2007:064

MASTER'S THESIS

A Study on

Adaptive Real Time Video over LTE

A.N.M. Zaheduzzaman Sarker

Luleå University of Technology

Master Thesis, Continuation Courses
Computer Science and Engineering
Department of Computer Science and Electrical Engineering
Division of Media Technology

2007:064 - ISSN: 1653-0187 - ISRN: LTU-PB-EX--07/064--SE

Abstract

LTE, The next generation mobile network system by 3GPP, only allows IP-based transport. Traditional telephony services such as voice and video real-time communication will be provided through the use of IMS (IP Multimedia Subsystem) Multimedia Telephony (MTSI) which imposes high demands on the transport channel. In a packet switched network any intermediary node can act like a congestion bottle neck leading to delay or packet loss. In this scenario a fast end to end adaptation scheme at the media layer can play a vital role to secure service performance by ensuring high quality throughout the session, even in extreme conditions, and keeps the service alive avoiding session termination. This thesis report describes a study of perceived video quality improvement using end to end congestion control depending on both “pre-warning” for congestion and packet loss observed at receiver in IMS video telephony over 3GPP LTE system according to 3GPP TS 26.114 V7.0.0. In the process of achieving the goal a new adaptation state machine was developed with different strategies to address the early congestion notification and packet losses. Simulations has been carried out with different adaptation schemes to compare results and benefits of newly proposed scheme. This thesis shows the effectiveness of early congestion notification for real-time video transmission and reveals some important aspects of adaptation procedure which need to be considered while designing adaptation for video traffic to avoid the congestion in a LTE network. The results found in the study suggest the inclusion of ECN for UDP traffic for future network like LTE.

Acknowledgement

At first the Author wants to thank the all mighty creator for giving such a wonderful life to enjoy. Next he want to express his heartily gratitude to Daniel Enström, Supervisor at Ericsson, for being such a great advisor and guide throughout the thesis work and helping the author to solve all kind of problems related or unrelated to this thesis work and to Stefan Håkansson for believing in him and allowing him to do the thesis at Ericsson. He also wants to thank Kåre Synnes, thesis examiner, without whose inspiration and help this work could not have been completed. Further more the author is grateful to Ingemar Johansson, Fredrik Hultin for their ideas and discussion to implement the solution in simulator and all those at Ericsson who have been so nice to him during his stay at Ericsson. And last but not the least he wants to especially thank his loving wife and parents who stood beside him in every situation while doing this work.

Table of Contents

List of Figures

List of Tables

Abbreviations

3GPP	3rd Generation Partnership Project
ACK	Acknowledgement
AE	with Adaptation and ECN
AIMD	Additive Increase Multiplicative Decrease
AIMD	Adaptive Increase Adaptive Decrease
AIPD	Additive Increase Proportional Decrease
AVC	Advanced Video Coding
AVP	Audio Visual Profile
AVPF	Audio Visual Profile with Feedback
CE	Congestion Experienced
DCCP	Dynamic Congestion Control Protocol
DDAI	Distance Dependent Adaptive Increase
DWAI/LDMD	Distance Weighted Additive Increase and Loss dependent Multiplicative Decrease
E_nodeB	Evolved Node B
ECN	Explicit Congestion Notification
FDFU	Fast bit-rate Decrease Fast bit-rate restore Up
FDSU	Fast bit-rate Decrease Slow bit-rate restore Up
GBR	Guaranteed Bit-Rate
HTTP	Hyper Text Transfer Protocol
IDR	Instantaneous Decoder Refresh
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
IP	Internet Protocol
ITU-T	International Telecommunication Union-Telecommunication
LIMD/H	Linear Increase Multiplicative Decrease with History
LTE	Long Term Evolution
MBR	Maximum Bit-rate
MIMO	Multiple Input Multiple Output
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
NA	No Adaptive increase
NACK	Negative Acknowledge
NE	No ECN based rate control
NN	No Congestion control No ECN
OFDM	Orthogonal Frequency Division Multiplexing
PLR	Packet Loss Rate
PSNR	Peck Signal to Noise Ratio
QoS	Quality of Service
RCP	Rate Control Protocol
RR	Receiver Report
RTCP	Real Time Control Protocol
RTP	Real Time Protocol
RTT	Round Trip Time
SAE	System Architecture Evolution
SC-FDMA	Single Carrier Frequency Division Multiple Access
SDFU	Slow bit-rate Decrease Fast bit-rate restore Up

SDSU	Slow bit-rate Decrease Slow bit-rate restore Up
SR	Sender Report
TCP	Transport Control Protocol
TFRC	TCP Friendly Rate Control
TMMBN	Temporary Maximum Media Bit-rate Notification
TMMBR	Temporary Maximum Media Bit-rate Request
TR	Technical Report
TS	Technical Specification
UDP	User Datagram Protocol
VCEG	Video Coding Expert Group
VoIP	Voice over IP

Chapter 01

1. Introduction

This chapter starts with a brief introduction of this thesis to the reader followed by problem statement, goal formation and scope of this thesis. The chapter ends with research methodology and by stating the structure of the thesis.

1.1 Overview

This thesis work presents an application layer adaptation mechanism for real-time communication services over IP (Internet Protocol). It is especially targeted towards the next generation mobile network standardized in 3GPP (3rd Generation Partnership Project), LTE (UTRA-UTRAN Long Term Evolution). Since LTE only allows IP-based transport, traditional telephony services such as voice and video real-time communication will be provided through the use of IMS (IP Multimedia Subsystem) Multimedia Telephony (MTSI). MTSI is an IP-based real time communication service with high demands on the transport channel. This is required in order to provide a high quality, both media quality in terms of bit-rate and packet loss rates and conversational quality in terms of latency. In this scenario a fast end to end adaptation scheme at the media layer can play a vital role to secure service performance by ensuring high quality throughout the session, even in extreme conditions, and keeps the service alive avoiding session termination. Especially regulating the transmission of media like video, which requires higher bandwidth relative to voice or other media, can help to improve the congested situation by releasing important network resources.

1.2 Problem statement

The need for communication has always been the driving force behind the invention of technologies which is bringing the people closer day by day. The telecommunication industry began with a wired connection and then the recent advancement in different technologies has made it possible to communicate using wireless technologies. Invention of computer, software, hardware, micro chips has changed the whole concept of communication. Today we see a blend of wired and wireless communication network using heterogeneous technologies.

The heterogeneity in communication networks has not only opened the door of different formats of communication but also has induced lot of issues that need to be addressed in order to provide high quality communications. Congestion is one of the major issues among them. A modern communication network is built using a number of well connected devices or nodes, which have limited local capacity and resources. Currently, two different transport paradigms are used – i) circuit switched transport and ii) packet switched transport. In a circuit switched domain a connection is established between two or more communicating peers before the actual communication takes place where as in a packet switched network no prior connection is established between peers. In this case both peers send data to each other in a form called "Packet" and these packets are routed by the different nodes in the network and eventually reach the receiver. The intermediate nodes between the communicating peers are suppose to route the packets to a suitable next node but can also drop the packets if necessary. Congestion, which is a sudden state of this communication network where one or more nodes reach their capacity limit and as a result they either drop the incoming packet or buffer them for a later transmission, induces delay in the arrival of packets at the receiver. None of these effects of congestion are desirable for media transport hence counter measures should be taken to prevent the occurrence of congestion in communication network.

Taking user experience into consideration it has been found that delay in data arrival and lost traffic can induce stuttering or stalling in audio and for video it will cause jerkiness, blurring, or a corrupted picture. Again if the network does not come out of congestion that will eventually lead to more delay and data loss. Therefore, it is desirable for the real time media application to adapt with the congestion to improve quality of user experience and help the network to reach desired condition as early as possible and keep its stability.

The transport layer protocol for real-time media application is UDP (User Datagram Protocol) [1] which has been preferred over TCP (Transmission Control Protocol) [2] for its suitability of being connectionless meaning that it does not ensure the ordered and reliable delivery of packet as TCP but ensures a timely delivery of the data packet. To support this nature of connectionless UDP does not allow any congestion control [1]. Hence, the UDP traffic will suffer more packet loss when the network is congested compared with TCP. For real time media or streaming media application RTP (Real Time Protocol) [3] is used over UDP along with RTCP (Real Time Control Protocol) which provides timely feedback to the sender of the media transport characteristics. Even if congestion control is not possible for UDP traffic, adaptation can be done in application layer of IP protocol suit depending on the feedback matrices (jitter, packet loss etc.) provided by RTCP. This will ultimately help the UDP traffic to backoff when the network is congested.

A general approach of handling congestion is to backoff in transmission bit-rate depending on the reported characteristics via RTCP. This has a clear disadvantage as only the end terminals react on the congestion but intermediary nodes like routers, gateways and base stations do not react on the current congested situations although they can sense the congestion prior to the end terminals. TCP addresses this scenario with the help of it's built in flow control mechanism. TCP also have the possibility to react to so-called ECN (Early Congestion Notification) messages [4]. But there is need of transport layer support to deploy ECN to have any effect on application behavior to avoid congestion before packets have been dropped. Though RFC 3168 [4] comes with a support mechanism for TCP but the support for ECN to use with UDP is unspecified. It is believed that features like ECN for UDP traffic can certainly help to avoid the congestion in the network if proper reactions mechanisms to backoff UDP traffic are specified.

With a vision of "All-IP" [5] network and increase of IP based services, 3GPP (3^{rd} Generation Partnership Project)[6] has decided to build their next mobile network system, LTE (Long Term Evolution), with an objective "to develop a framework for the evolution of the 3GPP radio-access technology towards a high-data-rate, low-latency and packet-optimized radio-access technology". Further, 3GPP has standardized multimedia telephony over IMS (IP Multimedia Subsystem) ([7], [8], [9]) as a pure packet-switched real-time communication service. Being a packet switched network, LTE gives us the opportunity to have more control over different network parameters (i.e. bandwidth, data rate etc) and more effective radio resource management specially optimized for IP traffic. The use of more robust retransmission procedure [10] and link adaptation technique [11] will also mitigate the possible loss over radio interface due to bad radio coverage. Although LTE aims at higher data rates up to 100Mbps and a user plane latency below 5ms (milliseconds) [12], a highly loaded mobile network or a user positioned in a bad radio coverage area the network still cannot guarantee perfect transport characteristics all the time. Hence, there is a need for the application on the mobile phone to monitor and adapt to the current transport characteristics to optimize the quality and keep the service attractive even in difficult radio environment.

In light of this, the research question of this thesis can be formulated as:

What quality improvement can be achieved in an IMS multimedia telephony session by adapting the video transmission rate in response to the ECN for UDP traffic and packet loss observed at end terminals in 3GPP LTE System?

1.3 Goals

This thesis will investigate the improvement of perceived video quality using end to end congestion control in IMS video telephony over 3GPP LTE system according to 3GPP TS 26.114 V7.0.0. The focus of the thesis will be to design an adaptation state machine to control the bit-rate of video transmission in a congested network and analyzing the feasibility of using ECN (Early Congestion Notification) for UDP traffic to avoid quality distortion due to packet losses in the network. The result of this thesis will provide an adaptation mechanism which can be used to address the sudden network load increase triggering ECN messages for UDP traffic.

1.4 Scope

In this thesis 3GPP Technical Specification TS 26.114 V7.0.0 will be followed. The studied scenario will contain video conversion based on IP Multimedia Subsystem (IMS) which has been standardized by 3GPP for multimedia telephony. Though video conversation contains both video and audio transmission, In the thesis emphasis will only be given to video transmission and adaptation as video transmission requires more bandwidth than that of audio transmission and in congested network where there is a necessity of releasing acquired resources to improve the congested situation, adapting video (i.e. going down in bit-rate) have a greater congestion alleviating potential since video codecs span over a much larger bandwidth than speech codecs. A simple policy for ECN marking will be implemented and used. The results will be produced in terms of perceived video quality which will include both an objective and a subjective video quality analysis. The result will give guidelines for rate adaptation in response to the ECN for UDP traffic to avoid packet loss hence improving perceived video quality for IMS video telephony over 3GPP LTE systems.

1.5 Research Methodology

The following structured research approach will be followed during this thesis work-

1.5.1 Investigate current state of art of handling the congestion.

1.5.2 Design the adaptation state machine based on the ideas gathered by study of literature and technological description.

1.5.3 Design and implement the necessary tools to deploy the adaptation state machine

1.5.4 Developing the simulation environment

1.5.5 Testing and collecting result

1.5.6 Analyzing the experimental result

1.5.7 Publishing the research outcome and future works.

1.6 Structure of the Thesis

In this book chapter 2 details the background study and literature review. It discusses the studied technologies, their advantages and pitfalls and reviews the different approaches adopted by different researcher community and tries to identify different issues regarding the stated problem.

Chapter 3 presents different adaptation approaches answering the questions of who, where, how and when adaptation can be done to provide high quality of service in extreme cases.

Chapter 4 outlines the design and implementation of adaptation state machines of media layer adaptation for IMS multimedia telephony. It also presents different scenarios using the adaptation.

Chapter 5 discusses the simulation setup and evaluates the experimental results. In this chapter, the results has been plotted and compared with base cases and findings from the analysis have been noted.

Chapter 06 summarizes the research work and proposes future work in the area.

Chapter 02

2. Background

This chapter describes different theories and technologies which have been studied and used in this thesis work. It starts by describing 3GPP technologies and continues with IETF protocols, adaptation, video codecs and video quality analysis tools. At the end of this chapter related work done by the research community regarding adaptation and congestion control has been summarized.

2.1 3GPP Technologies

2.1.1 LTE (UTRA and UTRAN Long Term Evolution)

In order to ensure the competitiveness of UMTS (Universal Mobile Telecommunication System) over other emerging technologies for next 10 years and beyond, 3GPP (3rd Generation Partnership Project) has introduced a study item to investigate the concept of UMTS Long Term Evaluation (LTE). The Objective of this study is to build a "framework towards high data-rate, low latency and packet optimized radio access technology" for the evolution of the 3GPP radio-access technologies. LTE has an aggressive set of requirements which can be found in [12]. The LTE system will operate only in the packet-switched domain. Hence all services will be running over IP which requires a replacement of the current circuit-switched voice communication to an IP based voice service. This transition will provide a single application domain serving whole range of access networks and will result in a convergence between different access technologies and networks into an All-IP network [5].

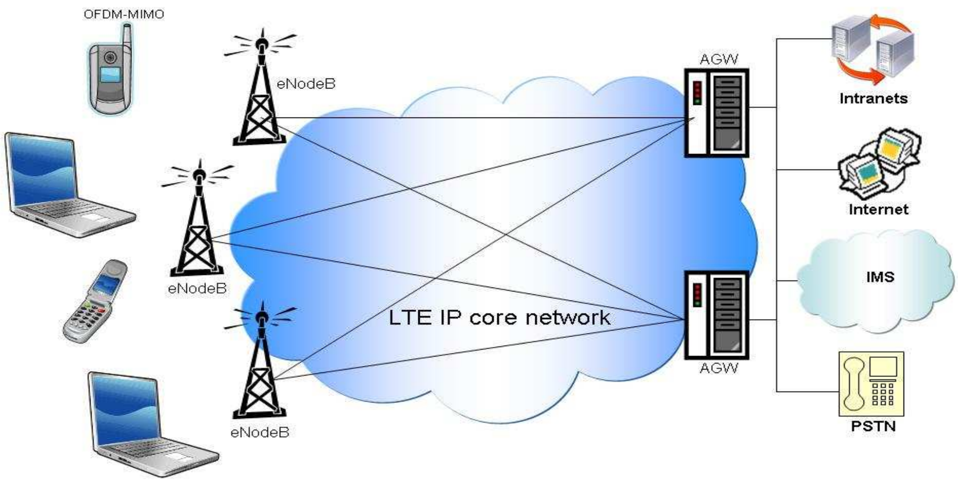


Figure 2-1: Overview of LTE

LTE will operate on a frequency bandwidth starting from 1.25MHz up to 20MHz. The use of advanced technologies like OFDM (Orthogonal Frequency Division Multiplexing) for downlink, SC-FDMA (Single Carrier Frequency Division Multiple Access) along with MIMO (Multiple Input Multiple Output) antennas will allow LTE of 3G to achieve high data rates up to 100 Mbps in the downlink and peak data rate up to 50 Mbps in the uplink with high spectral efficiency. More over, LTE will operate over a simple network architecture called SAE (System Architecture Evolution) which compared to the current network architecture will allow both the operator and the users to enjoy reduced cost per service with an evolved QoS [13].

