etc. and provides a higher correlation with subjective quality assessment.

The higher the scores obtained using PSNR and SSIM the better the quality, while in case of VQM the lower the score the higher the quality.

Table 7.9 presents the results obtained by SASHA and Mobile DCCP in the two networking scenarios considered. It can be observed in the table that in terms of PSNR SASHA performs better than Mobile DCCP. SASHA presents PSNR scores around 21 dB which is an acceptable value for video transmission over lossy wireless channels. Mobile DCCP presents poor PSNR scores with values as low as 18.38 dB and 15.35 dB.

In terms of VQM and SSIM, SASHA presents similar scores in both scenarios demonstrating its resilience to different number of wireless nodes engaged in data traffic simultaneously. Although Mobile DCCP performs much worse than SASHA in the first scenario, in the second scenario its performance presents a slight improvement over SASHA. This may be determined by the image distortion introduced by the interlaced splitting and merging process.

Figure 7.37 presents the average PSNR, VQM and SSIM scores obtained by SASHA and Mobile DCCP. The presented values are the average performed over the scores obtained in the two networking scenarios considered.

The better performance presented by SASHA can be clearly observed in each graph separately. In terms of PSNR SASHA presents a 26% improvement over Mobile DCCP. VQM shows an even better performance with a 29% improvement while SSIM presents a reduced 9% improvement.

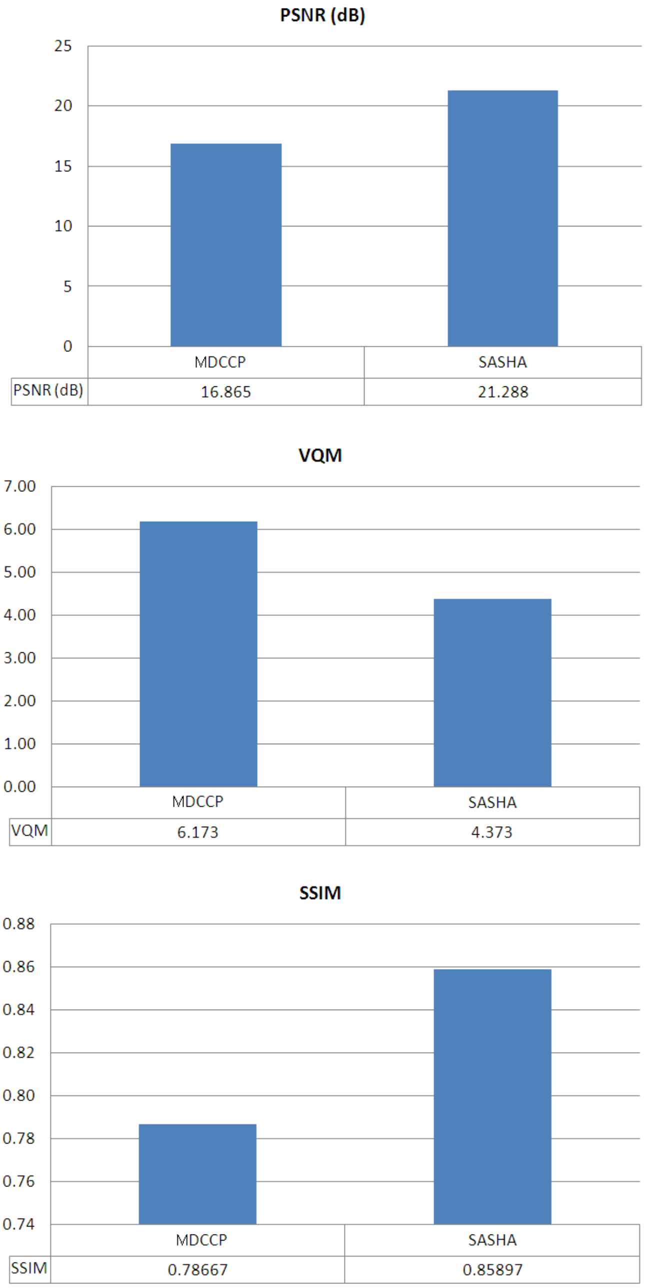

Figure 7.37: Objective video quality assessment.

7.8.3 Subjective Testing

Subjective video quality assessment has been performed including 22 human testers, a number between 6 and 40 subjects suggested in the ITU-T R. P.911 Recommendation [64].

The test environment has set according to above-mentioned recommendations. The

Table 7.10: Subjective video quality assessment, average values.

Scheme	Scenario	Video Quality	Video Continuity	Audio-Video Sync.
SASHA	Scenario 1	3.40	4.11	4.10
SASHA	Scenario 2	3.25	4.00	4.15
DCCP	Scenario 1	1.65	1.45	2.45
DCCP	Scenario 2	2.95	2.20	3.65

streamed multimedia clips are displayed on a average Notebook PS 13 inches monitor situated in a room with no natural light. The only source of light available was kept to a minimum intensity. The viewing distance was set to 5 times the hight of the picture.

The four distinct clips presented before, ”A-Team”, ”Nine”, ”Robin-Hood” and ”Salt”, have been used, each clip being processed by the prototype to emulate content delivery to mobile devices over wireless networks in the two scenarios discussed before (i.e. background traffic generated by one and four nodes per network respectively).

The users were asked to watch these clips and rate their perceived quality on a form. Three distinct aspects are considered. One is the overall perceived video quality, the second is the continuity of the video sequence and the last is the synchronization between audio and video. Each of these aspects are rated on a 5 points scale (ITU-T R. P.911 recommended scale) where 1 is the worst and 5 is the best level of quality, continuity and synchronization.

The results are presented in Table 7.10. The results are expressed as average values and are shown separately for each networking scenario.

The subjective testing shows a similar pattern like the objective video quality results presented in the previous section. SASHA perform similar in the two scenarios considered. The subjective video quality assessment results show that SASHA performs better than Mobile DCCP however DCCP shows a positive increase in performance when the four mobile nodes are used to generate background traffic.

Figure 7.38 presents the average values of the subjective video quality assessment results. It can be observed that on average the scores given by the test users to video clips delivered using SASHA are better than the ones delivered using Mobile DCCP. SASHA

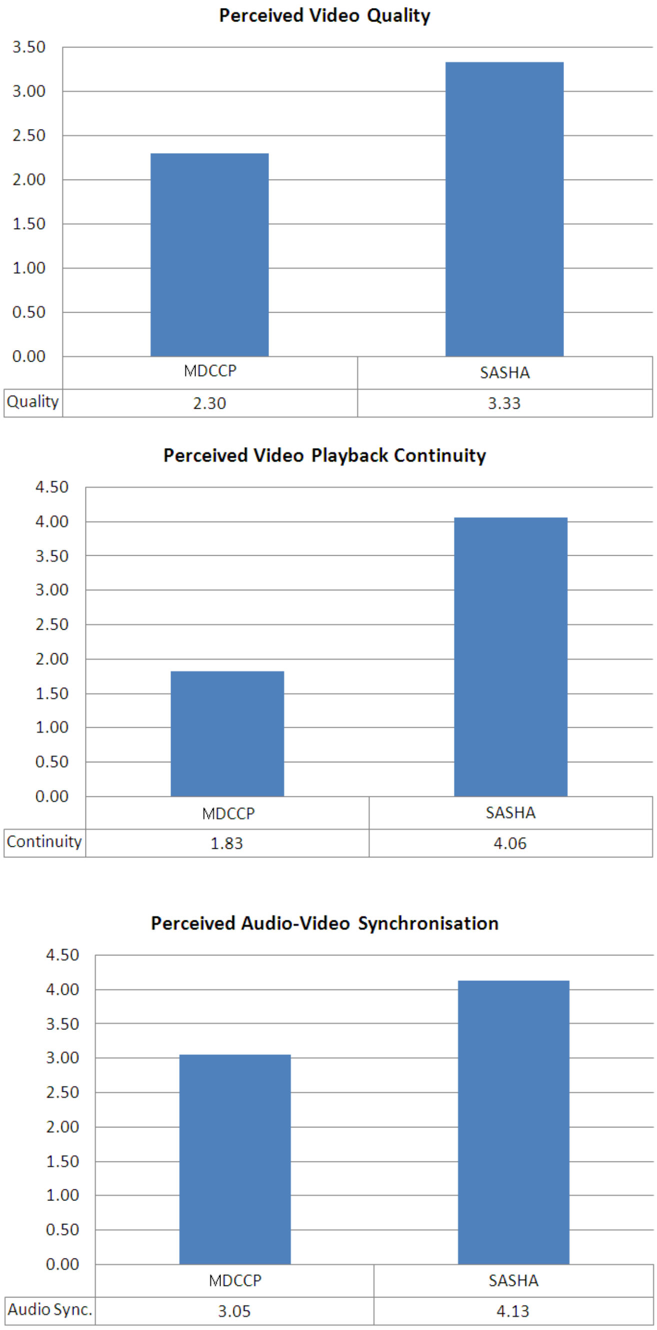

Figure 7.38: Subjective video quality assessment.

performs better in delivering a high quality video content with better playback continuity and little loss of synchronization between video sequence and the corresponding audio sequence.