2.1.2 IMS (IP Multimedia Subsystem)

IMS defines a control plane for IP-based service to control, route, charge and prioritize IP traffic in multimedia services. It is an access independent technology aimed to aid the deployment of ALL-IP networks. This technology will make most popular Internet services and application available anywhere and everywhere. IMS defines QoS, charging and integration of all the services in packet switched domain. Both 3GPP (3rd Generation Partnership Project) and 3GPP2 (3rd Generation Partnership Project 2) has defined their own set of standardization defining IMS but the core functions and protocols are the same. According to 3GPP TS 22.228 release 6 [14], IMS has the following requirements-

i) Support for a mechanism to negotiate Quality of Services (QoS)

ii) Support for variety of different media types and one or more IP multimedia applications in one IP Multimedia Session.

iii) Support for interworking with existing voice and data networks both for fixed and mobile users

iv) Support for roaming

v) Support for strong control imposed by the operator with respect to the services delivered to the end-user.

vi) Support for rapid service creation without requiring standardization

vii) Support for access independence.

The IMS architecture has been standardized in terms of functionalities so depending on implementation, one node can perform more then one functionalities using standard interfaces. A detailed IMS architectural description can be found in [15].

2.1.3 IMS Multimedia Telephony

IMS multimedia telephony is a communication service which has been standardized within 3GPP. As the name suggest IMS multimedia telephony aims to provide communication with multiple media (i.e. video, text, file sharing etc.) in addition of providing the most common service i.e. voice communication. IMS multimedia telephony has some challenging requirements such as it should able to provide service at least as good as the legacy circuit switched service in terms of coverage, capacity and quality, should have support for interoperability in multi-vendor and multi-operator environment and should be consistent in behavior for different IP networks and domains. In other words, the user will not be aware of the underlying access technologies while using the service and should enjoy same QoS as legacy systems. From the users' point of view, one of the attractive features of IMS multimedia telephony will be to be able to add or drop any kind of media (e.g. video, text) in the conversion session dynamically. The operators will also be able take advantages of the facilities provided by IMS architecture by having interoperability, reliability, security and meeting the users expectation. 3GPP has written several Technical Specifications (TS) and Technical Report (TR) while standardizing IMS multimedia telephony. 3GPP TS 26.114[9], 3GPP TS 24.229[16], 3GPP TR 26.914[17], 3GPP TR 22.973[18] respectively defines media handling, call control, optimization opportunities and services for IMS multimedia telephony[8].

2.2 Protocols

2.2.1 UDP (User Datagram Protocol)

UDP [1] is one of the main transport layer protocols in IP protocol stack. It is aimed to use with datagrams (data, in small fragmented packet entity) in a packet-switched interconnected computer networks environment with a minimum protocol mechanism. Except from multiplexing/de-multiplexing and simple error checking it does not add any thing to IP. As it maintains a connectionless state it does not provide any reliability in terms of ordered delivery and duplicate protection and does not also provide any means of congestion control. It is basically suitable for transporting delay sensitive data (e.g. voice, video) which have real time characteristics.

2.2.2 TCP (Transport Control Protocol)

TCP [19] is the most widely used transport layer protocol of the IP protocol stack. It guarantees ordered and reliable delivery of packets at the cost of delay. TCP assumes no or minimum reliability from the lower layer protocols so it maintains reliable inter process communication between communicating nodes. More over, TCP has its built in congestion management algorithm such as fast retransmission and fast recovery [19] which ensures that packet loss recovery due to congestion or a temporary link failure is fast. It is most suitable for the services where ordered reliable delivery is a requirement.

2.2.3 RTP (Real Time Protocol)/RTCP (Real Time Control Protocol)

RTP is a network protocol defined in RFC 3550 [3] by IETF (Internet Engineering Task Force) to provide end-to-end delivery for data that has real time characteristics which generally runs over UDP (figure 2-2).

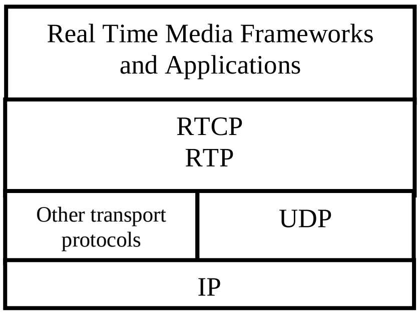


Figure 2-2: RTP architecture

RTP can be used in both unicast and multicast multimedia sessions. It provides sequence numbers which allows the receiver to reconstruct the senders sending sequence and makes it also possible to identify the type of the data received along with the source of the data being received.
RTCP is an auxiliary protocol comes in couple with RTP. It works as a control protocol for RTP. Though RTP does not guarantee any kind of QoS, the use of RTCP with RTP provides certain extension of the mechanism to monitor the quality of the data transmission. The sequence number and timestamp enables us to compute some important quality matrices such as packet loss ratio and delay jitter which can be reported back to the sender of the data with the help of timely feedback mechanism in terms of a SR (Sender Report) or a RR (Receiver Report)[3]. As RTP can be used with various types of applications, a profile document defining extensions is needed for a particular environment and class of media application.

2.2.3.1 RTP/AVP (Audio Visual Profile)

RTP/AVP is a profile defined in RFC 3551 [20] detailing interpretation of the generic fields of RTP and associate control protocol RTCP for audio and video conference with minimum control. In addition, this profile also defines a set of encodings and payload type for audio and video. The RTCP timing rule is relaxed in this profile.

2.2.3.2 RTP/AVPF (Audio Visual Profile with Feedback)

This profile defined in RFC 4585 [21] extends RTP/AVP and allows more relaxation in feedback timing without exceeding the bandwidth limitation for RTCP traffic especially in a single unicast multimedia session. Feedback from the data receiver can now be used more effectively by implementing fast adaptation and repair mechanisms as it is possible to immediately send a feedback to report a particular event. This profile also introduces two more feedback modes named "ACK mode" (only for point to point communication and heavy downlink bandwidth provided) and "NACK mode", statistically, for reporting any loss event as soon as possible. This profile document defines three kinds of feedback messages –

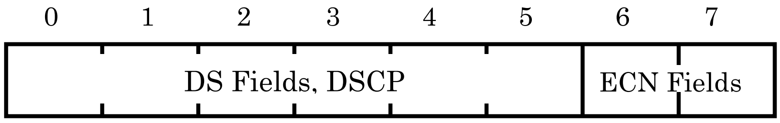
1. Transport layer feedback messages
2. Payload specific feedback messages
3. Application specific feedback messages

Transport layer feedback messages are independent of codecs or applications. They convey general purpose feedback information and are supposed to be interpreted at the transport or at the RTP layer. Payload and application specific feedback messages on the other hand, carry more specific feedback information for a particular payload type or application. Recently an IETF draft [22] has been submitted where a few extended transport layer and payload specific feedback messages have been proposed. One of the most significant transport layer feedback messages defined in that draft is Temporary Maximum Media Bit-rate Request (TMMBR) which works in couple with Temporary Maximum Media Bit-rate Notification (TMMBN). The TMMBR message is typically used by the data receiver, mixer or translator to request the sender of the data to limit the sending rate with a value calculated from the message. The receiver of TMMBR message is free to disregard or follow the request. In either case it is suppose to send a TMMBN message back to the sender of TMMBR message upon receiving any TMMBR. TMMBR messages can be used to both increase and decrease the bit-rate. Hence according to [22] TMMBR message could be useful to avoid serious congestion by doing transport link bit-rate reduction and restoring the bit-rate when condition is improved. However, the use of TMMBR must not conflict with traditional RTCP reports which, when reporting increasing packet losses, should make the sender reduce its transmission bit-rates. Hence, TMMBR is only allowed to be used when the receiver knows, that the available bit-rate to receive media has change. A simple packet loss monitoring scheme is not enough to base TMMBR message generation on.

2.2.4 ECN (Explicit Congestion Notification)

The ECN mechanism was proposed in RFC 2481 [23] by K. Ramkhisnan and S. Floyd in 1999 as an experimental protocol for TCP. Later, in RFC 3168 [4], this has been added as a standard protocol for the Internet community. Despite of its built in congestion control mechanism [19], TCP cannot help improving the service quality of the delay sensitive applications like telnet, transfer of audio and video data etc, as it treats the internet as black box [24] and loss of packets as an indication of congested network. TCP only reduces its transmission window when it observes packet loss in the network and tries to recover that loss by fast re-transmission or fast recovery thus can only improve the throughput in terms of packet loss. In this process it may also introduce unnecessary queuing delay for all other traffics sharing the same resource. But the use of active queue management [25] can let the routers inform end nodes about congestion before it occurs in the network.

One of the mechanisms of active queue management to inform end user about congestion is to use the Congestion Experienced (CE) codepoint in a packet header. This ECN mechanism uses 2 bits in the IP header (figure 2-4) to convey the congestion information with help of four ECN code points 00 to 11. Codepoint 00 indicates that the packet is not using ECN, senders of data are free to use codepoint 01 or codepoint 10 to indicate that the transport protocols are ECN-capable and codepoint 11 is used by the routers to let the end nodes know about the congestion and allows them to take counter measure to handle the congestion.



DSCP: differentiated services codepoint
ECN: Explicit Congestion Notification

Figure 2-3: ECN fields in IP

In RFC 3168 a transport layer support for ECN in TCP has been discussed leaving the other transport protocol (e.g. UDP) support open for further research. In case of TCP two new header fields have been defined, ECN-Echo (ECE) and CRW, which update the reserved bits in the TCP header [26]. Whenever a TCP end point receives an ECN marked bit in the IP header it starts sending ECE bit set in the next TCP ACK and continues until it gets a packet with CRW but set from the sender. The response of TCP has been more detailed in [27] by Floyd.

2.2.5 ECN for video traffic in LTE

This is an open research issue. In this thesis, the work done in [28] has been adopted to make ECN work with video traffic in LTE simulator.

2.2.5.1 ECN for UDP

To make the ECN fully functional, transport layer support is necessary. RFC 3168 has come with support for TCP traffic but has kept the support for UDP traffic open for research. There has been some work done in research community but nothing has been standardized yet. In [28], keeping the UDP protocol intact, minor modifications have been done in RTP packet header to convey the congestion marking to the receiver. In the simulator, no real packet header is used for simulation so only adding a Boolean variable in java class this congestion marking can be achieved. But in this thesis a real video client will be used hence it needs to support standard UDP and RTP specification. Thus this implementation needs more modification to used in this study. So the support of ECN for UDP is achieved in the following way. The real video client is connected with LTE simulator with UDP socket which will send standard RTP packets over UDP to the simulator and expected to receive the same. In this case when the LTE simulator is about to release one video packet at the socket, a check has been made to see that if this packet has been congestion marked in the LTE simulator and if so, one extra byte has been added to the payload of the UDP packet. That byte has later been examined and discarded before decoding the packet at the receiver. The rest of signaling associated with ECN has been done with the help of early RTCP RR report.

2.2.5.2 Implementing ECN at base station for LTE

Though LTE has not yet been fully standardized, enough work has been done to implement the major techniques which will be used. In a LTE system, the base station is likely to act as a bottle neck node as from this point radio transmission starts (see figure 2-1). So in LTE e-nodeB is the place where the congestion marking need to be done. The general approach of congestion marking is to monitor the buffer size in the routers and when it reaches a certain threshold value congestion marking will be triggered. But the buffer management in the used LTE simulator is different. Here one buffer is allocated for per user per service basis. Therefore, the buffer size will not be increased unless there is a heavy congestion in the radio network. This leads to a different approach for congestion detection. Instead of looking at buffer size at the base station the time one packet spends in the buffer has been monitored. Generally, a packet is likely to spend more time if there is congestion in the radio network. When the spend time in buffer for a packet exceeds a certain threshold value then that packet is marked as congestion experienced. Different threshold values for congestion marking has been studied and it has been seen that from a network point of view the earlier the packet is marked, the better cell throughput can be achieved.

2.3 H.264/AVC video codec

H.264/AVC (ITU-T H.264 / MPEG-4(part 10) Advanced Video Coding [29]) is the most recent and most powerful video coding of the series of international video coding standard. It has been jointly developed by experts from ITU-T's Video Coding Expert Group (VCEG) and ISO/IEC's Moving Picture Experts Group (MPEG). Having focused on "entertainment-quality" video it has a balance of the coding efficiency, complexity of implementation. Though it was targeted to be used in mobile devices, it has been standardized for wide variety of applications. It comes with four different profiles and has number of error resilience techniques described in [30].

This video coding falls in to hybrid video coding standard where time and space are treated separately. As the other video codecs in the series it defines five different types of slices named – I, P, B, SI and SP-slices. The traditional I or "Intra" slice is called IDR (Instantaneous Decoder Refresh) for H.264 which describes a full still image. Usually the first frame of a video sequence is made off I-slices but they are also used when there remains an abrupt change in video sequence. As, I-slices do not refer to any previously coded pictures the decoder can start decoding them as soon as they arrive without having decode the previous part of the incoming bitstream. P-slices or "Predictive" slices use one or more previously decoded slices as reference to draw a new picture. B-slices or "Bi-directional predictive" slices used one or more former or future I or P-slices to reconstruct one picture hence they need to be decoded after the following I or P-slices. SP and SI slices are used when there is a need to switch between two or more H.264 video streams. Usually, the following pattern of slices can be found in a video sequence (figure 2-4).

I	P	P	...P...	P	B	P	...P....	I	P	P

Figure 2-4: different slice pattern in video sequence

More detailed description of H.264/AVC standard coding can be found in [31] and [32].

2.4 Video Quality Analysis

Video Quality Analysis assesses the quality of video transmission over different mediums of transport. Often the transport characteristics, compression techniques and data loss can have a clear impact on the video quality which impact the quality difference between the original video and received video. Video quality analysis is very important both in research and real application development. There have been several different proposed methods of analyzing video quality and all those methods can be divided into two major categories – (i) subjective video quality analysis and (ii) objective video quality analysis.

2.4.1 Subjective video quality analysis

In subjective video quality analysis users/viewers of the video sequence are directly involved. They are asked to evaluate the quality of the transformed video (i.e. decoded received video) with one or more reference video sequence. ITU-T has defined many "subjective video quality measurements" in ITU-T recommendation BT .500 [33]. Mean Opinion Score (MOS) is the most common way to represent the result of a subjective video quality analysis. Subject quality estimation is often time consuming and expensive. Usually, there is a need to have a lot of samples to make the result statistically significant. The audiovisual content of the sequence also have impact on the subjective assessment of video. Recently it was found that good audio quality can improve the video quality [34].

2.4.2 Objective video quality analysis

In objective video quality analysis mathematical models are used to estimate the subjective quality of video. A computer program automatically compares the processed video sequence with a high quality original video sequence and objectively measures different matrices to evaluate the quality. Video frame rate, bit-rate and frame correlation are very important factors are when measuring objective video quality. Objective video quality analysis can be very useful as it is repeatable, inexpensive, fast and can be used as a benchmark. The choice of measurement matrices are also very important where wrong choices of matrices can lead to an erroneous conclusion. The most commonly used matrix is PSNR (Peck Signal to Noise Ratio) which is calculated as shown below.

PSNR is defined as:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

where, MAX_I is the maximum pixel value of the image. If pixels are represented with n number of bits then MAX_I is 2^n-1 and MSE (Mean

Squared Error) of two $m*n$ monochrome images I and K can be computed as

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}||I(i,j)-K(i,j)||^2$$

Though PSNR is easy to calculate and measure, it does not correlate the sensitivity of subjective evaluation of human eye. Moreover, PSNR calculation is heavily dependent on correlation between reference video and processed video so if there is any lack of synchronization in video clip PSNR evaluation can be misleading.

2.5 Related Works

2.5.1 Congestion and adaptation

Adaptation as a counter measure for congestion has been an interesting research topic since 80's [35]. This has been further investigated in a large number of reports; see e.g. [24], [36], [37]. All of the research at that time was aimed to TCP/IP. In 1999 Floyd and Fall [38] brought the danger of losing UDP packets into the light and since then efforts have been spent also on providing congestion control for UDP traffic, see e.g. [39] and [40]. Even, Dynamic Congestion Control Protocol (DCCP) [41] was proposed to provide transport for real time traffic replacing the UDP.