A statistical analysis of variance (ANOVA) [189] was also performed on the results to evaluate if there is a significant difference between the video quality, video continuity and audio-video synchronization scores respectively, achieved by the two handover schemes discussed. To determine which scheme performs better and in which conditions multiple comparison tests were performed based on Games-Howell procedure and when homogeneity of variance could be verified Tukey HDS was used as well.

In terms of video quality, the statistical analysis shows that there is no significant difference between the performance of SASHA in the two scenarios considered (with 1 traffic node and 4 traffic nodes respectively) with a confidence level of 95%. Mobile DCCP shows a significant difference (95% confidence level) between the two scenarios, with a better performance when the four traffic nodes are involved. Regarding the two different schemes SASHA is proved to perform better than Mobile DCCP however the performance of Mobile DCCP when four traffic nodes are used does not appear to be significant enough for a confidence level of 95%. However the average values presented in Table 7.10 suggest SASHA as best performing.

The reason why there is no significant difference between SASHA and Mobile DCCP in certain scenarios (4 traffic nodes) is the traffic splitting mechanism which slightly alters the image quality in the merging process. This can be overcome with a more advanced multiple description coding algorithm. However this does not represents the subject of this thesis.

In terms of video playback continuity, SASHA still does not present any significant variation in performance between the two scenarios considered while Mobile DCCP continues to perform significantly better when four traffic nodes are involved. Regarding the two handover schemes, in terms of video playback continuity SASHA presents a significant improvement with a confidence level of 95% compared with Mobile DCCP. This can also be observed by analyzing the average values presented in Table 7.10 where SASH's video continuity scores are reaching the value of 4 while Mobile DCCP was only awarded scores around 2.

A similar trend was observed when analyzing the scores given by the users to the synchronization between the video and the audio content where SASHA's scores are also in the area of "very good" (4) while Mobile DCCP only reaches 3.65.

7.9 Overhead Analysis

SASHA relies on QMS scores which are computed at the server side SASHA module, but requires information collected by the client-side module. They are delivered to the server module via feedback. This information updating requires a certain amount of data to be transported over the network, which generates a certain level of overhead.

QMS information update requires messages to be sent to the server at regular intervals, each message having a 128 bytes payload. Depending on the network dynamics, the frequency of these update messages may vary; however in the experiments performed a feedback message was sent every 500 ms on average. This generates an overhead of 256 bytes per second which is an insignificant amount of data compared with the bitrate required by the multimedia content.

For example, a 10 minute video clip encoded at 2 Mbps requires 150 MBytes to be sent while the feedback overhead required is only 0.14 MBytes, which represents 0.09% increase in traffic due to overhead.

However when an adaptive multimedia streaming application is used, the feedback required by the adaptive technique may be used by SASHA as well or vice versa. Depending on the level of correlation between the feedback data required by the adaptive solution an SASHA, the overhead induced by feedback may reduced by approximately 50%.

7.10 Chapter Overview

SASHA was tested in various scenarios in order to validate its concept and its capability to offer a better multimedia service to mobile users over wireless heterogeneous environments. The performance of SASHA was compared with the performance achieved by other similar solutions including Mobile IP, Mobile DCCP and Mobile SIP in the same networking and application scenarios.

Performance evaluation was done using both simulations and experiments performed with an emulated prototype. Evaluation was made using QoS-related metrics such as throughput and packet loss, and user perceived quality metrics, in both objective and subjective tests. The objective video quality metrics considered are PSNR, VQM and SSIM.

A wide range of networking scenarios were considered. These scenarios include multiple mobile devices performing handover simultaneously, handover performed between networks with various overlapping area sizes and scenarios when various numbers of mobile devices generate background traffic.

Testing results have shown that SASHA performs better than the other similar solutions considered in the testing scenarios and is capable of providing better video quality for the mobile user accessing a multimedia content over a heterogeneous wireless environment.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

Mobility management has become a very important component of the mobile Internet and in particular handover management plays a major role in providing high quality Internet-based services to end users in the increasingly popular mobile computational and communication environment.

User mobility and error prone nature of the wireless communication environment challenge service providers in their goal to provide a reliable and high quality data communication services to network-based user applications. Among these, the most demanding are the multimedia streaming applications with their bandwidth hunger and real-time constraints, important for supporting high user perceived quality.

In these circumstances this theses proposed the Smooth Adaptive Soft-Handover Algorithm (SASHA), a novel handover mechanism for multimedia content delivery in heterogeneous wireless environments. SASHA is an application layer solution which offers mobility and quality-aware content delivery service to multimedia applications. SASHA consists of a server-side component which attaches to the multimedia streaming server application and a client-side component which is deployed at the media player used by the user device. The proposed handover mechanism integrates and cooperates well with adaptive multimedia streaming solutions toward increasing user quality of experience when accessing multimedia services. Quality Oriented Adaptation Scheme (QOAS) proposed by Muntean in [59] was used during SASHA testing to demonstrate this concept; however other similar adaptation schemes can be successfully used in collaboration with SASHA.

SASHA's main task is to perform handover management. It performs handover by dynamically distributing the traffic load among all the available wireless networks in a smooth manner.

As presented in the Related Work chapter, handover plays a major role in load balancing and some techniques involve handover in the load allocation/reallocation process. Consequently SASHA incorporates this role to a certain extent as well, in the sense that it contributes to congestion avoidance and efficient load balancing. However SASHA is not a standalone load balancing solution. SASHA, by its basic functioning principle, aggregates the throughput achieved over multiple networks simultaneously, its design allowing it to act as a standalone throughput aggregation solution.

Quality of Multimedia Streaming (QMS) was also proposed in this thesis. QMS is designed as a general metric for estimating the amount of traffic a certain network can efficiently carry. In this context efficiency is an application related term and consequently QMS is a comprehensive and flexible metric allowing the application to tune its components and its consequent behavior. QMS incorporates several parameters such as QoS, monetary cost, energy consumption and user QoE. All these parameters are weighted and their weights can be set according to the application requirements.

The testing of the proposed handover management algorithm and the traffic allocation metric has been performed using simulations and experiments conducted on an emulator prototype. For simulation purposes, Network Simulator 2 (NS-2) was used. Simulation models for SASHA and QMS have been developed.

Using the above mentioned simulation platform, a methodology for tuning QMS parameters for integration with a multimedia application has been presented. A QOAS-based adaptive multimedia streaming application integrated with SASHA has been developed in NS-2 and was used as the basis for the experiments.

QMS was tested in various networking scenarios with parameter weights taking values over the entire possible range. Using subjective video quality metrics (i.e VQM and PSNR) user satisfaction was estimated for each combination of weight values. User satisfaction and the average throughput and loss rates achieved were used to choose the best weight combination to maximize user QoE. If single values could not be chosen to achieve the application requirements (e.g. for QoS and Energy), the set of weight values were presented and a methodology to chose the appropriate value dynamically according to application and user requirements was presented.