Significant research efforts have been spent to investigate how to provide QoS and guaranteed bandwidth for multimedia traffic. References [42] through [44] are examples in this area. Reservation based schemes to provide QoS and guaranteed bit-rate for application has been proposed in [45] and [46] but these schemes can not provide solutions to some of the important potential problems on Internet. Middleware layered architecture has been investigated in [47] but such schemes are by their nature not scalable and there will be a change necessary to existing applications. Active probing based approach has also been an interesting research topic to estimate the available bandwidth in the wireless networks [48], [49].

2.5.2 Rate control mechanism

TFRC (TCP Friendly Rate Control) caught the attention of the research community when Floyd et al introduced this in 2000 [50]. As TCP has its own congestion control management TCP traffic is more generous in nature and makes the way for other traffic in a congested network. Now, if other traffic flows do not address the generosity of TCP traffic then TCP traffic will be starved for network bandwidth. It is therefore desired that congestion control mechanisms for other transport protocols should be TCP friendly. A number of TFRC algorithms and variations on the same scheme can be found in [51] through [53]. The TFRC algorithm uses an equation based rate control ([50],[54]) where the source uses an equation which is a function of RTT (Round Trip Time) and packet loss probability to estimate the available bandwidth and adjust its sending rate. This is in contrast with the AIMD (Additive Increase and Multiplicative Decrease) [55] rate control algorithm used in TCP where the sending rate is halved if there is any packet loss observed and restored with a fixed additive value. However, it has been shown in [56] that the TCP friendly algorithms are not media friendly. Neither are the equation based algorithms suitable for links with high delays and high bit error rates. The AIMD algorithm has the unnecessary delay in restoring the bit-rate which can lead the under utilization of network resources. One of the main advantages of AIMD is that it allows a faster response to the packet loss. A comparative study on AIMD and loss proportional decrease algorithm can be found in [57].

In [58] a variation of AIMD algorithm has been proposed which they called LIMD/H (Linear Increase Multiplicative decrease with History). Here they maintain a history of packet loss and tries to differentiate the cause of packet loss whether by congested network or by other than congestion and react aggressively for losses due to congested network. Certainly, this approach helps to overcome some of the disadvantages of AIMD but in a mobile network it is very difficult to distinguish the cause of packet loss and to do that certain information is needed from lower layers (i.e. link layer).

A Multiple TFRC based scheme has been investigated in [59] where the sender maintain multiple connections and observes RTT values and

packet loss for all the connections and chose best the connection to send video. This have a clear performance gain over TFRC but switching among the connection leads to a fluctuating behavior.
[60] Describes a multi-state congestion control scheme which tries to combine the advantages of TFRC and AIMD in to one scheme. It uses a modified AIMD scheme name "DWAI/LDMD (Distance Weighted Additive Increase and Loss dependent Multiplicative Decrease)" [61] which performance is comparable with classic AIMD.

2.5.3 ECN based rate control

The use of ECN for UDP has been studied in a scattered way to avoid congestion caused by real time application traffic due to their use of higher bandwidth. A TCP friendly multimedia flow control using ECN marking mechanism can be found in [65] and [66] where the authors argued that if a TCP friendly rate estimation is calculated based on ECN-marked packet probability instead of packet loss probability then the throughput degradation can be prevented in wireless networks. In this paper the authors have not shown the percentage of loss ECN marked packet due to link error. Hence, even though they have shown an improvement in throughput and video quality their results can be questioned.
Another study uses the multi level ECN for UDP where they propose to use the 2 bits of IP header for ECN to use as a means of expressing the aggressiveness of congestion for UDP traffic [67], hence allowing the end nodes to adapt their rate according to need specified by 2 bits in ECN. In this implementation the routers need to be smarter to measure congestion in the network. Another problem is that only the last estimation of aggressiveness of congestion will have effect on rate adaptation as any node can alter the value according to their estimation which may not be desired situation. A similar kind of study will be found in [68].

2.5.4 RTCP feedback based congestion control

RTCP based feedback messages have been used in [68] through [71] as a triggering means of adaptation mechanisms. In [69] authors have amended the basic RTCP feedback functionality with modified RTCP transmission reporting frame by frame assessment and recovers the missing packets by retransmission. This strategy works fine with video as video uses higher bandwidth so RTCP feedback has more bandwidth to use as back link transmission channel to send frequent feedback messages. However, this is a clear violation of standard RTCP mechanism as it does not use RTP/AVPF.

After being put as an IETF draft on August, 2004 Extended RTP profile for RTCP-based Feedback (RTP/AVPF) has been a frequent research topic in the community. Researchers have tried to enjoy the relaxed time rule for sending feedback and advantage of payload specific and application specific feedbacks.

[63] and [70] examines the RTP/AVPF for multimedia streaming for 3G networks. It uses the buffer level occupancy to trigger adaptation. The buffer level information has been send by the use of RTCP/AVPF.

[71] Shows that how the RTP/AVPF's relaxed timing rule and payload specific feedback message can be used with H.264 video codecs' error resilience mechanism to gain substantial video quality improvement.

2.5.5 Other congestion control

All the adaptation schemes discussed so far takes counter measure for congestion when the loss has been observed but there has also been proposals where the authors tries to avoid the congestion rather allowing it to occur in network. These adaptation schemes either uses active probing from the router or intermediary nodes, or estimates the buffer occupancy or uses ECN (Explicit Congestion Notification) discussed in section 2.1.6.

In [62] the local buffer information is used trough customized robust feedback mechanism for changing the rate in encoders and decoder. More standard way of feedback mechanism and buffer occupancy calculation was used in [63].The average hop distance is currently 16 in the Internet and will increase in future [64]. As each intermediary node will induce queuing and processing delay which might corrupt the buffer occupancy estimation.

A more application oriented and user plan adaptation has also been investigated where user perceived video quality [72] and content of the video [73] triggers the adaptation. These kind of schemes need user input to initiate the adaptation procedure which might not be desirable especially in an IMS video telephony scenario

Chapter 03

3. Adaptation and Rate Control Techniques

This chapter starts with defining the term adaptation for this thesis and continues describing the state of art adaptation techniques used to handle the congestion in the network with a question/answer approach. At the end of this chapter the basic philosophy followed in this thesis to design adaptation machine is described.

3.1 Congestion and Adaptation

Before defining adaptation for congestion we need to look at what is congestion and how it might occur in the network. Internet is a connected network of several heterogeneous small networks. This heterogeneity includes different access technologies, system capacities, QoS requirements etc. When the data traffic leaves the sender it is going to traverse through a set of interconnected communication links which have different available bandwidths, QoS, and capacities. It is not uncommon to overload a network link with traffic coming from different sources and leaving for different destinations which will eventually lead to either excessive delays or packet drops. This scenario is termed as congestion and the network state is called a congested network. The only way to prevent this congestion is to release network resources such as router processing power or increase the link capacity. The latter option is not a realistic solution as the maximum capacity limit of a particular link is often fixed and installing links with higher capacity will of course terminate the current session. But to dynamically release network resources is possible through adaptive behavior in the transmission of data. An adaptive behavior will allow the sender to manipulate its sending rate depending of availability of network resources. Hence, adaptation can be defined as follows-

> ***"Adaptation is a technique for monitoring network utilization and manipulating transmission or forwarding rates for data frames to keep traffic levels from overwhelming the network medium."*** *[74]*

A cellular environment is more complicated since it seeks to provide services in the context of variable available bandwidth, location dependencies, bursty wireless link errors and user mobilities with different speeds. Here, the user might reach the cell edge which will need a significant amount of retransmissions, i.e. network resources, to deliver the data from the base station to the destination and vice versa. These network links or radio links will often act as a bottleneck for the rest of the network which will eventually lead to delays or packet drops in the bottleneck links. An efficient retransmission or link adaptation mechanism can reduce the packet loss but there will still be some packet loss and delay variations. Thus ensuring the maximum network utilization and providing constant level of QoS is a big challenge.

Moreover, future communication applications and systems will impose higher QoS requirements than that of the current state of the art. Especially, in next generation cellular systems like LTE higher call blocking probability and forced call termination needs to be kept at an absolute minimum. In this circumstance, adaptive applications or services can play a vital role to provide better quality of experience. The QoS scheme in LTE defines two rate specific QoS attributes; Guaranteed Bit-rate (GBR) and Maximum Bit-rate (MBR). GBR denotes the bit-rate upon which the admission decision is based while MBR denotes the maximum bit-rate which the user can utilize. A granted GBR confirms that the network admits one call and has reserved enough resources to provide at least bit-rate specified in GBR. The call can then use bit-rates up to the MBR. GBR and MBR can be used by the adaptive application for a better utilization of resources and improved QoS.

3.2 Implementation issues in Adaptation

To implement adaptation algorithms for congested networks, four questions need to be answered –

I) *Who should do the adaptation?*
II) *Where the adaptation should be done?*
III) *How should the adaptation be performed?*
IV) *When should the adaptation take place*?

The following sections will describe the state of art adaptation techniques to address those questions.

3.2.1 Who should do the adaptation?

- **Sender side adaptation**-- Sender can adapt with the network condition changes by changing transmission rate of the application according to the available feedback from the network or receiver. Buffer occupancy information from network nodes or packet loss and delay observed at receiver end can be fed to the sender through some feedback channel or by periodic probing from the sender. Then the sender has to devise a model to estimate the network condition and adapt accordingly. This approach is suitable for heterogeneous environments as the receiver may have different access technologies compared with the sender and may observe a wide range of congestion scenarios during the session. The main advantage of this approach is that it is more scalable than any other approach.

- **Receiver side adaptation**-- In receiver side adaptation the receiver of the stream has the option to select one stream from number of layered transmitted or layered encoded streams, depending on current network condition. It involves sender side streaming techniques and layered encoding. Multicasting is used for this technique. Multiple copies of data are encoded and transmitted on different multicast channels providing multiple quality levels of the services and the receiver selects the appropriate transmission channel by subscribing to a particular multicast group. The receiver can dynamically change the multicast group depending on its needs. Layered encoding is often done off-line so this is not suitable for real time communication. This approach provides a distributed solution managing local heterogeneity.

- **Intermediate node dependent adaptation**-- In this approach an intermediate node is placed in a convenient place to cover a set of networks with common rules for QoS and device constrains. These nodes are also known as “proxy gateways”. The receiver connects with these proxy gateways to get services form the sender or servers. These gateways can perform two kinds of adaptation. First, the proxy gateway can invoke sender side adaptation based on feedback form the receivers and second, is the gateway can modulate the transmission rate or encoding techniques of the stream suitable for the receivers connected to it. This approach allows deployment of complex adaptation mechanisms without changing the receiver or sender and allows serving a good number of clients with same set of rules. This could also prove to be a bottleneck problem as one gateway need to handle a large number of different streams and it might be hard to make optimal decisions for each different stream depending on the same set of parameters. It also has implementation complexity.

3.2.2 Where the adaptation should be done?

- **Adaptive protocols** -- Adaptive protocols have their built in congestion control management. The transport layer protocols are involved in the process. All the adaptive decisions and feedback should be interpreted at the protocol level. Protocols like TCP, DCCP, RCP falls into this criterion.

- **Adaptive applications** -- In this case, adaptation is done on application layer with a feedback channel transporting the congestion related message back to the sender. Their might be a need for help from the transport layer for transportation of feedback messages but all the messages should be interpreted by the applications and the applications need to devise a adaptation model to adapt with the congested situation.

- **Cross layer adaptation** -- This approach integrates two or more layers of communication to provide notification mechanisms about congestion. Typically, information from the lower layers is used to estimate the network condition and the adaptation decision is taken by the upper layers i.e. the application layer.

3.2.3 How should the adaptation be performed?

The general procedure to adapt with a congested network situation follows three steps- i) congestion detection, ii) congestion notification, iii) congestion reaction.

3.2.3.1 Congestion detection

This is the prerequisite for adaptation. The more efficient the congestion detection process is, the more robust the congestion handling will be. There are several proposed ways to detect the congestion. Packet loss, jitter value, buffer occupancy information, queuing delay or queue time in routers or retransmission delay in base station are the most commonly mentioned ways of detecting the congestion in the network. All these methods have their own advantages and disadvantages. With the help of packet loss and jitter values in packet arrival it is more challenging to detect the congestion in prior time while others are more direct methods better suited to detect the congestion prior to its occurrence.

3.2.3.2 Congestion notification

Usually, the adaptation decision is taken in another place other than the place of congestion detection. The sender of the stream needs to modify its sending rate, encoding rate or retransmit the lost packets but the congestion detection is usually done in intermediate nodes (i.e. routers, base stations etc) or in the receiver end so the notification is important. The notification procedure can be application dependent or transport layer dependent. For example TCP has its own notification procedure where as UDP does not have any so any application using UDP as a transport protocol must use RTCP or other application defined feedback messages to notify the sender about the congestion.

3.2.3.3 Congestion reaction

There can be several procedures to react on congestion notification such as limiting the sending rate, implicit bandwidth renegotiation, connection drop or blocking. Among these possibilities, rate adaptation is quite attractive since it has a high potential of maintaining acceptable quality while not being too complex. Further, it is an easy way for multimedia communication as in this case only the sender and/or the receiver are involved. Advanced video codecs such as H.264/AVC and audio codecs like AMR-NB/AMR-WB, this can easily be implemented. The normal behavior is to go down in bit-rate when there is a notification of congestion, and when the network condition has improved, the sending rate is increased to restore the initial sending rate. The following rate control procedures are often utilized in many cases of adaptation-

- **Additive Increase and Multiplicative Decrease (AIMD)**

In this scheme the sender of the traffic is notified about the congestion in the network by means of feedback messages. The sender will decrease the sending rate with a constant multiplication value and if the sender is notified about the congestion reduction in the network it will increase the sending rate with a constant additive value. So if we define C_i as a congestion notification and R_i as the sending rate at time t_i then AIMD can be defined as

$$R_{i+1} = R_i + \alpha, \text{ if } C_i = 0$$

$$\text{And} \quad R_{i+1} = R_i * (1 - \beta), \text{ if } C_i \,! = 0 \quad \text{------------} \; (1)$$

Where, α is the increase constant and β is the decrease constant. Typically AIMD converges to a optimize rate allocation with "-1/x" utility function. A detailed convergence analysis of AIMD can be found in [75]. In a classic AIMD algorithm the typical values of α = 1.0 and β = 0.5. From the definition of AIMD it is clear that the decrease and increase of sending rate is independent of the number of packet losses in one round of the adaptation cycle.

- **Additive Increase and loss Proportional Decrease (AIPD)**

AIPD schemes adjust their sending rate proportional to a fraction of the packet loss. If we assume f_i as the loss fraction and R_i as the sending rate at round i then according to AIPD

$$R_{i+1} = R_i + \alpha, \text{ if } f_i = 0$$

$$\text{And} \quad R_{i+1} = R_i * (1 - \beta.f_i), \text{ if } f_i > 0 \quad \text{------------(2)}$$

Where, α is the increase constant and β >=1 is the decrease constant. Here β is an adjustable parameter which determines the variation of rate change. It is obvious that from equation 2 that a smaller value of $\beta.f_i$ means a more aggressive decrease and a bigger value means slow decrease. The value is dependent on fractional packet loss.

- **Adaptive Increase and Adaptive Decrease (AIAD)**

As the name suggest, in the AIAD rate control scheme both the increase and decrease of rate is non linear. The increase of rate may be dependent of one or more variables such as distance from initial state, increase in PLR (Packet Loss Rate), available bandwidth etc. The α values in equations 1 and 2 are tunable depending of certain parameters. This is also true for going down in sending rate. The main advantage of this scheme is that it takes the current network condition into account and tries to take the best possible decision to avoid or overcome the congestion situation. On the other hand, is it relatively difficult to design and implement as it is always difficult to estimate current network condition, especially in wireless networks. These kinds of schemes are presented as variations of AIMD or AIPD.

- **Equation based rate adaptation**

Equation based rate control schemes follow a particular equation to set the current sending rate of data. A classic equation for these schemes is given below –

$$X = S/R * sqrt(2bp/3) + RTO[3 * sqrt(3bp/8) * p(1+32*p^2)] \qquad \text{---------}(3)$$

Where X is the transmission rate, s is the packet size, R is the round-trip time, p is the loss event rate as a fraction of the number of packets transmitted, *RTO* is the TCP retransmission time out value, and b is the number of packets acknowledged by a single ACK.

In general, equation based rate adaptation is known as TFRC (TCP friendly Rate Control). It has been shown that equation based rate control does not respond well to packet reordering. It is not suitable for environments with high error rates and is not media friendly.