The simulation-based experiments showed that QMS can be used in traffic distribution to meet the application and user specific requirements in terms of QoS, QoE, monetary cost and device battery lifetime.

SASHA, using QMS, was tested both in the simulation environment and using the prototype emulator. The simulation scenarios considered were intended to cover a wide range of situations where SASHA can be used to deliver multimedia content to mobile users. The scenarios include multiple mobile nodes performing handover simultaneously, networks with variable overlapping area sizes, multiple nodes generating background traffic and high level of collisions and scenarios where networks with different wireless communication technologies were used.

The performance of the proposed solution was evaluated using user QoE estimated by video quality and QoS measured by throughput and packet loss. The video quality was assessed using both objective metrics and subjective testing. The objective metrics used were PSNR, VQM and SSIM. Subjective tests were performed using 22 subjects in a controlled laboratory environment. The role of the subjective tests was to verify and confirm the objective testing results, which they did.

SASHA's performance was compared against the results obtained by other three handover management solutions in the same networking and application scenarios. These solutions are Mobile IP (MIP), Mobile DCCP and Mobile SIP. Each solution resides at a different protocol stack layer, network, transport and session layer, respectively.

SASHA performed better than the other three solutions considered in all the scenarios involved. To exemplify, in the context of multiple mobile nodes performing handover simultaneously, SASHA achieved a performance gain in terms of PSNR of 37% for 3 mobile nodes, 20% for 5 mobile nodes and 10% for 10 mobile nodes as compared with the best competitor which is Mobile DCCP. In the scenarios involving network congestion generated by background traffic nodes the performance of SASHA load balancing capability shows performance improvements of 51% in terms of throughput and 157% in terms of PSNR when both networks are fully loaded.

Objective video quality assessment performed in the context of two network scenarios involving high network congestion levels also demonstrated SASHA's better performance against Mobile DCCP. In this case, as it is the case of the subjective testing as well, the emulator prototype was used to emulate the transmission of video over wireless networks to the mobile user using SASHA and QMS.

In terms of PSNR, SASHA presents a 26% improvement, with an improvement of 29% measured by VQM and a 9% improvement presented by the results obtained using SSIM.

The results obtained from objective video quality assessment were confirmed by the subjective testing were SASHA has achieved a 44% improvement against Mobile DCCP in terms of video quality as rated by the users.

There could be identified several limitations of SASHA, the proposed handover mechanism. One is the limited number of network interfaces available on the mobile device which in certain circumstances may prevent SASHA from connecting to multiple networks and therefore performing an optimum traffic distribution. Another limitation is the increased amount of energy determined by using multiple network interfaces in parallel. Even in an ideal context, where the mobile device is equipped with the required wireless networking resources for communication over all available networks distributing the traffic over too many networks may lead to a reduced performance due to traffic splitting overhead.

To summarize, the contributions of this thesis consist in introducing a novel handover management algorithm (SASHA) based on dynamic traffic distribution over multiple simultaneous connections, proposing a novel comprehensive and flexible quality metric (QMS) for traffic distribution in wireless environments and development of a new quality aware mobile data communication architecture and framework for multimedia content delivery to mobile users over heterogeneous wireless environments.

The thesis was structured in eight chapter as follows.

Chapter 1 introduces the motivation of the research activity conducted, the problem statement is exposed and then a brief overview of the solution is presented. The contributions to the advancement of the state of the art are detailed as well in this chapter.

Chapter 2 discusses the background technologies related to wireless data communications, multimedia streaming and mobile devices, all these being related to the subject investigated.

Chapter 3 presents the related works in tight relationship with SASHA and the QMS metric.

Chapter 4 introduces the QMS metric and discusses its parameters. Each parameter is presented separately and its mathematical formula is introduced.

Chapter 5 presents SASHA architecture and related algorithms.

Chapter 6 describes the methodology used to evaluate QMS parameter weighting and suggests default values. Test results are detailed, experimental data for QMS tuning in different application contexts being provided.

Chapter 7 discusses SASHA performance evaluation and analyzes testing results. Simulation models and scenarios are presented. The testing results are detailed and the better performance of SASHA against its competitors is proved.

Chapter 8 concludes the thesis and presents possible future work directions.

8.2 Future Works

The work presented in this thesis was mainly focused on proposing the SASHA (and QMS) concept as a viable handover management solution, and demonstrating its capacity to offer a better multimedia streaming service to mobile users in the context of a heterogeneous wireless environment.

Several research directions on the short term can be identified and will be pursuit to further develop SASHA and QMS.

A first direction will be to investigate SASHA as a load balancing and throughput aggregation technique. Currently SASHA contributes to load balancing in the wireless environment. However it is a fully distributed approach with no direct interaction between distinct mobile hosts involved in data communication simultaneously and with no knowledge of the overall status of the wireless network environment. To overcome this, inter-host cooperation will be investigated as well as a possible centralized approach as an alternative to the current distributed one.

A second direction will target SASHA deployment on wireless MESH networks. MASH networks are increasingly popular and being wireless multi-hop networks pose some particular challenges for real-time application in particular, which is exactly SASHA's target application scenario.

The third aspect targets the development of a fully functional prototype incorporating SASHA and QMS. A testbed will be prepared and the prototype will be deployed on this testbed. This step is extremely important for real-life testing and evaluation of SASHA and QMS aiming at providing the path toward a possible integration of SASHA and QMS in a commercial multimedia streaming application. The current evaluation prototype relies on simulations for network delivery.

The fourth direction targets a comprehensive evaluation of QMS and its parameters on the experimental testbed. This step aims at evaluating SASHA with all possible weights combination and tuning parameters values in a wider range of networking scenarios. Various video encoding schemes will be considered in order to investigate the impact of QMS parameters and their weights on the video quality. This task is of crucial importance for product development and will provide a complete data set for tuning QMS for best performance in the context of specific application requirements.

Based on the testbed and the findings of the experiments mentioned above the stability of SASHA and QMS will be tested. An enhanced threshold-based mechanism will be developed to avoid the ping-pong effect and to increase the overall system stability. In the same context the adaptability and the reaction speed of the system will be evaluated and methods for its fine tuning will be proposed.

The impact of SASHA and especially the video quality assessment algorithms on the mobile terminal energy consumption will be evaluated and methods to optimize energy consumption will be sought by moving as many computational intensive tasks as possible to the server and by investigating the correlation between the QMS parameters in order to reduce complexity.

A possible method for reducing mobile host energy consumption would be to extract the relevant parameters (i.e. quantization factor, motion vectors) from the received stream and send them as feedback to the server which will further model and compute the video quality.

On the long term the research activity presented in this thesis makes room for investigation in several adjacent fields.

One direction will be energy saving. Currently QMS incorporates a energy related parameter but this only considers the energy consumption per megabit transmitted or received. Although the wireless transceiver is among the most energy hungry components of the mobile device there are other modules which consume significant amount of power. One of these is the processor and another is the display.

The processor is involved in decoding the video streams and the characteristics of the encoding scheme may impact on the amount of processing power required and consequently the energy consumption. To exemplify, consider a video encoding scheme with error concealment. The error concealment feature will allow for a higher maximum loss rate than a encoding scheme without this feature and consequently will allow SASHA to distribute a higher amount of traffic on lossy links. Although the loss concealment feature will overcome the transmission errors it will require more processing power to achieve this which in turn will increase the energy consumption of the mobile device. This and other similar scenarios will be investigate in the future, and QMS will be further developed to incorporate the findings.

An important module within SASHA architecture which was not investigated in detail in this work is the traffic splitter and merger. This is necessary for traffic distribution and is crucial for the success of any data communication approach using multiple simultaneous connections or networks. This aspect can be approached in various ways, however the Multiple Description Coding (MDC) is among the most robust and best performing. Various techniques for MDC will be investigated and new ones will be proposed to best fit SASHA and QMS in the context of heterogeneous wireless environments.