3.2.4 When should the adaptation take place?

Adaptation can be done in two ways. First, adapting with the network condition so that the congestion does not occur in the network and second, reacting after observing the congestion occurrence in the network. In the first case, the adaptation mechanism needs explicit indication from the network that the congestion is going to occur and if no counter measure is taken there will be a decrease in the level of QoS. As the adaptation mechanism gets the information about the congestion prior to any congestion related transport problem, this approach is helpful for such applications those are packet loss and delay sensitive such as real time media (e.g. video telephony).

In the later case, the receiver informs the sender about the congestion after observing some packet loss or delay jitter values in the received data traffic. Here, the sender receives the feedback message some time later than the event of congestion had occurred and then it decides the adaptation approach and react accordingly. Hence, there always remain delays associated with the management of the congestion which will eventually decrease the QoS level of the service.

3.3 Proposed Adaptation Approach

The basic philosophy behind the proposed adaptation approach is to **provide a fast response to the congestion and slow rate restoration.** Looking at the scalability and ease of implementation, a sender side adaptation will be used in this thesis. The adaptation will be done on application layer with help of RTCP feedback of RTP running over UDP/IP.

With a slogan of "no packet loss due to congestion" the ECN for UDP traffic will be used and the congestion will be addressed before it occurs in the network However, since we will have a mix of wireless and wired networks there is a possibility of loss of ECN marked packets on the radio interface so there will also be a backoff state to handle the packet loss due to radio link error. As we want to use ECN for UDP and there is no standard defined for this usage we have to define signaling support to use ECN for UDP. We will use the RTP/AVPF's relaxed timing rule for feedback message, which allows us immediately report on event back to the sender, for notifying the sender upon receiving ECN marked UDP packets at the receiver. Rather than to define a new APP FB (APPlication Based Feedback) message to serve this purpose TMMBR messages will be used in early feedback mode to notify the sender about the reception ECN marked packet for UDP traffic. One can argue that TMMBR messages are not meant to be used as congestion control mechanism but in this thesis, it is assumed that there is an opportunity to do so.-

First, according to [22] TMMBR can be sent from the receiver side to the sender of the data requesting bit-rate limitation upon observing sudden network changes reported by any network entity. The receiver can only sense the network changes if the network sends specific information about the network load or calculated from lower layer matrices. In our case the e_nodeB (evolved NodeB) will set the ECN bit for the UDP traffic depending on the decision rule used in [28]. This can be seen as an indication from the network entity about the network condition in terms of congestion which is a sudden network change. Hence, use of TMMBR upon receiving an ECN marked packet does not violate the rules in [22].

Second, when the sender receives a TMMBR message from the receiver of the data traffic, the sender of data traffic need to notify or send an acknowledgement to the receiver about the reception of the TMMBR message. TMMBR is used along with TMMBN which is scheduled to send as a notification for TMMBR message. We can use the basic functionality of TMMBR and TMMBN for our case.

Third, we wanted to design an adaptation system with a set of standard protocols and signaling specifications to that the system itself becomes a part of standard.

Besides, The regular feedback RR or SR will be used to report any packet loss or delay jitter value back to the sender.

How to react when the sender gets a TMMBR or observes packet loss is a crucial component in this design. As it has been shown that slow responsive adaptation like TFRC is not media friendly [60], and TFRC has lower throughput and resource efficiency in wireless environment, TFRC is not a good candidate for our network environment. Again the classic AIMD is not suitable for media traffic for its throughput and the constant behavior to address the congestion and restoration of data rate. In this thesis, in addition of ECN market packet a packet loss rate will also trigger the adaptation and an AIAD procedure (described in section 3.2.3.3) will be followed to address the congestion.

Chapter 04

4. Media layer Adaptation for Real Time Conversational Video

In this chapter the proposed adaptation mechanism is described in details. The necessity of adaptation for real time conversational video has been described followed by a usecase scenario describing a real life phenomenon where adaptation can come in to play. The rest of the chapter details the adaptation mechanism with flowcharts.

4.1 Adaptation for video telephony

This thesis gives emphasis on video telephony since

- Video transmission typically requires higher bandwidth and available bit-rate relative to that of audio transmission.
- The effect of packet loss, due to congestion and delay, is instantly noticeable in video.
- It is important to provide a telephony-grade service as part of IMS Multimedia Telephony.

Packet loss and jitter are two important issues related to the transport processing of real time media data. For video, jitter does not produce a dramatic change in quality as it does for voice. Jitter in video transmission will cause jerkiness which has minor effects on the conversational video quality. However, it is totally different in the case of packet loss. Modern video coding relies heavily on prediction mechanisms which increase the coding efficiency at the cost of higher sensitivity to packet losses. Again, video transmission involves higher data rate than speech transmission, so loss of packets has greater impact on the quality of the video as the packets contain big chunk of data stored in them. In a cellular environment where we have error prone wireless links, the larger packets like to have more probability of getting lost in the air. This is also true for a congested network. A bursty or distributed packet loss can be observed throughout the session. As discussed in chapter 2, H.264 encoded video pictures are divided into several number of slices which needs to be re-synchronized to avoid data loss. Let us take figure 2-5 of chapter 2 as an example video sequence and let us assume that a bursty and distributed packet loss was observed during transmission of the video sequence. The packet loss can cause loss of several P-slices and some B and I-slices. Figure 2-4 will then look like figure 4-1.

I	P	P	...~~P~~...	P	~~P~~	P	...P..	~~P~~	P	P

Figure 4-1: packet loss in video transmission

Obviously, this will affect the decoding efficiency of decoder. The loss of different slices will be compensated by either spreading the effect of loss all over the picture or refreshing the decoder with another IDR picture and cut the predictive chain. This is going to affect the perceived video quality. Further, if a number of users in a system try to setup a connection with high bandwidth requirements and start to operate in an high bit-rate, the likelihood of congestion will be high if all the users continue to send data at the initial rate. A successful adaptation mechanism is needed to avoid congestion for the system as well remedy congestion effects for the individual user.

Now, the question is what should trigger the adaptation? Several options are available – packet loss, packet arrival jitter, network information from lower layer matrices. It has been already mentioned that the packet arrival jitter is not so crucial for the user experience in the case of video. Network information can be retrieved by periodic probing or from lower layer matrices but both of these methods have their own extra implementation cost and retrieving lower layer information will break the IP protocol layering convention. The packet loss information on the other hand, is readily available through RTCP SR or RR. In the context of information retrieval, an increase in packet loss is a suitable adaptation trigger since video is more sensitive to packet loss and the encoding bit-rate of sender should be adapted relative to packet loss rate observed in the receiver. The major drawback with basing adaptation on packet loss is that the measure can only be calculated after packets have been lost irrespective to the reason of loss. The initial loss of packets will therefore have effect on the perceived video quality and an adaptation algorithm can only try to ensure no further packet loss. The reasons of packet loss can be a congested network, a failure in the radio link, a handover to incompatible radio access technology etc. As the packet loss is not desirable in video transmission, the network can help to reduce the packet loss due to congestion by informing the sender or receiver about the congestion in prior to its occurrence. The sender can then take counter measures according to the adaptation algorithm to change its bit-rate to avoid congestion in the network, ultimately resulting in no packet loss. Unfortunately, this explicit notification of congestion is not available for UDP traffic. Since real-time video is transported using UDP, no standardized congestion notification mechiasm is available.

Although the utilization of higher bit-rates and the high sensitivity to packet loss makes the video adaptation more challenging than that of speech, video codecs used in IMS Multimedia Telephony does not have an adaptive heritage similar to the speech codecs AMR and AMR-WB. But, the use of H.264/AVC baseline profile in IMS video telephony allows better video quality than previous standards even at low bit-rates. Therefore, going down in bit-rate according to the adaptation algorithm in a congested network situation might allow us to provide services with relatively better video quality than utilizing older encoding schemes with a higher packet loss rate. Hence, the prospect of using media layer adaptation to improve service quality is good in IMS Multimedia Telephony.

In this study we will use state of the art technologies to do rate adaptation for real time conversational video with an aim to investigate the benefit of using a "pre-warning" compared to allowing packet loss.

4.2 UseCase scenario

In this thesis we will consider the following simple IMS Multimedia Telephony scenario in which a unicast video transmission will be established between two communicating peers.

> *Adam has been invited by one of his old school friends, Lena, for lunch at Stockholm City Restaurant. Before going to the restaurant they want to meet with each other at a place then go to the restaurant together. Adam has reached the place of their meeting on time but he doesn't see his friend there. Now, Adam is confused about the place "is this the correct place?" He calls Lena who is traveling through the subway. Lena gets Adams call when she was about to change the subway train. There were a lot of noises and Lena was not hearing Adams voice, but she figured out that Adam has already reached their*

> *agreed meeting place but he is confused about it. On the other hand Adam understood that Lena is not hearing him correctly but he remembered that last time Lena told him about her new mobile phone. Fortunately, Adam also has the same and he decided to show his current place by adding video to the session. He started the video and Lena confirms the place by just saying "yes" after viewing the video while she is in a tunnel of the subway. Thanks to the new phone with these features, Adam cuts the call with a big smile on his face.*

Although this scenario looks very simple, it has some characteristics that are relevant to this investigation.

- It shows that video can be important and a useful mean of communication

- As one user in the scenario wants another user to recognize some place by looking at the video, the video quality is very important. The video should provide enough quality so that objects in the video can be recognizable just by having a quick glance at it. In this case packet losses due to bad radio coverage or congested network might not give the desired video quality. Packet loss can make the video quality degraded resulting in not enough information being available to recognize the objects in the video.

- The multimedia telephony session should be able to continue to provide a reliable service even in a hostile situation. Usually the subway tunnels have bad radio coverage so when the actor is in the tunnel and the call might be dropped, either due to bad radio coverage or a congested network. The user will be unsatisfied because the new features in the phone are not good enough to provide the expected service.

4.3 Adaptation design guidelines

The proposed adaptation scheme follows design guidelines found through study and observations of current state of art solutions –

- Fast reaction to loss and slow bit-rate restoration
- Use of ECN (Explicit Congestion Notification) for UDP traffics to achieve no packet loss for video transmission
- Transport support for ECN
- Support for adaptation if ECN fails or lost in radio interface
- Lowest allowed bit-rate is 30kbps
- 2% of packet loss is tolerable in the whole video session
- Measuring results in terms of video quality

4.4 Media layer Adaptation Scheme

4.4.1 Adaptation States

The proposed adaptation scheme has five different adaptation states and 2 sub-states. Figure 4-2 illustrates the five different states and their sub-states.

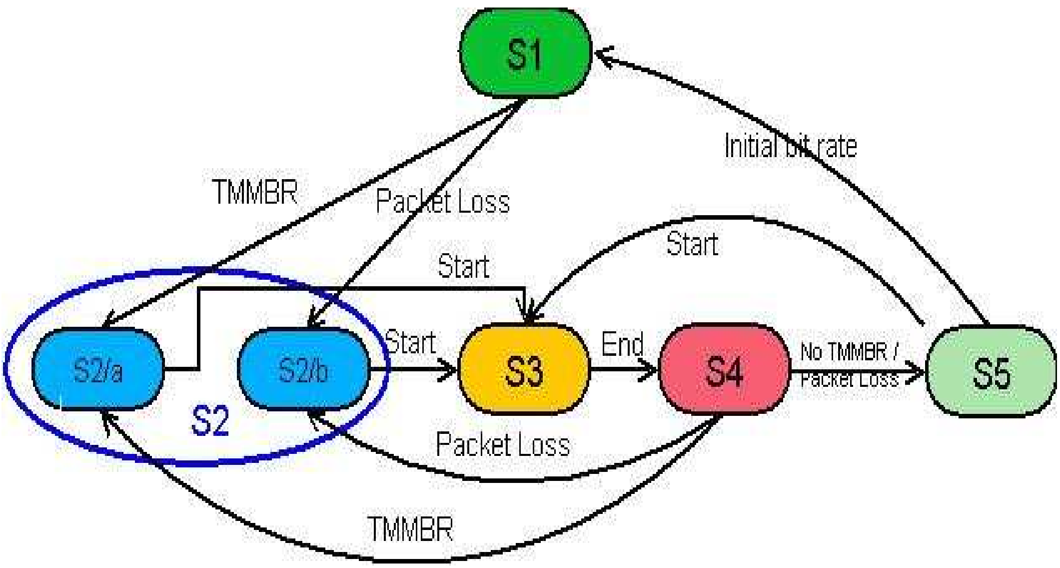


Figure 4-2: Adaptation state machine

As shown in the figure both TMMBR (ECN marked packet) in early RR or packet loss reported in regular RR can trigger the adaptation. The adaptation scheme stays only in one state at a specific time. The purposes of different are stated below-

- **Initial state (S1)**—this is the default state of this scheme. Highest coding and bit-rate is used for data transfer.
- **Backoff state (S2)**—in this state the bit-rate will be reduced by the sender of the data stream. Depending on the cause of the visit, this state is further divided into two sub states.
    - **Fast Backoff state (S2/a)**—this state will be visited upon receiving a TMMBR message form the receiver. In this state an aggressive bit-rate reduction will be performed.
    - **Slow Backoff state (S2/b)**—this state will be visited depending on Packet Loss Rate (PLR) calculated form packet loss reported via RR. A slightly smoothed bit-rate reduction strategy will be followed in this state.
- **Wait state (S3)**—this state will be visited upon any change in bit-rate in both backoff states and probing state. When the encoding bit-rate is changed by the sender, some time is needed to realize the effect on the network characteristics by the receiver. The main purpose of this state is to provide enough time to detect the bit-rate reduction and increment at the receiver end. While the scheme is in this state all the TMMBR message and packet loss reported in the RR will be ignored.
- **Detect state (S4)**—this state is an auxiliary state. In this state the control algorithm will observe network conditions for taking future decisions.
- **Probing state (S5)**—in this state the bit-rate will be increased with the time interval allowed by the detect state. Here the bit-rate will be increased in adaptive manner.

Table 4-1 depicts the state transition conditions and actions taken in those states.

Transition		Entry condition	Actions	Exit condition	Next state
from	to				
S1	S2/a	TMMBR from receiver	Reduce bit-rate	Bit-rate is reduced	S3
S1	S2/b	Packet loss reported in regular RR and PLR >=2%	Reduce bit-rate according to PLR table	Bit-rate is reduced to the rate according to the PLR table	S3
S2/a or S2/b	S3	Bit-rate decreased	Wait until the waiting period has elapsed, ignore any TMMBR or Packet loss	Waiting time has elapsed	S4
S3	S4	Waiting time has elapsed	Check current RR for TMMBR or fractional loss	RR has been checked	S2 or S5
S4	S2/a	TMMBR from receiver	Reduce bit-rate	Bit-rate is reduced	S3
S4	S2/b	Packet loss reported in regular RR and PLR >=2%	Reduce bit-rate according to PLR table	Bit-rate is reduced to the rate according to the PLR table	S3
S4	S5	No TMMBR or packet loss reported in RR	Increase bit-rate	Bit-rate increased	S3
S5	S1	Current bit-rate = initial bit-rate	Maintain steady bit-rate and wait for TMMBR or packet loss reported in RR to start a new adaptation cycle	Received TMMBR or packet loss reported in RR	S2/a or S2/b

Table 4-1: State transition and actions in state machine

4.4.2 Adaptive Increase and Adaptive Decrease

An Adaptive Increase and Adaptive Decrease (AIAD) mechanism has been proposed to change the sending bit-rate of the sender. In this adaptation scheme the only counter measure followed to address the congestion in the network is reduction in bit-rate. Naturally, congestion is more likely to occur when sending bit-rate is high. Looking at this fact an adaptive technique has been adopted to increase the bit-rate. According to this mechanism, when the condition is in favor of increasing the sending bit-rate the step to which the rate will be increased will depend on the difference between current bit-rate and initial bit-rate. So, this technique will be called as "Distance Dependent Adaptive Increase (DDAI)". The advantage of DDAI is that it will take bigger steps when the sender is running at a lower bit-rate and will take smaller steps to increase the bit-rate when the sender is running at a higher bit-rate, which means it will take more careful steps while restoring the initial bit-rate as a result the frequency of congestion occurrence will be reduced.