Finally, another research direction will target investigation of MPEG4 region-oriented coding and the integration of previous research conducted by the author in the field of Region of Interest (RoI) adaptive multimedia streaming with SASHA and multi-connection data transfer. RoI-based adaptive streaming identifies regions within the video image in which the users are most interested in and treat them distinctly in the adaptation process in order to increase user perceived quality. QoE will be further investigated in this context aiming at prioritizing components of the multimedia content (e.g. RoI, Audio track) based on their impact on user perceived quality. This approach combined with SASHA can lead to a solution which prioritizes regions within the video image and allocates the most important ones for transmission to the best connections available while the least important ones to connections with lower QoS levels.

Bibliography

[1] A. M. Bock, *Video compression systems*. Institution Of Engineering & Technology (IET), 2009.

[2] R. Engelmann, "The Origins of Radio," *Potentials, IEEE*, vol. 19, no. 4, pp. 41 –44, October 2000.

[3] J. Haartsen, "The Bluetooth Radio System," *IEEE Personal Communications*, vol. 7, no. 1, pp. 28 –36, February 2000.

[4] *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks (WPANs)*, IEEE 802.15.1 Std., 2002.

[5] F. Reynolds, "Whither Bluetooth?" *IEEE Pervasive Computing*, vol. 7, no. 3, pp. 6 –8, July 2008.

[6] *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)*, IEEE 802.15.4 Std., 2003.

[7] *ZigBee Specification v1.0*, ZigBee Alliance Std., 2006.

[8] H. Zhai, "QoS Support over UWB Mesh Networks," in *IEEE Wireless Communications and Networking Conference (WCNC 2008)*, April 2008, pp. 2283 –2288.

[9] *High rate ultra wideband phy and mac standard, 1st edition*, Standard ECMA-368 Std., December 2005.

[10] *ISO/IEC 26907 Information technology Telecommunications and information exchange between systems High Rate Ultra Wideband PHY and MAC Standard*, International Standards Organisation Std., 2007.

[11] G. Mulligan, "The 6LoWPAN Architecture," in *Proceedings of the 4th Workshop on Embedded Networked Sensors (EmNets '07)*. New York, NY, USA: ACM, 2007, pp. 78–82.

[12] *IEEE Standard for Local and Metropolitan Area Networks Specfic Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specfications*, IEEE 802.11 Std., June 1999.

[13] *IEEE Standard for Local and Metropolitan Area Networks Specfic Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specfications High Speed Physical Layer Extension in the 2.4 GHz Band*, IEEE 802.11b Std., September 1999.

[14] *IEEE Standard for Local and Metropolitan Area Networks Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications High Speed Physical Layer in the 5 GHz Band*, IEEE 802.11a Std., 1999.

[15] *IEEE Standard for Local and Metropolitan Area Networks Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band*, IEEE 802.11g Std., June 2003.

[16] *IEEE Standard for Local and Metropolitan Area Networks Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications MAC Enhancements for QoS*, IEEE 802.11e Std., 2005.

[17] Y. Xiao, "Performance Analysis of Priority Schemes for IEEE 802.11 and IEEE 802.11e Wireless LANs," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1506 – 1515, July 2005.

[18] *IEEE Draft Standard for Local and Metropolitan Area Network–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 5: Enhancements for Higher Throughput*, P802.11n/D7.0 Std., September 2008.

[19] L. Eastwood, S. Migaldi, Q. Xie, and V. Gupta, "Mobility Using IEEE 802.21 in a Heterogeneous IEEE 802.16/802.11-based, IMT-advanced (4G) Network," *IEEE Wireless Communications*, vol. 15, no. 2, pp. 26 –34, April 2008.

[20] *IEEE Draft Standard for Information Technology Telecommunications and Information Exchange Between System LAN/MAN Specific Requirements Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 10: Mesh Networking*, IEEE P802.11s/D4.0 Std., December 2009.

[21] G. Hiertz, D. Denteneer, S. Max, R. Taori, J. Cardona, L. Berlemann, and B. Walke, "IEEE 802.11s: The WLAN Mesh Standard," *IEEE Wireless Communications*, vol. 17, no. 1, pp. 104 –111, February 2010.

[22] G. Hiertz, D. Denteneer, L. Stibor, Y. Zang, X. Costa, and B. Walke, "The IEEE 802.11 Universe," *IEEE Communications Magazine*, vol. 48, no. 1, pp. 62 –70, January 2010.

[23] M. S. Kuran and T. Tugcu, "A Survey on Emerging Broadband Wireless Access Technologies," *Computer Networks*, vol. 51, no. 11, pp. 3013 – 3046, 2007.

[24] D. H. Lee, K. Kyamakya, and J. Umondi, "Fast Handover Algorithm for IEEE 802.16e Broadband Wireless Access System," in *1st International Symposium on Wireless Pervasive Computing*, 2006, p. 6 pp.

[25] *Broadband Radio Access Net-works (BRAN) HIPERACCESS System Overview*, ETSI HiperACCESS Std., March 2002.

[26] *Broadband Radio Access Networks (BRAN); Functional Requirements for Fixed Wireless Access Systems Below 11 GHz: HIPERMAN*, ETSI HiperMAN Std., March 2001.

[27] D. Kim, "WiBro Overview and TTA Activities," TTA, Tech. Rep., 2005. [Online]. Available: http://docbox.etsi.org/workshop/gsc/gsc10archive/GSC10_GRSC3/gsc10_grsc3_29%20WAS_TTA.ppt

[28] E. Cianca, R. Prasad, M. De Sanctis, A. De Luise, M. Antonini, D. Teotino, and M. Ruggieri, "Integrated Satellite-HAP Systems," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 33 – 39, December 2005.

[29] G. Chouinard, "Status of work in the IEEE 802.22 WG," -, Tech. Rep., 2005. [Online]. Available: www.rabc.ottawa.on.ca/e/Files/RABC%20802.22%20Presentation.ppt

[30] S. Ortiz, "4G Wireless Begins to Take Shape," *Computer*, vol. 40, no. 11, pp. 18 –21, November 2007.

[31] M. Anas, C. Rosa, F. Calabrese, P. Michaelsen, K. Pedersen, and P. Mogensen, "QoS-Aware Single Cell Admission Control for UTRAN LTE Uplink," in *IEEE Vehicular Technology Conference (VTC 2008).*, May 2008, pp. 2487 –2491.

[32] J. Gozalvez, "Ultra Mobile Broadband [Mobile Radio]," *IEEE Vehicular Technology Magazine*, vol. 2, no. 1, pp. 51 –55, March 2007.

[33] W. Bolton, Y. Xiao, and M. Guizani, "IEEE 802.20: Mobile Broadband Wireless Access," *IEEE Wireless Communications*, vol. 14, no. 1, pp. 84 –95, February 2007.

[34] U. Ladebusch and C. Liss, "Terrestrial DVB (DVB-T): A Broadcast Technology for Stationary Portable and Mobile Use," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 183 –193, January 2006.

[35] *Transmission System for Handheld Terminals (DVB-H), ETSI EN 302304 v1.1.1*, Digital Video Broadcasting (DVB) Std., November 2004.

[36] M. Kornfeld and K. Daoud, "The DVB-H Mobile Broadcast Standard [Standards in a Nutshell]," *IEEE Signal Processing Magazine*, vol. 25, no. 4, pp. 118 –122, 127, July 2008.

[37] M. Kornfeld and G. May, "DVB-H and IP Datacast mdash;Broadcast to Handheld Devices," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 161 –170, March 2007.

[38] *ISO/IEC 10918-1 - ITU-T Recommendations T.81, Information technology - Digital compression and coding of continous-tone still images: Requirements and guidelines*, ISO - ITU Std., January 1992.

[39] M. Isnardi, "Historical Overview of Video Compression in Consumer Electronic Devices," in *Digest of Technical Papers. International Conference on Consumer Electronics (ICCE 2007)*, January 2007, pp. 1 –2.

[40] *ISO/IEC 15444-1 - ITU-T Recommendations T.800, Information TechnologyJPEG 2000Image Coding SystemPart 1: Core Coding System*, International Standards Organisation Std., 2000.

[41] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG 2000 Still Image Coding System: An Overview," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103–1127, November 2000.

[42] *ISO/IEC 15444-3 Information Technology - JPEG 2000 Image Coding System: Motion JPEG2000*, International Standards Organisation Std., 2002.

[43] *International Standard ISO/IEC 11172-2 MPEG-1 Video*, ISO/IEC Std., 1993.

[44] *International Standard ISO/IEC 13818-2 MPEG-2 Video*, ISO/IEC Std., 1995.

[45] *International Standard ISO/IEC 14496-2 MPEG-4 Visual*, ISO/IEC Std., 2004.

[46] *International Standard ISO/IEC 14496-10 MPEG-4 AVC*, ISO/IEC Std., 2005.

[47] *ITU-T Recommendation H.261 - Video codec for audiovisual services at p x 64 Kb/s*, International Telecommunication Union Std., March 1993.

[48] *ITU-T Recommendation H.262 - Information technology - Generic coding of moving pictures and associated audio information: Video*, International Telecommunication Union Std., February 2000.

[49] *ITU-T Recommendation H.263 - Video coding for low bit rate communication*, International Telecommunication Union Std., February 1998.

[50] *ITU-T Recommendation H.264 - ISO/IEC 14496-10 AVC - Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG Std., 2003.

[51] *VC-1 Compressed Video Bitstream Format and Decoding Process - SMPTE 421M-2006*, Society of Motion Picture and Television Engineers Std., 2006.

[52] *Information technology - advanced coding of audio and video - Part 2: Video*, GB/T 20090.2-2006 Std., 2006.

[53] F. Hartung, U. Horn, J. Huschke, M. Kampmann, T. Lohmar, and M. Lundevall, "Delivery of Broadcast Services in 3G Networks," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 188 –199, March 2007.

[54] P. Frojdh, U. Horn, M. Kampmann, A. Nohlgren, and M. Westerlund, "Adaptive Streaming Within the 3GPP Packet-Switched Streaming Service," *IEEE Network*, vol. 20, no. 2, pp. 34 –40, March 2006.

[55] A. Cuevas, J. Moreno, P. Vidales, and H. Einsiedler, "The IMS Service Platform: a Solution for Next-Generation Network Operators to be More Than Bit Pipes," *IEEE Communications Magazine*, vol. 44, no. 8, pp. 75 –81, August 2006.

[56] *Multimedia Broadcast/Multicast Service (MBMS); Stage 1 (Release 7) Tech. Rep. 3G TS 22.146 V7.1.0*, 3GPP Std., March 2006.

[57] *Broadcast Multicast Service for CDMA2000 1x Systems C.S0077 Rev. 1.0*, 3GPP2 Std., April 2006.

[58] K. Kang and T. Kim, "Improved Error Control for Real-Time Video Broadcasting Over CDMA2000 Networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 188 –197, January 2009.

[59] G.-M. Muntean, P. Perry, and L. Murphy, "Subjective Assessment of the Quality-Oriented Adaptive Scheme," *IEEE Transactions on Broadcasting*, vol. 51, no. 3, pp. 276 – 286, September 2005.

[60] ANSI T1.801.02, "Digital Transport of Video Teleconferencing/Video Telephony SignalsPerformance Terms, Definitions, and Examples," 1996.

[61] N. Montard and P. Bretillon, "Objective Quality Monitoring Issues in Digital Broadcasting Networks," *IEEE Transactions on Broadcasting*, vol. 51, no. 3, pp. 269 – 275, September 2005.

[62] *Recommendation ITU-R BT 500-10 - Methodology for the Subjective Assessment of the Quality of the Television Picture*, ITU-R Std.

[63] *ITU-T Recommendation P.910 - Subjective Video Quality Assessment Methods for Multimedia Applications*, ITU-T Std., Sept 1999.

[64] *ITU-T Recommendation P.911 - Subjective Audiovisual Quality Assessment Methods for Multimedia Applications*, ITU-T Std., 1998.

[65] S. Winkler, A. Sharma, and D. McNally, "Perceptual Video Quality and Blockiness Metrics for Multimedia Streaming Applications," in *Proceedings of the International Symposium on Wireless Personal Multimedia Communications*, 2001, pp. 547–552.

[66] D. Miras, “Network QoS Needs of Advanced Internet ApplicationsA Aurvey,” Internet2 QoS Working Group, 2002.

[67] S. Winkler and P. Mohandas, “The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics,” *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660 –668, September 2008.

[68] VQEG, “Objective Perceptual Assessment of Video Quality: Full Reference Television,” March 2005. [Online]. Available: ftp://vqeg.its.bldrdoc.gov/Documents/VideoQualityTutorial/TutorialJ144-1149March05.doc

[69] O. Verscheure, P. Frossard, and M. Hamdi, “User-Oriented QoS Analysis in MPEG-2 Video Delivery,” *Real-Time Imaging*, vol. 5, no. 5, pp. 305 – 314, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/B6WPR-45FJXX0-8/2/5cc1dde3573eb165707d3720698fcad4

[70] M. Knee, “The Picture Appraisal Rating (PAR) - A Single-ended Picture Quality Measure for MPEG-2,” Snell and Wilcox, Tech. Rep., January 2000.

[71] S. Daly, “The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity,” *Digital Images and Human Vision*, vol. -, pp. 179–206, 1993.

[72] D. F. J. Lubin, “Sarnoff JND Vision Model,” T1A1.5 Working Group Document 97-612, ANSI T1 Standards Committee, 1997.

[73] C. J. van den Branden Lambrecht and O. Verscheure, “Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System,” 1996.

[74] S. Winkler, “A Perceptual Distortion Metric for Digital Color Video,” in *in Proc. SPIE*, 1999, pp. 175–184.

[75] Z. Wang, L. Lu, and A. C. Bovik, “Video Quality Assessment Based on Structural Distortion Measurement,” 2004.

[76] M. Pinson and S. Wolf, “A New Standardized Method for Objectively Measuring Video Quality,” *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312 – 322, September 2004.

[77] S. Kanumuri, P. Cosman, A. Reibman, and V. Vaishampayan, “Modeling Packet-Loss Visibility in MPEG-2 Video,” *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 341 – 355, April 2006.

[78] *ITU-R Recommendation BS.1284-1, General methods for the subjective assessment of sound quality*, InternationalTelecommunications Union Std., 1997-2003.

[79] *ITU-T Recommendation G.107 - The E-model, a computational model for use in transmission planning*, ITU-T Std., 2000.

[80] *ITU-R Recommendation BS.1387, Method for objective measurements of perceived audio quality*, InternationalTelecommunications Union Std., 1998.

[81] R. Huber and B. Kollmeier, “PEMO-Q A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1902 –1911, November 2006.

[82] A. Barbedo, Jayme Garcia Arnal; Lopes, “A New Cognitive Model for Objective Assessment of Audio Quality,” *Journal of Audio Engineering Society*, vol. 53, no. 1/2, pp. 22–31, 2005. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13386

[83] A. Rix, J. Beerends, D.-S. Kim, P. Kroon, and O. Ghitza, “Objective Assessment of Speech and Audio Quality - Technology and Applications,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1890 –1901, November 2006.

[84] M. Kassar, B. Kervella, and G. Pujolle, “An Overview of Vertical Handover Decision Strategies in Heterogeneous Wireless Networks,” *Computer Communications*, vol. 31, no. 10, pp. 2607 – 2620, 2008.

[85] H. Kim and Y. Han, "An Opportunistic Channel Quality Feedback Scheme for Proportional Fair Scheduling," *IEEE Communications Letters*, vol. 11, no. 6, pp. 501 –503, June 2007.

[86] H. Wang, R. Katz, and J. Giese, "Policy-Enabled Handoffs Across Heterogeneous Wireless Networks," in *Proceedings of the Second IEEE Workshop on Mobile Computing Systems and Applications (WMCSA '99)*, February 1999, pp. 51 –60.