According to the adaptation state machine there are two events when the bit-rate needs to be reduced. First, upon receiving a TMMBR message from the receiver of the data stream and secondly, observing packet loss in the Receiver Report (RR). These two events do not depict the same network scenario. The TMMBR message in our scenario works as an early congestion notification issued by the receiver of the data stream upon receiving ECN marked packet. This means if sender does not reduce the bit-rate in response to the TMMBR message then congestion will occur in network and it is expected that the first TMMBR will be issued by the receiver before any packet loss happens in the transmission. It has been decided to reduce the bit-rate aggressively upon receiving a TMMBR message from the receiver so that the congestion will not occur and no packet loss will be observed in the ongoing session. Here the reduction step will depend on time and number of TMMBR issued. Hence, it will be adaptive to the network condition. On the other hand, a fractional loss in the RR states the fraction of packets lost since last RR was sent. While operating with ECN this scenario is not expected to be happen but as we can have radio interface in one or both sides of the communicating session ECN marked packets or TMMBR can get lost at any time. In our case, packet loss is expected to occur suddenly due to bad radio conditions and will be handled either aggressively or less aggressively depending on the packet loss rate. This is assuming that ECN will report the risk for congestion in due time for a proper response. The response to the packet loss reported in RR will also be adaptive to the amount of Packet Loss Rate (PLR). Of course, a higher PLR will result in a higher bit-rate reduction and a lower PLR will result in a lower reduction in bit-rate. The PLR calculation is a crucial factor in this case and will be covered in the next section.

4.4.3 Packet Loss Rate (PLR) calculation

Packet Loss Rate (PLR) is an important matrix in this adaptation scheme. The PLR value at a certain time indicates the percentage of packet loss at that time. In the proposed scheme the PLR will be calculated at the sender end and as RTCP RR will be used to convey feedback from receiver we are left with two parameters in the RR report to calculate the PLR. They are i) fractional loss and ii) cumulative loss. The fractional loss reports the fraction of packets lost since the last RR was sent while cumulative loss reports the total number of packets lost since the session started. Certainly, there are some issues related to calculation of PLR with either fractional loss or cumulative loss. PLR calculation with only fractional loss can make the adaptation respond too fast to packet loss. For example, let us consider a session at 100 kbps and 20 packets per second. According to RFC 3551 and RFC 4585, it is recommended to reserve 5% of the total bandwidth to be used for RTCP reports which gives 4 packets per RR on average. If using the fractional loss as basis for the PLR calculation, such a flow would result in 25% of packet loss if one packet is lost. This calculation will make the adaptation scheme “trigger happy”. On the other hand, if we calculate only with cumulative loss this will make the adaptation slow to respond to packet loss and the current network condition will not be reflected in the PLR calculation. Let us consider the same session as stated above. Now, if we don’t have any packet loss in first minute but have 10 packet losses during last two seconds this will give 0.8% of packet loss rate. This is a quite low number which probably not triggers the adaptation. But the current situation should force the adaptation to trigger since loosing 25% of all packets during the last few seconds can be taken as a quite clear indicator of congestion. Taking these extreme cases into account, a new procedure to calculate PLR is defined which involves both fractional loss and cumulative loss, hence addressing the description of the current network condition in a more correct way.

The proposed PLR calculation involves a "history window" which is a sliding window that keeps the history of packet loss reported in a number of previous RRs; a number equal to the given window size. The cumulative loss in the window is calculated and the current fractional loss is divided by the calculated cumulative loss. The result is called "Smoothed Fractional Loss (SFL)". This smoothed fractional loss is then used to calculate current PLR. Hence, PLR is in this report defined as shown below—

$$\text{Smoothed Fractional Loss (SFL)} = {}^{\text{Current Fractional Loss}} / {}_{\text{cumulative loss in the window}} \text{---}(4)$$

$$\text{PLR} = ({}^{\text{SFL}}/{}_{\text{total number of packet send in this window}})\ \% \qquad \text{-------------}(5)$$

The fractional loss reported in RR is not an integer value so at the sender it is necessary to calculate the number of packet loss from the fractional value.

4.4.4 Support for ECN

As it has already been mentioned, ECN support for UDP traffic is not available. Therefore, we propose transport layer signaling support for ECN to be used with UDP traffic. When the receiver of the UDP traffic receives a congestion-experienced marked packet from the network it has to notify the sender of the traffic about this phenomenon. The sender has to inform the receiver about the reception of the notification meaning that it will take the necessary adaptive measure in response to the congested situation. We propose to accomplish this task with the help of the TMMBR and TMMBN message proposed in [22]. The reason for choosing TMMBR and TMMBN for this purpose has been discussed in section 3.3. Upon receiving the congestion-experienced marked packet the receiver will schedule a TMMBR message as soon as possible. Most often this message will be send as early feedback to the sender if a regular RR is scheduled later. When the sender will receive the TMMBR message from receiver it will send a TMMBN message back to the receiver. And when the receiver receives the TMMBN message it knows that the sender has received its request and will terminate the retransmission of TMMBR message. As we can see, here TMMBR is used as congestion notification to the sender and TMMBN is used as an acknowledgement of the rate request.

4.5 Communication signaling

In this thesis a unicast scenario will be studied where one of the peers participating in the conversation sends video to the other peer. The first peer will be called sender and the other peer will be called receiver. The sender will send video data using RTP/UDP/IP. The receiver will send feedback about the quality of the data received with RTCP/UDP/IP. Nodes in the network will be responsible to mark packets as congestion experienced when such situations occur and will do so in due time. If packet losses are observed by the receiver, they will be reported back to the sender with the regular RTCP RR. The sender will calculate PLR each time it receives a RR from the receiver and regulate the bit-rate accordingly. Figure 4-3 and figure 4-4 illustrates the signaling involved in this procedure.

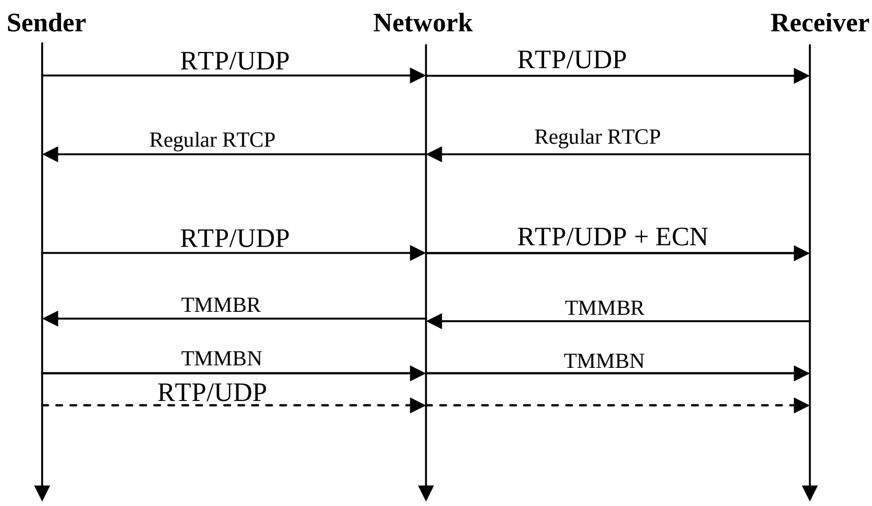


Figure 4-3: signaling ECN between receiver and sender

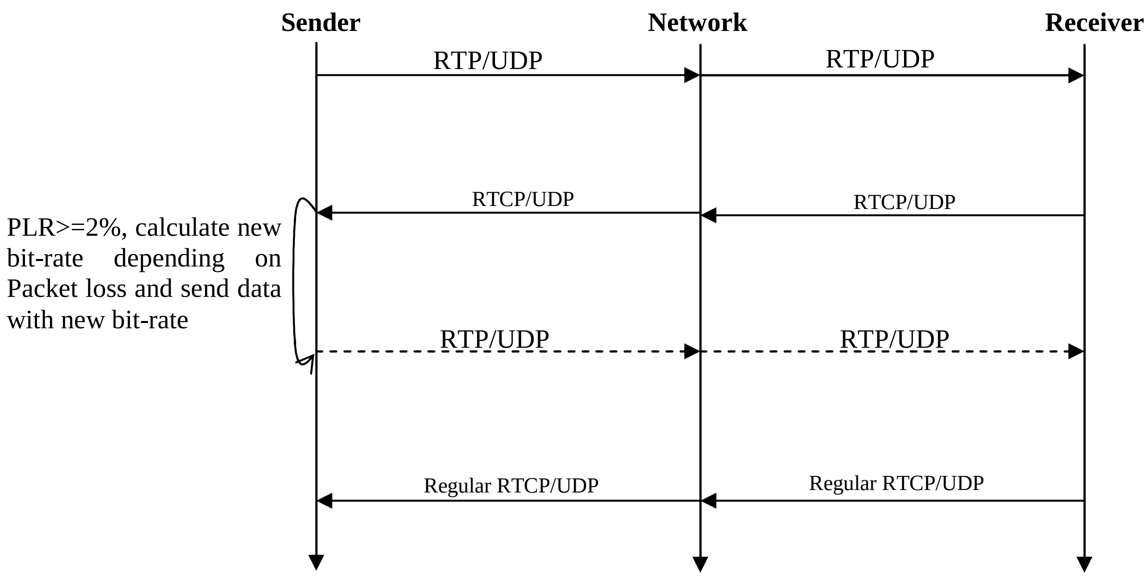


Figure 4-4: signaling packet loss between receiver and sender

4.6 Operational Overview

4.6.1 Receiver side interaction

In this proposed adaptation scheme the receiver performs one regular activity and one special activity. Calculating matrices for RTCP RR and sending RR with regular interval falls in to the regular activity. But receiving one ECN and sending this to sender is a special activity. When ever a receiver gets an ECN from the network it schedules a TMMBR message in either early RR or regular RR depending on the convention specified in RFC 4585. In both cases it expects a TMMBN from the sender in acknowledgement to the last TMMBR message. If it gets a TMMBN in reply from the sender then it checks for another ECN and if there is any it again sends TMMBR in an early or regular RR. If the receiver does not get a TMMBN in response to a TMMBR then it should take the procedure stated in RFC 4585 and try to retransmit the TMMBR. Now, the receiver might get number of consecutive ECN marked packets from network. In this case waiting for TMMBN before sending the next TMMBR might restrict the process of conveying real information about the congestion in the network. Thus, the receiver will not stop sending TMMBR if it gets consecutive number of ECN marked packet rather it will try to schedule TMMBR as soon as it can. Figure 4-5 depicts different decisions and activities those take place in receiver.

Besides, the receiver is also allowed to send a SIP re-invite or SIP update to renegotiate the session parameters at any time either automatically or by the active participation of user.

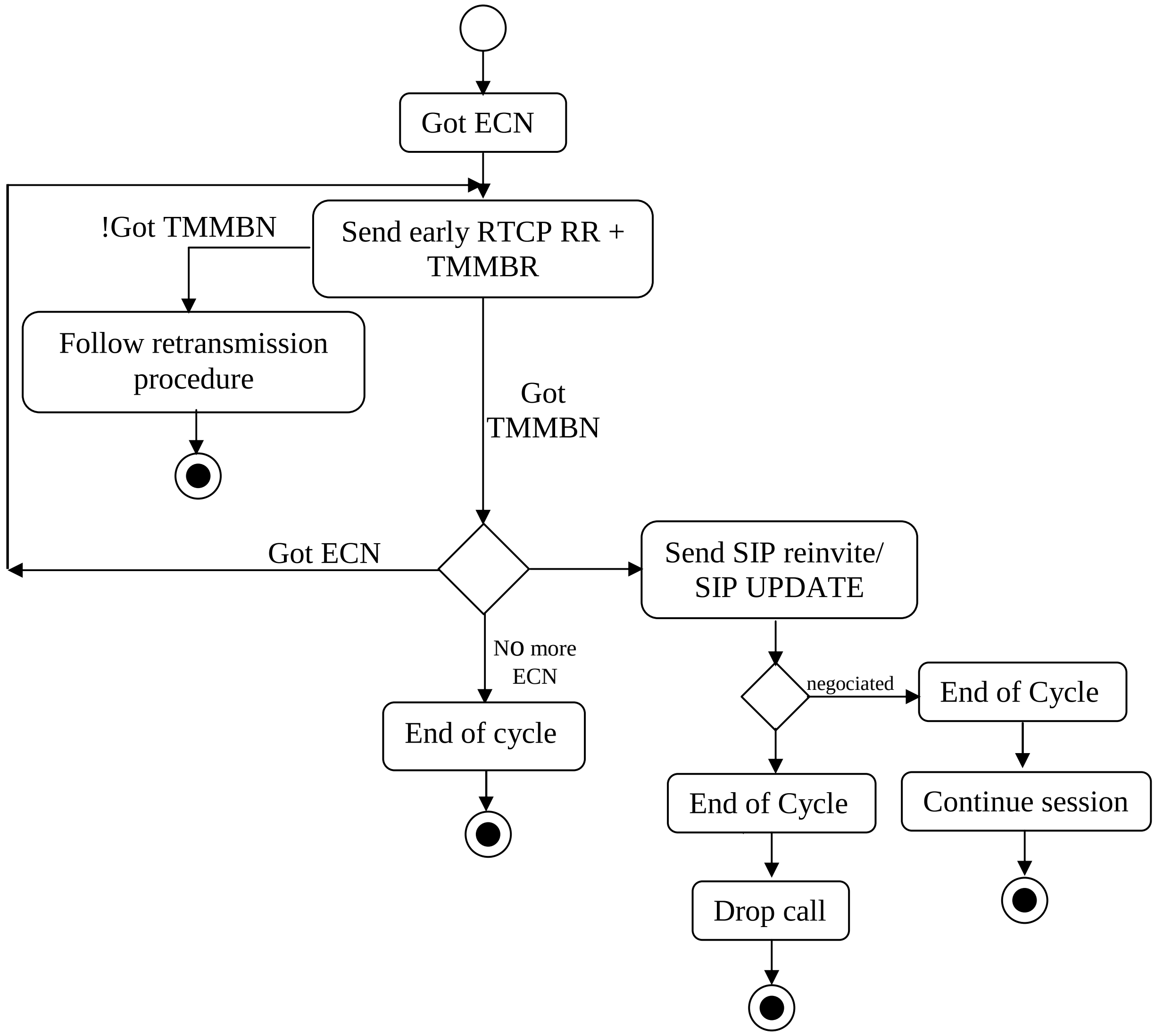


Figure 4-5: Receiver side activities in response to ECN

4.6.2 Sender side interaction

Being a sender side adaptation scheme, the sender has to perform many activities while remaining in different adaptation states. In the initial state the sender maintains highest possible bit-rate. When it gets a TMMBR message from the receiver it switches its current state to the fast backoff state. According to this scheme, in the fast backoff state the bit-rate needs to be reduced aggressively so that we stay on the safe side and avoid congestion in the network. Each time the bit-rate is reduced to 50% of the current bit-rate. If the reduced bit-rate is less than the minimum allowed bit-rate of the session, the new bit-rate will be set to the minimum allowed bit-rate. The sender then goes to the waiting state and when finished, it detects the current network condition by looking at the TMMBR value or fractional packet loss in recent RR. If it senses a good network condition it starts probing. Figure 4-6 depicts the activities taken by the sender in response to a TMMBR message.

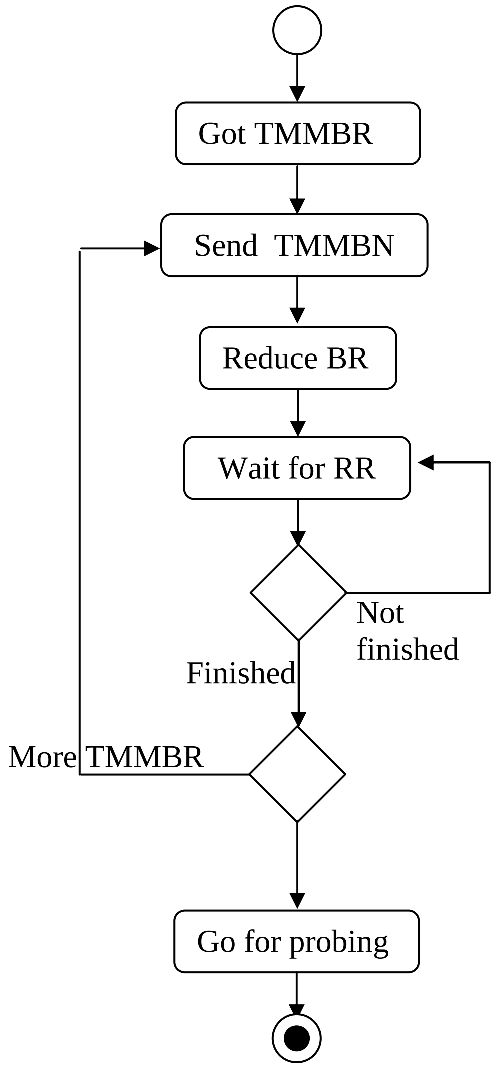


Figure 4-6: Sender side activities in response to TMMBR (fast backoff)

The sender visits the slow backoff state when a packet loss is reported in the RR. In the slow backoff state the sender needs to calculate the PLR value as described in section 4.4.3. The adaptation decision is taken based on the PLR value. There is a PLR threshold value and any PLR values greater than the threshold value will trigger the adaptation procedure. The sender also needs to maintain a table called “PLR table” consisting of different PLR value and bit-rate couples. The idea is to find the PLR value in the table and adjust the bit-rate according to the corresponding bit-rate found in the table. The actions taken by the sender will vary depending on the previous state of visit.