[87] Q. Wei, K. Farkas, C. Prehofer, P. Mendes, and B. Plattner, "Context-Aware Handover Using Active Network Technology," *Computer Networks*, vol. 50, no. 15, pp. 2855–2872, 2006.

[88] T.-H. Lin, P. Huang, H.-H. Chu, and C.-W. You, "Energy-Efficient Boundary Detection for RF-Based Localization Systems," *IEEE Transactions on Mobile Computing*, vol. 8, no. 1, pp. 29 –40, January 2009.

[89] *IEEE Draft Standard for Local and Metropolitan Area Networks: Media Independent Handover Services*, IEEE P802.21/D9.0 Std., January 2009.

[90] I. Papanikos and M. Logothetis, "A Study on Dynamic Load Balance for IEEE 802.11b Wireless LAN," in *Proceedings of COMCON*, 2001.

[91] W.-T. Chen, J.-C. Liu, and H.-K. Huang, "An Adaptive Scheme for Vertical Handoff in Wireless Overlay Networks," in *Proceedings of the Tenth International Conference on Parallel and Distributed Systems (ICPADS 2004)*, July 2004, pp. 541 – 548.

[92] F. Zhu and J. McNair, "Optimizations for Vertical Handoff Decision Algorithms," in *IEEE Wireless Communications and Networking Conference (WCNC 2004)*, vol. 2, March 2004, pp. 867 – 872 Vol.2.

[93] R. Tawil, O. Salazar, G. Pujolle, and J. Demerjian, "A Decision Scheme for Vertical Handoff in Overlay Wireless Networks," in *4th International Conference on Innovations in Information Technology, 2007. IIT '07.*, 2007, pp. 436 –440.

[94] W. Shen and Q.-A. Zeng, "Cost-Function-Based Network Selection Strategy in Integrated Wireless and Mobile Networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3778 –3788, November 2008.

[95] A. Calvagna and G. Di Modica, "A User-Centric Analysis of Vertical Handovers," in *Proceedings of the 2nd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH '04)*. New York, NY, USA: ACM, 2004, pp. 137–146.

[96] O. Ormond, J. Murphy, and G.-M. Muntean, "Utility-based Intelligent Network Selection in Beyond 3G Systems," in *IEEE International Conference on Communications, (ICC '06)*, vol. 4, June 2006, pp. 1831 –1836.

[97] D. Niyato and E. Hossain, "Dynamics of Network Selection in Heterogeneous Wireless Networks: An Evolutionary Game Approach," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 2008 –2017, May 2009.

[98] M. Rehan, M. Yousaf, A. Qayyum, and S. Malik, "A Cross-Layer User Centric Vertical Handover Decision Approach Based on MIH Local Triggers," in *Wireless and Mobile Networking*, ser. IFIP Advances in Information and Communication Technology, J. Wozniak, J. Konorski, R. Katulski, and A. Pach, Eds. Springer Boston, 2009, vol. 308, pp. 359–369. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-03841-9_32

[99] A. Sehgal and R. Agrawal, "QoS Based Network Selection Scheme for 4G Systems," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 560 –565, May 2010.

[100] E. Stevens-Navarro and V. Wong, "Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks," in *IEEE 63rd Vehicular Technology Conference(VTC 2006)*, vol. 2, May 2006, pp. 947 –951.

[101] Q. Song and A. Jamalipour, "An Adaptive Quality-of-Service Network Selection Mechanism for Heterogeneous Mobile Networks," *Wireless Communications*

and Mobile Computing, vol. 5, pp. 697–708, 2005. [Online]. Available: http://dblp.uni-trier.de/db/journals/wicomm/wicomm5.html#SongJ05

[102] G. Godor and G. Detari, "Novel Network Selection Algorithm for Various Wireless Network Interfaces," in *16th IST Mobile and Wireless Communications Summit*, July 2007, pp. 1 –5.

[103] F. Bari and V. Leung, "Multi-Attribute Network Selection by Iterative TOPSIS for Heterogeneous Wireless Access," in *4th IEEE Consumer Communications and Networking Conference (CCNC 2007)*, January 2007, pp. 808 –812.

[104] K. Pahlavan, P. Krishnamurty, A. Hatami, M. Ylianttila, J.-P. Makela, R. Pichna, and J. Vallström, "Handoff in Hybrid Mobile Data Networks," *IEEE Personal Communications Magazine*, vol. 7, pp. 34–47, April 2000.

[105] P. Chan, Y. Hu, and R. Sheriff, "Implementation of Fuzzy Multiple Objective Decision Making Algorithm in a Heterogeneous Mobile Environment," in *IEEE Wireless Communications and Networking Conference (WCNC2002)*, vol. 1, March 2002, pp. 332 – 336 vol.1.

[106] L. Xia, L. ge Jiang, and C. He, "A Novel Fuzzy Logic Vertical Handoff Algorithm with Aid of Differential Prediction and Pre-Decision Method," in *IEEE International Conference on Communications (ICC '07)*, June 2007, pp. 5665 –5670.

[107] Q. Guo, J. Zhu, and X. Xu, "An Adaptive Multi-Criteria Vertical Handoff Decision Algorithm for Radio Heterogeneous Network," in *IEEE International Conference on Communications (ICC 2005)*, vol. 4, May 2005, pp. 2769 – 2773 Vol. 4.

[108] S. Balasubramaniam and J. Indulska, "Vertical Handover Supporting Pervasive Computing in Future Wireless Networks," *Computer Communications*, vol. 27, no. 8, pp. 708 – 719, 2004, advances in Future Mobile/Wireless Networks and Services. [Online]. Available: http://www.sciencedirect.com/science/article/B6TYP-4B3JKTS-2/2/89e36a981ffa2f008415b094606569b4

[109] T. Ahmed, K. Kyamakya, and M. Ludwig, "Architecture of a Context-Aware Vertical Handover Decision Model and Its Performance Analysis for GPRS - WiFi Handover," in *Proceedings of the 11th IEEESymposium on Computers and Communications (ISCC '06)*, June 2006, pp. 795 – 801.

[110] L. Mokhesi and A. Bagula, "Context-aware Handoff Decision for Wireless Access Networks Using Bayesian Networks," in *Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists SAICSIT '09:*. New York, NY, USA: ACM, 2009, pp. 104–111.

[111] T. Henderson, "Host Mobility for IP Networks: A Comparison," *IEEE Network*, vol. 17, no. 6, pp. 18 – 26, November 2003.

[112] D. Famolari and S. Baba, "Performance Evaluation of the ITSUMO Mobility Protocols for RTP/UDP Multimedia Sessions Across Subnet Boundaries," in *IEEE International Conference on Communications (ICC 2001)*, vol. 8, 2001, pp. 2483 –2487 vol.8.

[113] W. Eddy, "At What Layer Does Mobility Belong?" *IEEE Communications Magazine*, vol. 42, no. 10, pp. 155 – 159, October 2004.

[114] A. Mishra, M. Shin, and W. Arbaugh, "An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process," *SIGCOMM Comput. Commun. Rev.*, vol. 33, no. 2, pp. 93–102, 2003.

[115] H. Velayos and G. Karlsson, "Techniques to Reduce IEEE 802.11b Handoff Time," in *Proceedings of the International Conference on Computer Communications (ICC 2004)*, 2004.

[116] C.-T. Chou and K. G. Shin, "An Enhanced Inter-Access Point Protocol for Uniform Intra and Intersubnet Handoffs," *IEEE Transactions on Mobile Computing*, vol. 4, pp. 321–334, 2005.

[117] G. Lampropoulos, N. Passas, L. Merakos, and A. Kaloxylos, “Handover Management Architectures in Integrated WLAN/Cellular Networks,” *IEEE Communications Surveys Tutorials*, vol. 7, no. 4, pp. 30 – 44, 2005.

[118] H. Balakrishnan, S. Seshan, R. H. Katz, and Y. H. Katz, “Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks,” 1995.