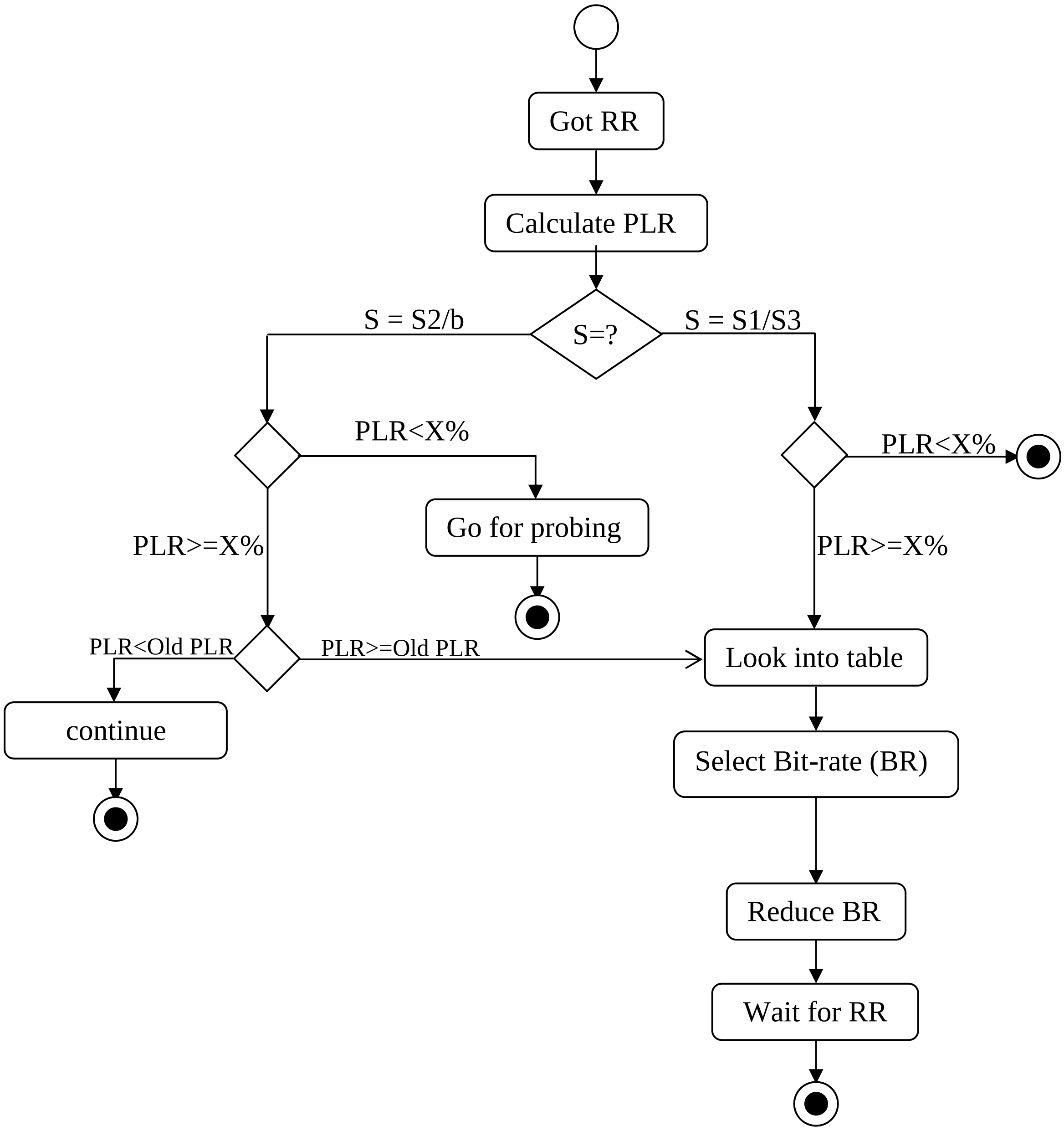


Figure 4-7: Sender side activities in response to packet loss reported in RR

Figure 4-7 depicts the different activities at different states followed by the sender when packet loss is reported by the receiver.

The probing procedure to restore the bit-rate to the initial bit-rate must be done carefully as the higher the bit-rate, the more the probability increase of congestion occurrence. Hence, the step to increase the bit-rate needs to be adaptive. The adaptive increase scheme is described in section 4.4.2. If any TMMBR message or RR indicating a packet loss is reported while probing, the related backoff state will be visited accordingly. The detailed operation of probing state can be found in figure 4-8.

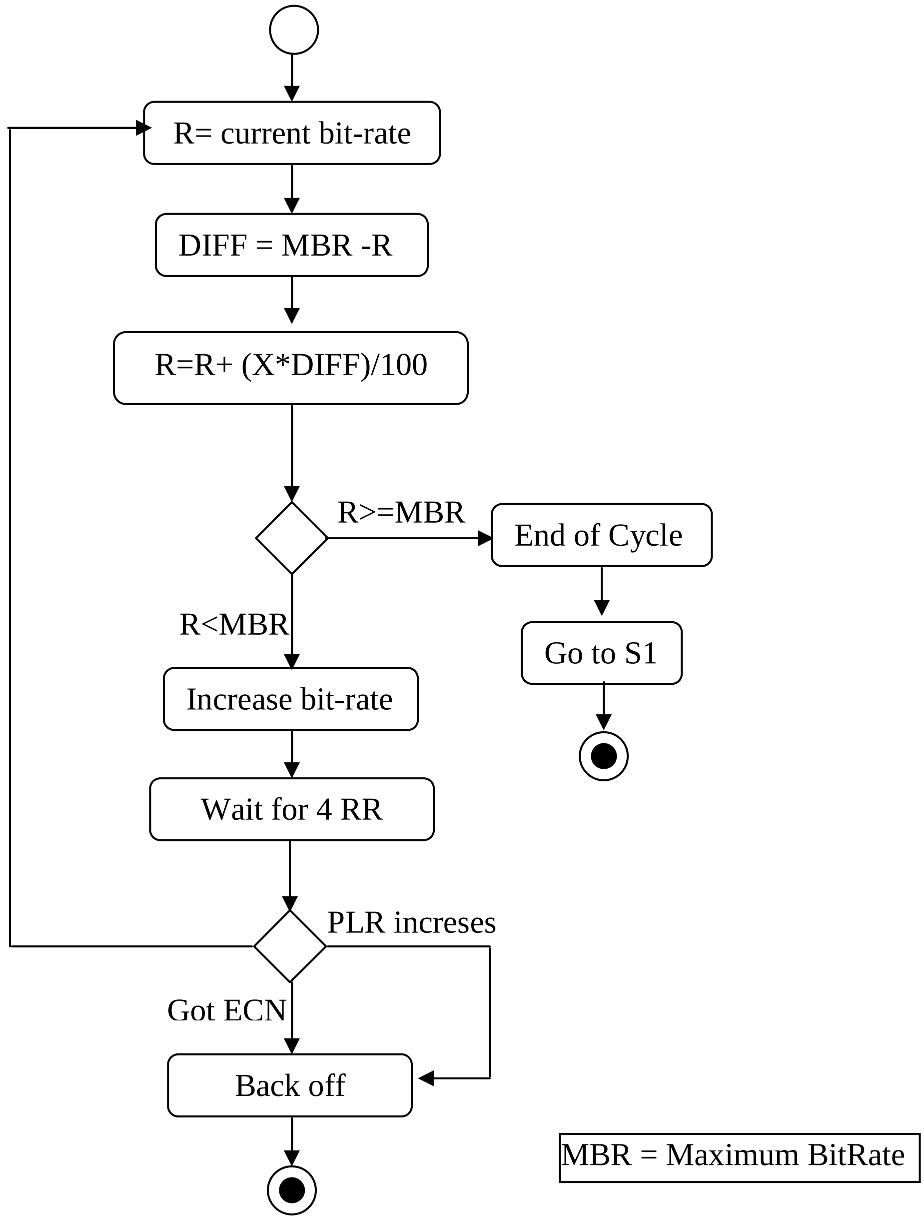


Figure 4-8: Activities in probing state

It needs to be mentioned that as this adaptation scheme has two triggers for backoff (i.e. ECN and packet loss) the sender can encounter both triggers in any sequence which may lead to a malfunctioning situation. Therefore, no direct transition from the fast backoff state to the slow backoff state or vice versa is allowed. This transition can only be possible after visiting the probing state. This constraint also prevents an oscillating behavior in terms of sending bit-rate.

Chapter 05

5. Simulations and Results

In this chapter, simulation results for this thesis are discussed. The chapter starts with describing the simulation environment used in this thesis including the simulation settings. Results of different simulations for every experiment performed in this study are compared with each other and discussed in separate sections.

5.1 Simulation Models

5.1.1 Network model

The network model used in this study includes a 3GPP LTE supported packet switched network with one e_node_B (evolved node B) covering 3 cells. This LTE network is connected with the internet through a gateway. Mobile users are connected through the LTE radio access technology while the fixed users are connected with a wired connection. Figure 5-1 depicts the network model.

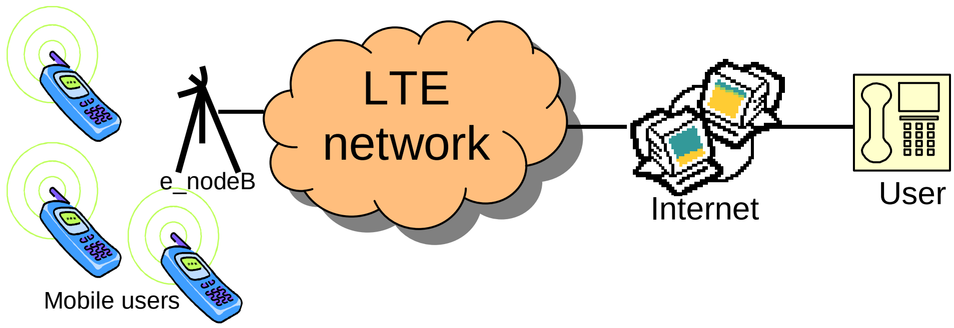


Figure 5-1: Simulated network model

The REDHAWK network simulator has been used in the simulation.

5.1.2 Node model

In the above network model three nodes have been modeled in quite detail; mobile users, fixed users and the e_nodeB.

Both mobile and fixed users share the same model (see figure 5-2), hence they have the same functionalities. The mobile users are allowed to move with various speeds while the fixed users are stationary.

The e_nodeB is regarded as the bottle neck in the network since a radio link is dynamic in its behavior. Radio spectrum is typically also a limited resource. Therefore, ECN marking capability has been implemented in the e_nodeB [28]. The e_nodeB will set the congestion notification bit in the RTP packets based on the queueing time in the scheduler buffer in e_nodeB. Figure 5-3 shows the node model of the e_nodeB used in this study.

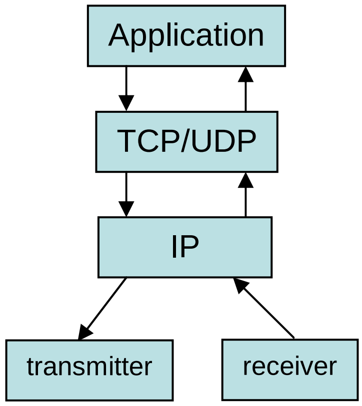


Figure 5-2: User node model

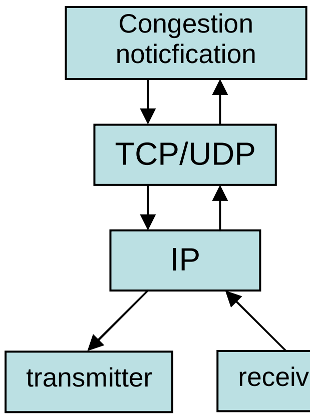


Figure 5-3: e_nodeB node model

5.1.3 Application model

Although the network model can be simulated using the REDWAWK network simulator, the simulator doesn't have support for real video packet. The users (either mobile or fixed) can only transmit and receive simulated RTP packets with dummy data. As in this study perceived video quality is big factor hence sending and receiving real video packet is a necessity. So, to be able to send real video packets an external video client has been built which was attached with the network simulator through UDP sockets to be able to send and receive data to and from the simulator (figure 5-4). The real video client has one sender module and one receiver module. The sender module consists of an h.264 *encoder*, one *RTPpacketizer* which is responsible for creating RTP packets with encoded video data and sending packets, an *RTCPreceiver* which will receive the RTCP RRs and send QoS information to the *Adaptation module* to take necessary adaptive decisions. The adaptive decisions will be fed to the encoder to change the current encoding bit-rate. The receiver module of the real video client consists of one *RTPreceiver*, one *RTCPgenerator* which will generate the regular or early RTCP RRs depending on the output form *ECNwatcher* and an h.264 Decoder which will decode the received encoded RTP packets and write them to the disk in the yuv file format.

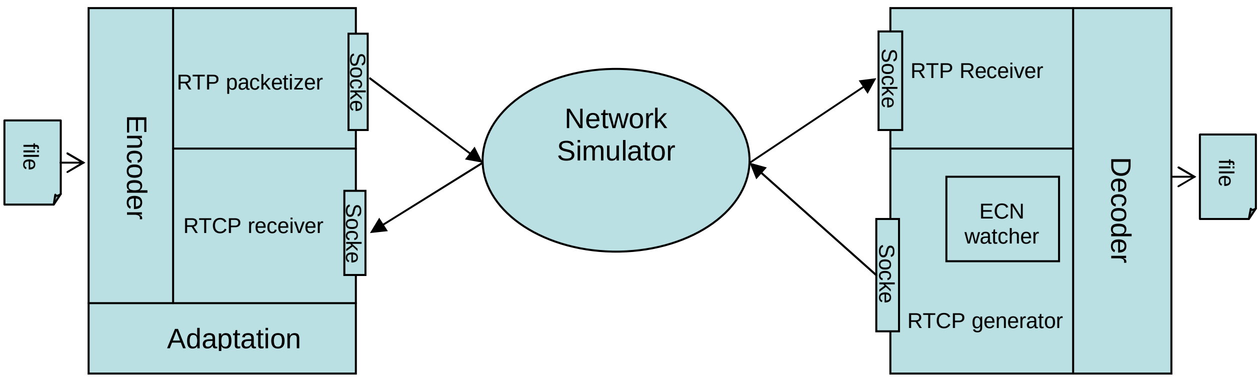


Figure 5-4: Real video client and simulator

5.2 Experimental setup

The LTE network has been configured in REDHAWK with the parameters in table 5-1. The parameter values have been chosen with the objective of representing a deployed, future LTE system. The deployed frequency bandwidth was kept at 1.25 MHz to reduce the system capacity due to some complexity issues of getting band radio coverage.

Parameters	**Values (units)**
Frequency band bandwidth	1.25 (MHz)
Maximum downlink transmission power	20(W)
Number of sub-bands	5
Number of symbols per sub-band	Downlink = 720, Uplink = 660
Number of base station	1
Number of cell per base station	3
Cell radius	500(m)
Number of virtual users	110
Number of real video client	1
Buffer time threshold	.005(s)

Table 5-1: Network setting in simulator

Two different types of users are used in the simulation. The virtual users are the users built into REDHAWK which do not have support for real video streaming. The real video clients are external users who are capable of generating and receiving real video streams. Both kinds of users in the simulation use the same adaptation mechanism. The video target rate defined for virtual user is between 38.4 kbps (minimum) to 128kbps (maximum) and for real video client 30kbps (minimum) to 100 kbps (maximum). Here one thing should be noted that the minimum allowed bit-rate is the 30% of maximum allowed bit-rate. For ease of computation the maximum and minimum bit-rate for real video client was kept 100 kbps and30 kbps respectively and packets are transmitted at 25 frames per second. Three different video sequences have been used in the experiments. Video sequence “Jacob” illustrates a real time video telephony scenario. The” Foreman” sequence also illustrates a video conversation and the last video sequence is “Silent” where the actress is using sign language to communicate. In all three video sequences the quality of video is very important and the viewer should be able to recognize each of the objects without difficulties. All of them have QCIF (176X144 pixels) resolution.

The following PLR table (table 5-2) has been used in the simulations for real video client where the adaptation was allowed. The values in the PLR table were chosen carefully to simulate the slow responsive behavior to the packet loss in this study.

PLR	0%-1%	2%	3%	4%	5%	6%	7%	8%
Bit-rate	100	100	80	70	60	50	40	30

Table 5-2: PLR table used

5.3 Results

The results shown in this section are all based on observing the real video client but it should be noted that the results of real video client includes the effect of different procedure taken to handle congestion at all other virtual users in different cells. Again, all of the plots in this section are generated from same set of data of respective scenario. For technical reasons the whole simulation process needed to be slowed down to avoid packet losses in the communication socket between real video client and network simulator which broke the synchronization of fixed frame sending rate in time division. Thus the results of bit-rate and packet loss are presented with respect to packet sequence number. This has also made the x axis of different plots variable in length as the change in bit-rate will also change the number of total packet send during the session. But the PSNR calculation is independent of this fact and is a post processed result.

5.3.1 Experiment 1 – Suitability of proposed adaptation scheme

5.3.1.1 No congestion control, No ECN (NN)

In this simulation no congestion control was used which means that the sender will not react to any ECN or packet loss reported through RTCP RR. The maximum bit-rate is used throughout the session.

- Video sequence – Jacob

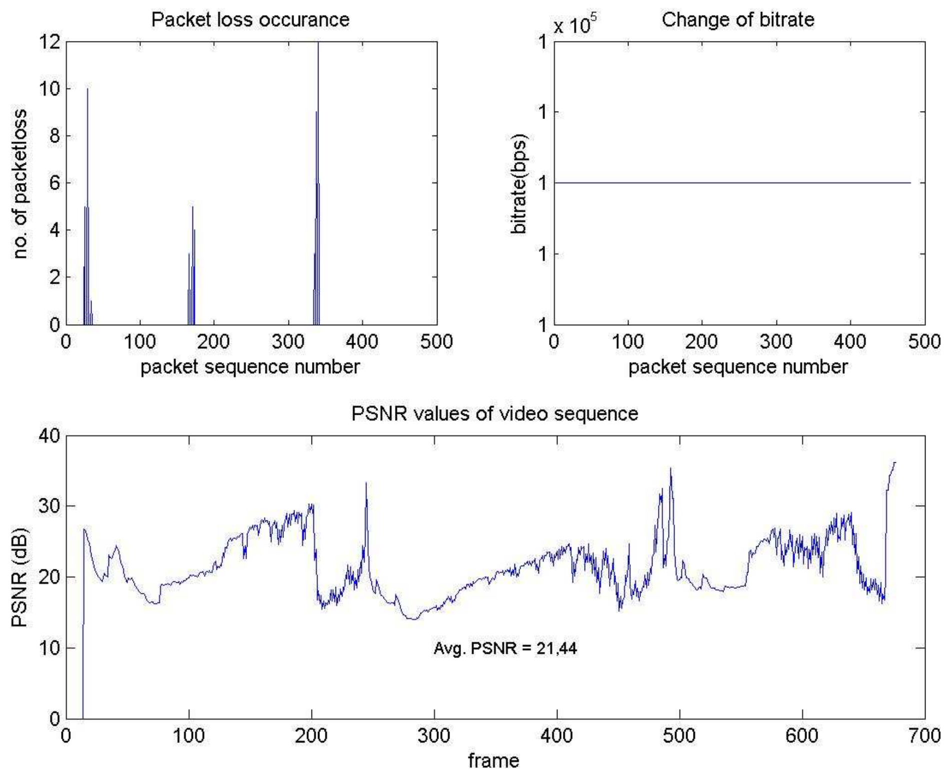


Figure 5-5: simulation result with no adaptation and no ECN for Jacob video sequence

We can see there was no change in the bit-rate (figure 5-5) and total 52 packets (10.9% of total packets) have been lost during the session. The PSNR values also represent the effect of packet loss during the session and the dips in the PSNR value represent the packet loss effect during the session.

- Video sequence – Foreman

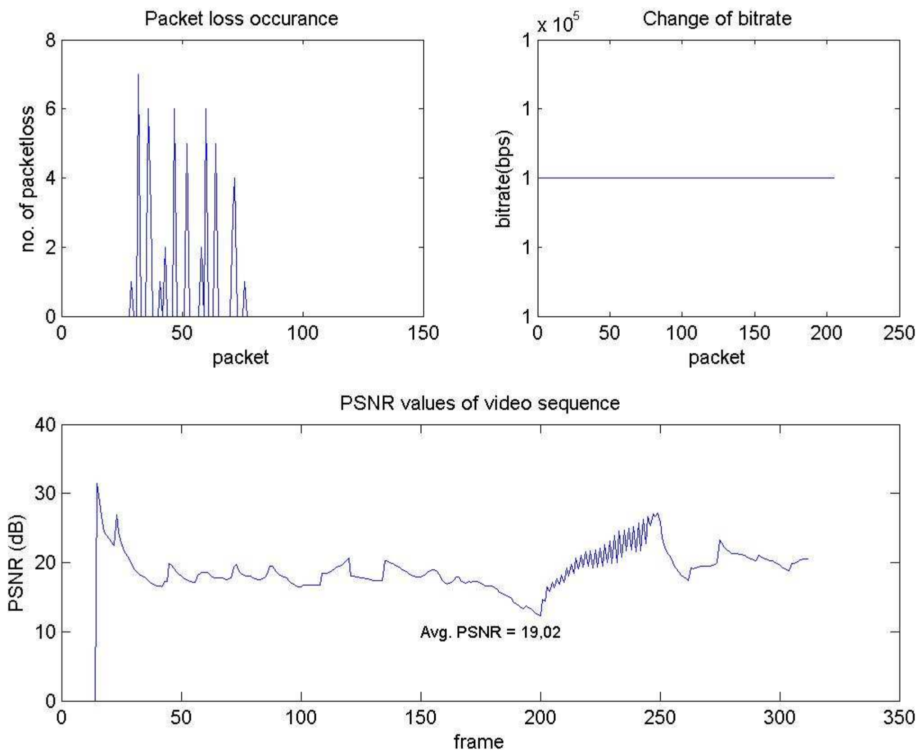


Figure 5-6: simulation result with no adaptation and no ECN for Foreman video sequence

Figure 5-6 shows the number of packet loss occurred during this session. The effect of this packet loss can be realized by looking at the dip of PSNR values. PSNR values have gone down when the packet loss occurred and stayed low for a longer period as the packet losses were observed for a longer period of time. As expected there was no change in bit-rate.

- Video sequence – Silent

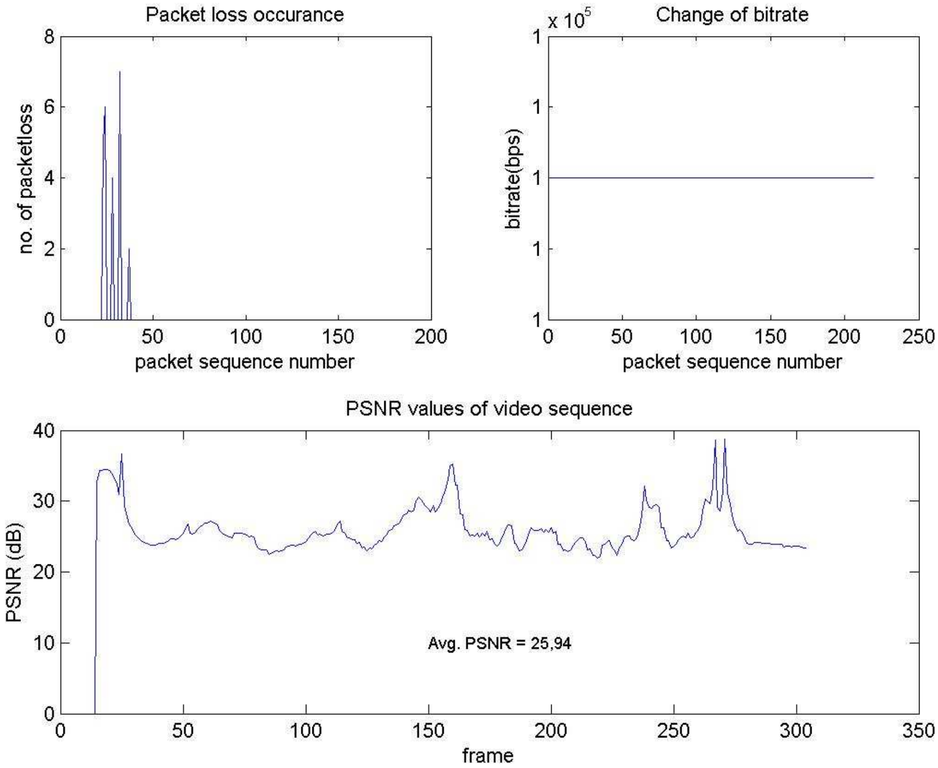


Figure 5-7: simulation result with no adaptation and no ECN for Silent video sequence

In this session we observe packet loss at the beginning of the session. The average PSNR value is 25.94 which is better than that of other two sessions due to a lower packet loss rate.

5.3.1.2 Packet loss based congestion control – No ECN (NE)

In this simulation a congestion control algorithm only based on packet loss is used. The PLR has been calculated according to equation 4-1 and corresponding bit-rate was used accordingly table 5-2.

- Video sequence – Jacob

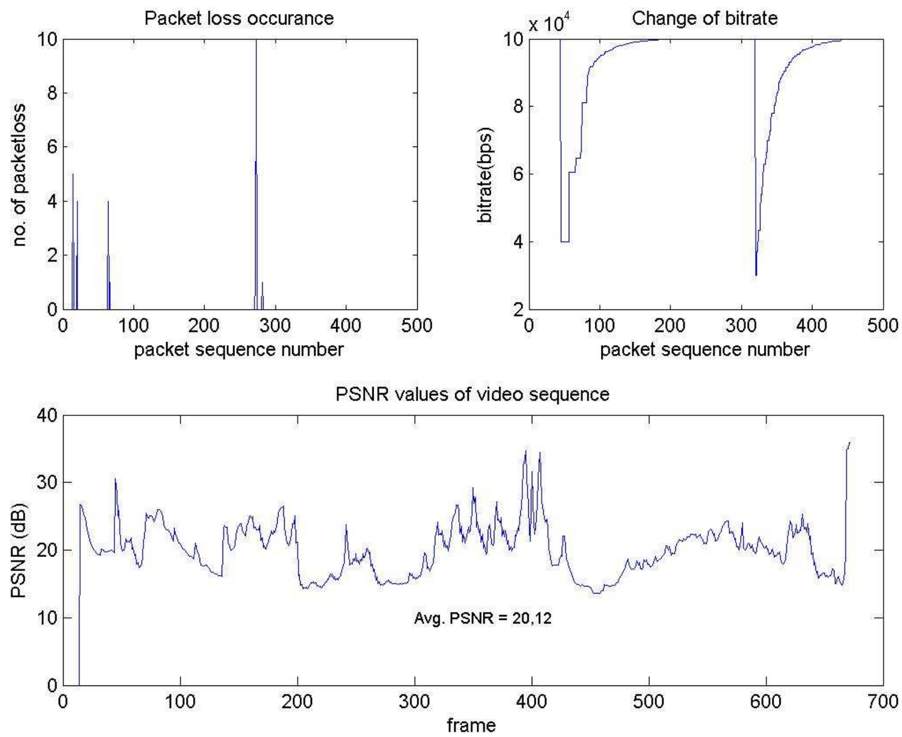


Figure 5-8: simulation result with only packet loss based adaptation for Jacob video sequence

Figure 5-8 shows that the packet loss based adaptation has reduced the total number of packet losses. But as this adaptation scheme only reacts after the loss occurs, it cannot remove the possibility of packet losses. Looking at figure 5-8 we can see an initial rate reduction when the first packet loss occurred. The sender has then restored the bit-rate before the next packet loss event occurs. Although the number of packet loss has been reduced by this adaptation scheme, the average PSNR increase is quite small which was unexpected. We can also see that there is a delay in response to the congestion. This delay depends on the RTCP RR interval and delay in generation of RTCP RR at receiver and reception of that report at sender.

- Video sequence – Foreman

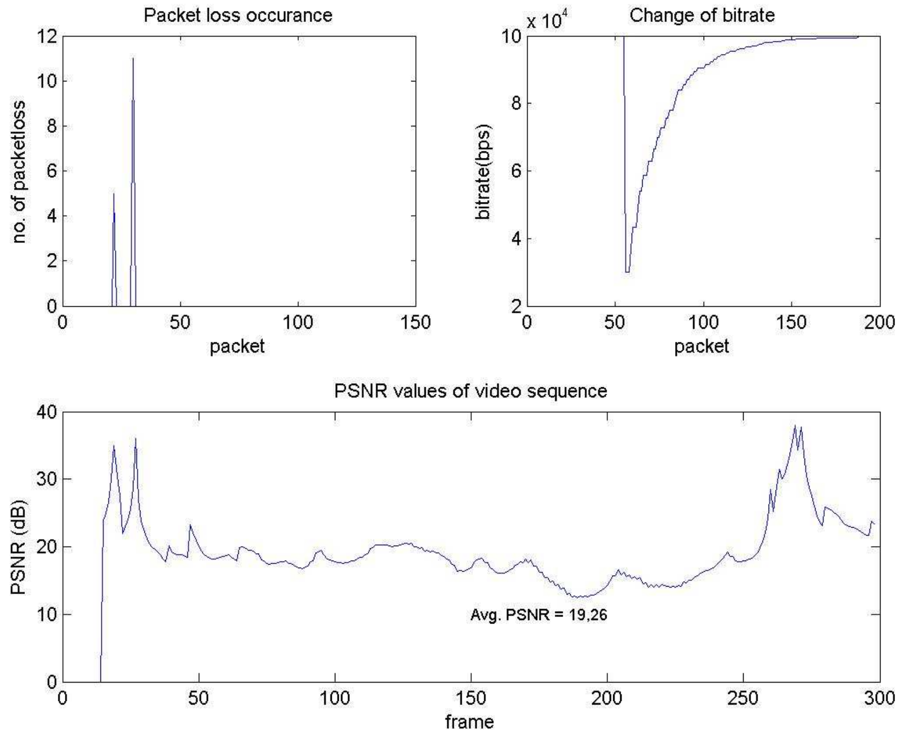


Figure 5-9: simulation result with only packet loss based adaptation for Foreman video sequence

The packet loss was substantially reduced by this congestion control mechanism compared to the packet loss observed in simulations with no congestion control. We can see that the sending bit-rate was reduced down to 30 kbps and then gradually restored to the initial bit-rate as there was no more packet loss. The rest of the session was run with the higher bit-rate.

- Video sequence – Silent

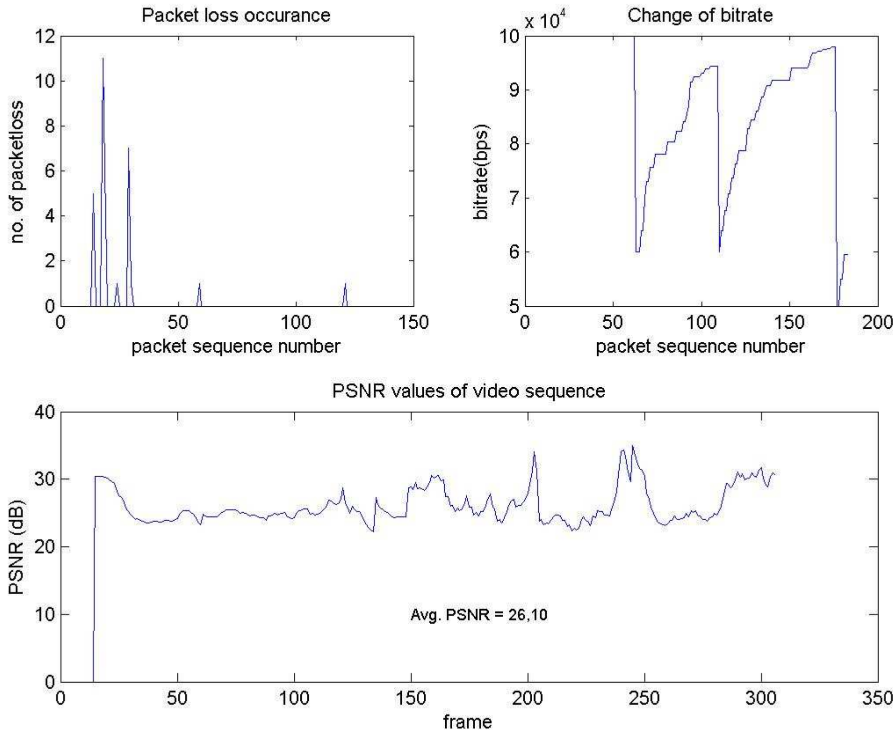


Figure 5-10: simulation result with only packet loss based adaptation for Silent video sequence

In this case the total number of packet losses has not really been compensated by the scheme and the bit-rate has not been reduced drastically when the packet loss was observed. Instead, it was gradually reduced to the lowest bit-rate at the end of the session. The average PSNR value is 26.10 which is not much higher than that with no adaptation and ECN (avg. PSNR = 25.94). Although the total number of packet loss has not been reduced significantly, further losses in the session have been compensated after the adaptation.