[119] I. Radovanovic, R. Verhoeven, and J. Lukkien, “Improving TCP Performance Over Last-Hop Wireless Networks for Live Video Delivery,” *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 1139 –1147, August 2008.

[120] C. Perkins, “IP Mobility Support for IPv4,” RFC 3344 (Proposed Standard), Internet Engineering Task Force, August 2002, updated by RFC 4721. [Online]. Available: http://www.ietf.org/rfc/rfc3344.txt

[121] D. Johnson, C. Perkins, and J. Arkko, “Mobility Support in IPv6,” RFC 3775 (Proposed Standard), Internet Engineering Task Force, June 2004. [Online]. Available: http://www.ietf.org/rfc/rfc3775.txt

[122] A. Campbell, J. Gomez, S. Kim, C.-Y. Wan, Z. Turanyi, and A. Valko, “Comparison of ip micromobility protocols,” *IEEE Wireless Communications*, vol. 9, no. 1, pp. 72 – 82, February 2002.

[123] F. Chiussi, D. Khotimsky, and S. Krishnan, “Mobility Management in Third-Generation All-IP Networks,” *IEEE Communications Magazine*, vol. 40, no. 9, pp. 124 – 135, September 2002.

[124] A. G. Valkó, “Cellular IP: A New Approach to Internet Host Mobility,” *SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 1, pp. 50–65, 1999.

[125] R. Ramjee, K. Varadhan, L. Salgarelli, S. Thuel, S.-Y. Wang, and T. La Porta, “HAWAII: A Domain-Based Approach for Supporting Mobility in Wide-Area Wireless Networks,” *IEEE/ACM Transactions on Networking*, vol. 10, no. 3, pp. 396 –410, jun. 2002.

[126] N. Montavont and T. Noel, "Handover Management for Mobile Nodes in IPv6 Networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 38 – 43, August 2002.

[127] H. Soliman, C. Castelluccia, K. E. Malki, and L. Bellier, "Hierarchical Mobile IPv6 Mobility Management (HMIPv6)," RFC 4140 (Experimental), Internet Engineering Task Force, August 2005, obsoleted by RFC 5380. [Online]. Available: http://www.ietf.org/rfc/rfc4140.txt

[128] R. Jan, T. Raleigh, D. Yang, L. F. Chang, C. Graff, M. Bereschinsky, and M. Patel, "Enhancing Survivability of Mobile Internet Access Using Mobile IP with Location Registers," in *Proceedings of the Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '99)*, vol. 1, March 1999, pp. 3 –11 vol.1.

[129] P. Nikander, J. Lundberg, C. Candolin, and T. Aura, "Homeless Mobile IPv6," Internet Engineering Task Force, Tech. Rep., February 2001, internet Draft, work in progress.

[130] R. Moskowitz and P. Nikander, "Host Identity Protocol (HIP) Architecture," RFC 4423 (Informational), Internet Engineering Task Force, May 2006. [Online]. Available: http://www.ietf.org/rfc/rfc4423.txt

[131] D. Le, X. Fu, and D. Hogrefe, "A Review of Mobility Support Paradigms for the Internet," *IEEE Communications Surveys Tutorials,*, vol. 8, no. 1, pp. 38 –51, 2006.

[132] D. Funato, K. Yasuda, and H. Tokuda, "TCP-R: TCP Mobility Support for Continuous Operation," in *Proceedings of the International Conference on Network Protocols*, October 1997, pp. 229 –236.

[133] D. Maltz and P. Bhagwat, "MSOCKS: An Architecture for Transport Layer Mobility," in *Proceedings of IEEE Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '98)*, vol. 3, Apr. 1998, pp. 1037 –1045 vol.3.

[134] M. Stangel and V. Bharghavan, “Improving TCP Performance in Mobile Computing Environments,” in *IEEE International Conference on Communications (ICC 1998)*, vol. 1, June 1998, pp. 584 –589 vol.1.

[135] A. C. Snoeren and H. Balakrishnan, “An End-to-End Approach to Host Mobility,” in *MOBICOM*, 2000, pp. 155–166. [Online]. Available: http://dblp.uni-trier.de/db/conf/mobicom/mobicom2000.html#SnoerenB00

[136] W. E. Y Swami, K Le, “Lightweight Mobility Detection and Response (LMDR) Algorithm for TCP,” Internet Draft (work in progress), draft-swami-tcp-lmdr-06, 2005.

[137] A. Bakre and B. Badrinath, “I-TCP: Indirect TCP for Mobile Hosts,” in *Proceedings of the 15th International Conference on Distributed Computing Systems*, may. 1995, pp. 136 –143.

[138] T. Goff, J. Moronski, D. Phatak, and V. Gupta, “Freeze-TCP: A True End-to-End TCP Enhancement Mechanism for Mobile Environments,” in *Proceedings of the Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000)*, vol. 3, March 2000, pp. 1537 –1545 vol.3.

[139] S. Liang and D. Cheriton, “TCP-RTM: Using TCP for Real Time Multimedia Applications,” in *International Conference on Network Protocols*, 2002.

[140] I. Kim, Y. Kim, M. Kang, J. Mo, and D. Kwak, “TCP-MR: Achieving End-to-End Rate Guarantee for Real-Time Multimedia,” in *Second International Conference on Communications and Electronics (ICCE 2008)*, June 2008, pp. 80 –85.

[141] K. Brown and S. Singh, “M-UDP: UDP for Mobile Cellular Networks,” *SIGCOMM Comput. Commun. Rev.*, vol. 26, no. 5, pp. 60–78, 1996.

[142] R. Stewart, “Stream Control Transmission Protocol,” RFC 4960 (Proposed Standard), Internet Engineering Task Force, September 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4960.txt

[143] S. J. Koh, M. J. Chang, and M. Lee, "mSCTP for Soft Handover in Transport Layer," *IEEE Communications Letters*, vol. 8, no. 3, pp. 189 – 191, March 2004.

[144] R. Stewart, Q. Xie, M. Tuexen, S. Maruyama, and M. Kozuka, "Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration," RFC 5061 (Proposed Standard), Internet Engineering Task Force, September 2007. [Online]. Available: http://www.ietf.org/rfc/rfc5061.txt

[145] I. Aydin, W. Seok, and C.-C. Shen, "Cellular SCTP: A Transport-Layer Approach to Internet Mobility," in *Proceedings of the 12th International Conference on Computer Communications and Networks (ICCCN 2003)*, October 2003, pp. 285 – 290.

[146] S. Fu, M. Atiquzzaman, L. Ma, W. I. andYong Jin Lee, J. S. Jones, and S. Lu, *TraSH: A Transport Layer Seamless Handover for Mobile Networks*, Telecommunications and Networks Research Lab, School of Computer Science,University of Oklahoma, January 2004.

[147] E. Kohler, M. Handley, and S. Floyd, "Datagram Congestion Control Protocol (DCCP)," RFC 4340 (Proposed Standard), Internet Engineering Task Force, March 2006, updated by RFCs 5595, 5596. [Online]. Available: http://www.ietf.org/rfc/rfc4340.txt

[148] E. Kohler, "Generalized Connections in the Datagram Congestion Control Protocol," Internet Draft (work in progress), draft-kohler-dccp-mobility-02, June 2006.

[149] B. Landfeldt, T. Larsson, Y. Ismailov, and A. Seneviratne, "SLM, A Framework for Session Layer Mobility Management," in *Proceedings of the Eight International Conference on Computer Communications and Networks*, 1999, pp. 452 –456.

[150] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol," IETF, RFC 3261, Jun. 2002.

[151] P. Eronen, "IKEv2 Mobility and Multihoming Protocol (MOBIKE)," RFC 4555 (Proposed Standard), Internet Engineering Task Force, June 2006. [Online]. Available: http://www.ietf.org/rfc/rfc4555.txt

[152] C. Kaufman, "Internet Key Exchange (IKEv2) Protocol," RFC 4306 (Proposed Standard), Internet Engineering Task Force, December 2005, updated by RFC 5282. [Online]. Available: http://www.ietf.org/rfc/rfc4306.txt

[153] K. Andersson, D. Granlund, and C. Ahlund, "M4: Multimedia Mobility Manager: A Seamless Mobility Management Architecture Supporting Multimedia Applications," in *Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia (MUM '07)*. New York, NY, USA: ACM, 2007, pp. 6–13.

[154] S. Bose and A. Kannan, "Adaptive Multipath Multimedia Streaming Architecture for Mobile Networks with Proactive Buffering Using Mobile Proxies," *Journal of Computing and Information Technology CIT 15*, vol. 3, pp. 215–226, 2007.

[155] D. D. Nguyen, Y. Xia, M. N. Son, C. K. Yeo, and B. S. Lee, "A Mobility Management Scheme with QoS Support for Heterogeneous Multihomed Mobile Nodes," in *IEEE Global Telecommunications Conference (GLOBECOM 2008)*, November 2008, pp. 1 –6.

[156] A. Balachandran, P. Bahl, and G. Voelker, "Hot-Spot Congestion Relief in Public-Area Wireless Networks," in *Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications*, 2002, pp. 70 – 80.

[157] Y. Bejerano, S.-J. Han, and L. Li, "Fairness and Load Balancing in Wireless LANs Using Association Control," *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 560 –573, June 2007.

[158] T. Scully and K. N. Brown, "Wireless LAN Load Balancing with Genetic Algorithms," *Knowledge-Based Systems*, vol. 22, no. 7, pp. 529 – 534, 2009, artificial Intelligence 2008 - AI-2008, The twenty-eighth SGAI International

Conference on Artificial Intelligence. [Online]. Available: http://www.sciencedirect.com/science/article/B6V0P-4VBDK7H-B/2/f01d459fe490964fe932620fe2197880

[159] G. Nunzi, S. Schuetz, and M. Brunner, "Design and Evaluation of Distributed Load-Balancing for Wireless Networks," in *10th IFIP/IEEE International Symposium on Integrated Network Management (IM '07)*, May 2007, pp. 478 –486.

[160] J. Ha, J. Kim, J.-U. Kim, and S.-H. Kim, "Dynamic Load Balancing Architecture in Heterogeneous Wireless Network Environment," in *9th International Symposium on Communications and Information Technology (ISCIT 2009)*, September 2009, pp. 248 –253.

[161] L. Yun, J. Hong, L. Xi, and C. Daojin, "A Novel Load Balancing Algorithm in IEEE 802.11 Wireless LANs with Cell Breathing," in *5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCom '09)*, 2009, pp. 1 –4.

[162] K. Ishizu and H. Harada, "A Load-Balancing Framework for Cognitive Wireless Network to Coexist with Legacy WiFi Systems," in *IEEE 69th Vehicular Technology Conference, 2009. VTC Spring 2009*, April 2009, pp. 1 –5.

[163] E. H. Ong and J. Khan, "An Integrated Load Balancing Scheme for Future Wireless Networks," in *IEEE GLOBECOM Workshops*, November 2008, pp. 1 –6.

[164] H. Velayos, V. Aleo, and G. Karlsson, "Load Balancing in Overlapping Wireless LAN Cells," in *IEEE International Conference on Communications*, vol. 7, June 2004, pp. 3833 – 3836 Vol.7.

[165] J. Chen, T. Rappaport, and G. de Veciana, "Iterative Water-filling for Load-balancing in Wireless LAN or Microcellular Networks," in *IEEE 63rd Vehicular Technology Conference (VTC 2006)*, vol. 1, May 2006, pp. 117 –121.

[166] D. Sarkar and U. Sarkar, "Balancing Load of APs by Concurrent Association of Every Wireless Node with Many APs," in *Fifth International Conference on Networking and Services (ICNS '09)*, April 2009, pp. 320 –325.

[167] M. Allman, V. Paxson, and E. Blanton, *TCP Congestion Control*, RFC 5681 (Draft Standard), Internet Engineering Task Force, September 2009. [Online]. Available: http://www.ietf.org/rfc/rfc5681.txt

[168] M. Handley, S. Floyd, J. Padhye, and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification," RFC 3448 (Proposed Standard), Internet Engineering Task Force, January 2003, obsoleted by RFC 5348. [Online]. Available: http://www.ietf.org/rfc/rfc3448.txt

[169] S. Shakkottai, E. Altman, and A. Kumar, "Multihoming of Users to Access Points in WLANs: A Population Game Perspective," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 6, pp. 1207 –1215, August 2007.

[170] X. Zhu, P. Agrawal, J. P. Singh, T. Alpcan, and B. Girod, "Distributed Rate Allocation Policies for Multihomed Video Streaming Over Heterogeneous Access Networks," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 752–764, 2009.

[171] R. Chow, C. wei Lee, and J. Liu, "Traffic Dispersion Strategies for Multimedia Streaming," in *Proceedings of the Eighth IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS 2001)*, 2001, pp. 18 –24.

[172] A. El Al, T. Saadawi, and M. Lee, "A Transport Layer Load Sharing Mechanism for Mobile Wireless Hosts," in *Proceedings of the Second IEEE Annual Conference onPervasive Computing and Communications Workshops*, March 2004, pp. 87 – 91.

[173] Y. Wang, L.-P. Chau, and K.-H. Yap, "Bit-Rate Allocation for Broadcasting of Scalable Video Over Wireless Networks," *IEEE Transactions onBroadcasting*, vol. 56, no. 3, pp. 288 –295, September 2010.

[174] V. Goyal, "Multiple Description Coding: Compression Meets the Network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74 –93, September 2001.

[175] V. Vaishampayan and S. John, "Balanced Interframe Multiple Description Video Compression," in *Proceedings of the International Conference on Image Processing (ICIP 1999)*, vol. 3, 1999, pp. 812 –816 vol.3.

[176] V. Vaishampayan, "Design of Multiple Description Scalar Quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821 –834, May 1993.

[177] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple Description Coding for Video Using Motion Compensated Prediction," in *Proceedings of the International Conference on Image Processing (ICIP 1999)*, vol. 3, 1999, pp. 837 –841 vol.3.

[178] T. Tillo, M. Grangetto, and G. Olmo, "Redundant Slice Optimal Allocation for H.264 Multiple Description Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 59 –70, January 2008.

[179] D. Rakhmatov, S. Vrudhula, and D. Wallach, "A Model for Battery Lifetime Analysis for Organizing Applications on a Pocket Computer," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 11, no. 6, pp. 1019 – 1030, 2003.

[180] S.-B. Lee, G.-M. Muntean, and A. Smeaton, "Performance-Aware Replication of Distributed Pre-Recorded IPTV Content," *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 516 –526, June 2009.

[181] NS-2 Network Simulator-2, http://www.isi.edu/nsnam/ns.

[182] NO Ad-Hoc Routing Agent (NOAH), http://icapeople.epfl.ch/widmer/uwb/ns-2/noah/.

[183] M. Fioare, "Realistic Radio Patch for NS-2," http://www.tlc-networks.polito.it/fiore.

[184] G.-M. Muntean and N. Cranley, “Resource Efficient Quality-Oriented Wireless Broadcasting of Adaptive Multimedia Content,” *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 362 –368, March 2007.

[185] MSU Video Quality Measure, http://www.compression.ru.

[186] H. J. Kim and S. G. Choi, “A Study on a QoS/QoE Correlation Model for QoE Evaluation on IPTV Service,” in *12th International Conference on Advanced Communication Technology (ICACT 2010)*, vol. 2, February 2010, pp. 1377 –1382.

[187] SIP implementation for NS-2, http://www.dcc.fc.up.pt/ rprior/ns/.

[188] Z. Wang, A. Bovik, and L. Lu, “Why is Image Quality Assessment so Difficult?” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’02).*, vol. 4, 2002, pp. IV–3313 – IV–3316 vol.4.

[189] M. H. Kurz, Ludwik; Benteftifa, *Analysis of variance in statistical image processing*. Cambridge University Press, 1997. [Online]. Available: http://dx.doi.org/10.1017/CBO9780511530166