5.3.1.3 ECN based congestion control – No Adaptive increase (NA)

Only ECN is considered as the adaptation triggering event in this simulation. Each time the sender gets a TMMBR message as a result of an ECN, the sender will reduce its bit-rate to 50% of current bit-rate until it reaches the minimum allowable bit-rate but the bit-rate was not restored.

- Video sequence – Jacob

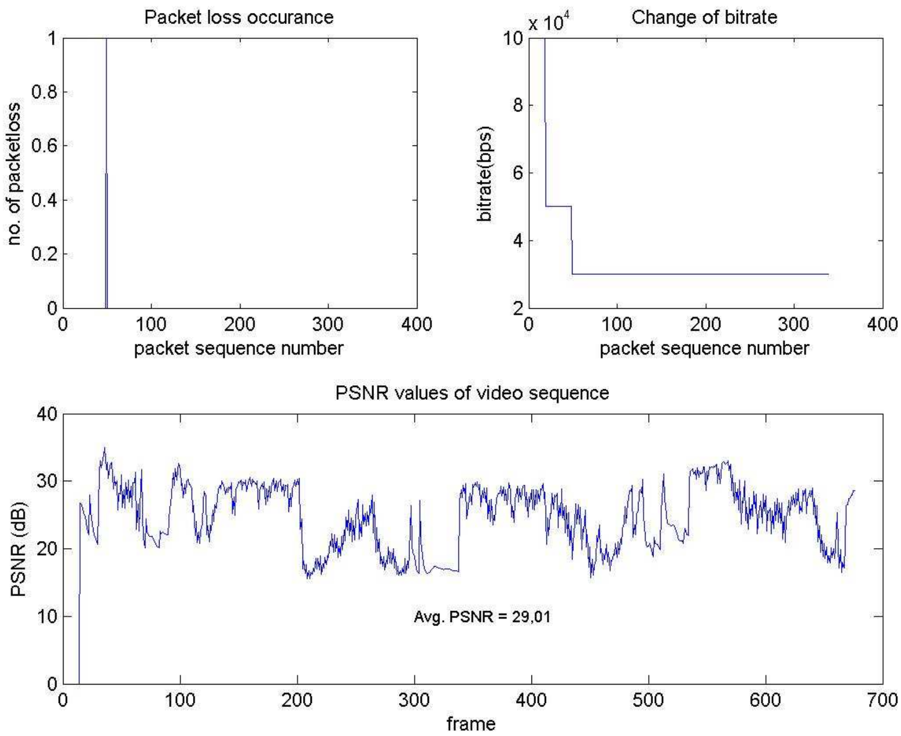


Figure 5-11: simulation result with only ECN based adaptation for Jacob video sequence

As we can see in figure 5-11 there were some packet losses but the number is reduced to only one packet loss. The PSNR value has also increased in this case. But we can see that ECN did not prevent the initial packet loss in this case

- Video sequence – Foreman

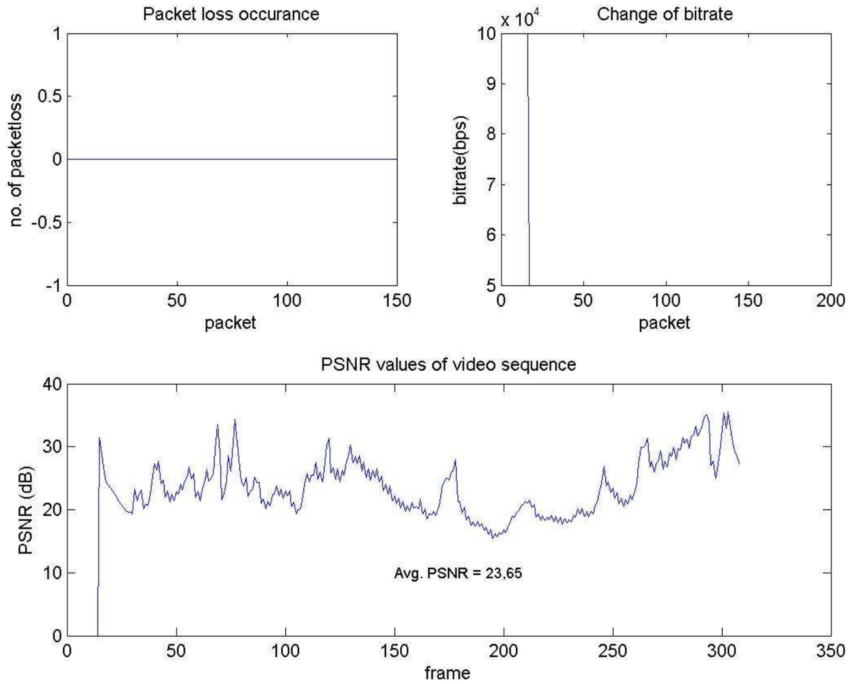


Figure 5-12: simulation result with only ECN based adaptation for Foreman video sequence

In figure 5-12 we can see there is no packet loss observed during the whole session and by looking at the bit-rate change it can be seen that the bit-rate has gone down very early in the session and as there was no probing it remained same for the rest of the session. This might have lead to no packet loss during the session. Hence the average PSNR value has been raised.

- Video sequence – Silent

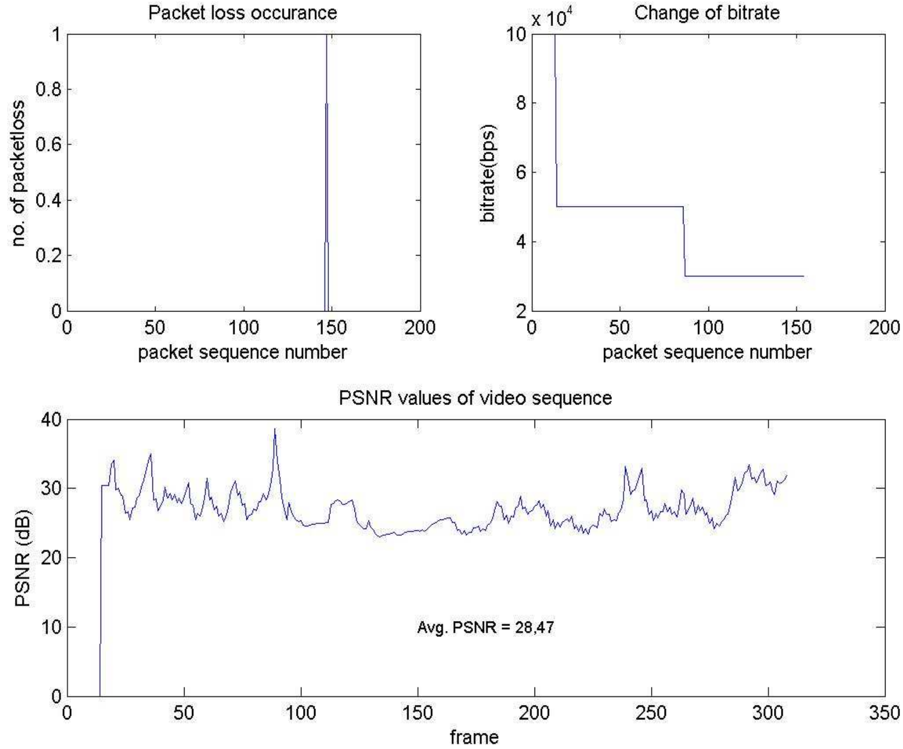


Figure 5-13: simulation result with only ECN based adaptation for Silent video sequence

Figure 5-13 shows that going down in bit-rate upon receiving the early congestion notification has reduced the packet loss at the beginning of the session if we compare it with the simulation results with no adaptation and ECN. But there was a packet loss at the end of the session. As the packet loss was prevented early in the session that gave a raise to the average PSNR value (28.47) compared to the other simulation results in both case of without congestion control and with only packet based congestion control.

5.3.1.4 Proposed congestion control- with Adaptation and ECN (AE)

The congestion scheme used in this simulation is the proposed congestion control scheme where both ECN and packet loss or either of them can be used as the adaptation triggering event. When the sender gets a TMMBR message form the receiver it lowers its bit-rate to 40% of current bit-rate and when the network condition allows, it goes up in bit-rate slowly taking 10% of the difference of current bit-rate and maximum-bit-rate as the increasing step.

- Video sequence – Jacob

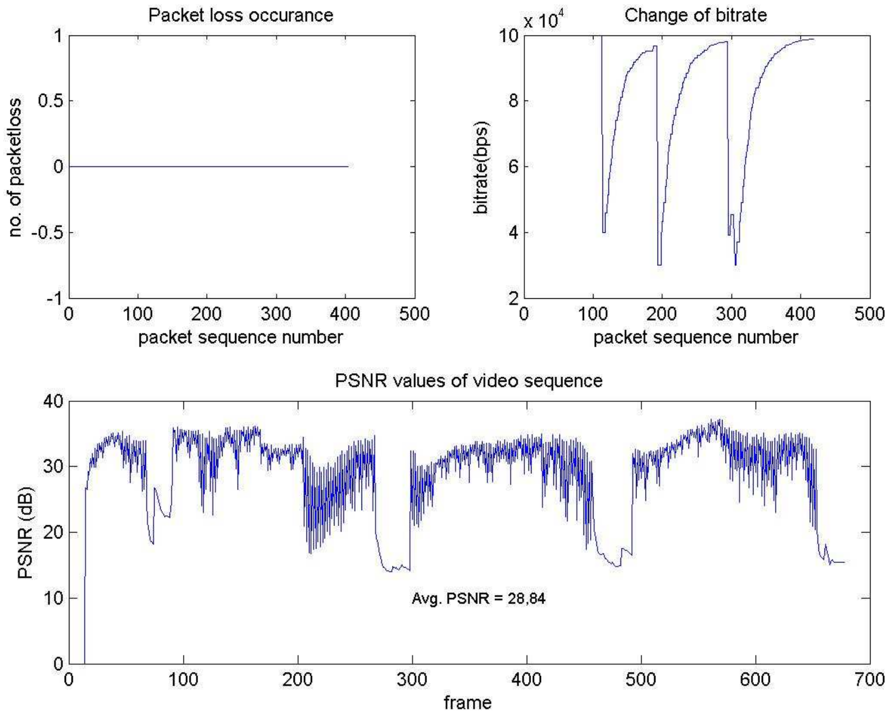


Figure 5-14: simulation result with proposed adaptation scheme for Jacob video sequence

Looking at the figure 5-14 we can see that there was no packet loss during the whole session. We can see the fluctuation of bit-rate going down and up during the middle of the session. The bit-rate is going down up to 30 kbps which is the minimum allowed operable bit-rate. The average PSNR value has also increased compared to the other adaptation scheme and there is 7.4% rise in average

PSNR than that of with no congestion control. This result can be taken as an indication of the suitability of the proposed scheme.

- Video sequence – Foreman

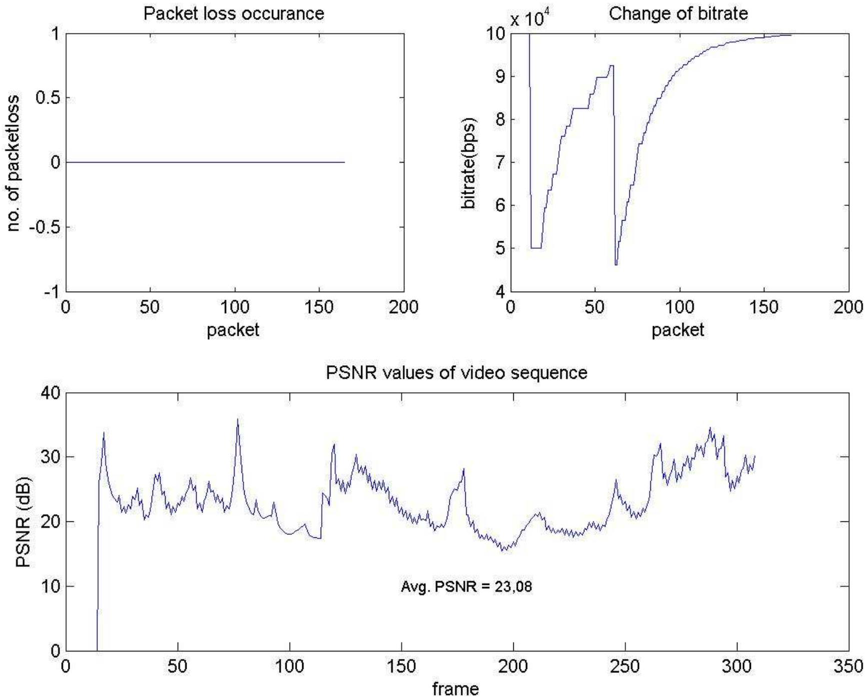


Figure 5-15: simulation result with proposed adaptation scheme for Foreman video sequence

The figures here resemble the result we have observed for Jacob sequence. There was no packet loss observed during the session and there is a little improvement in the average PSNR in compared to no adaptation and no ECN scheme. The bit-rate was changed two times with the arrival of two TMMBR message from receiver.

- Video sequence – Silent

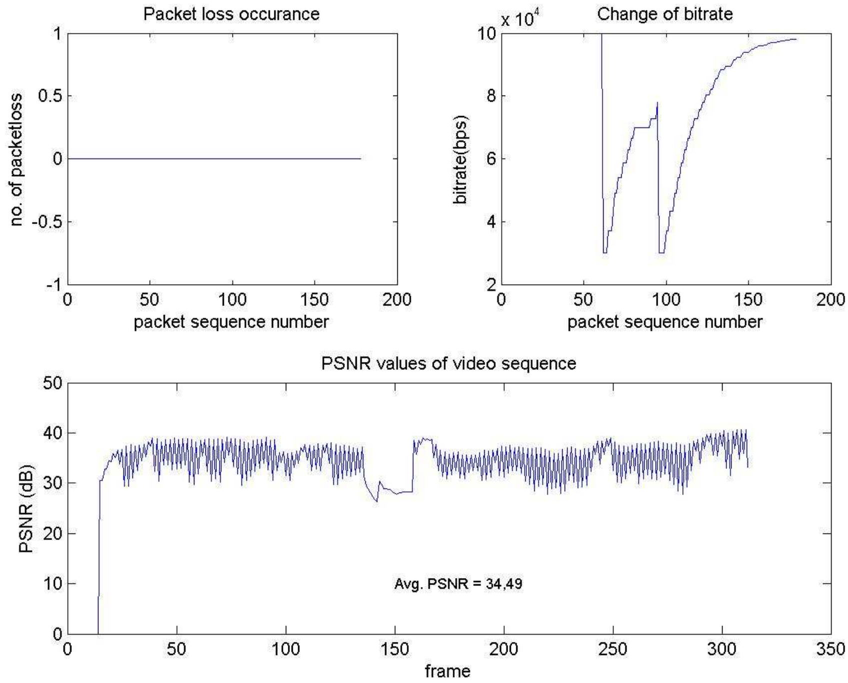


Figure 5-16: simulation result with proposed adaptation scheme for Silent video sequence

Here we can see a similar result that we have seen in the case of Jacob video sequence. The average PSNR has increased compared to the other simulations and compared with the case of no congestion control it gave a quite significant increase in avg. PSNR.

5.3.1.5 Discussion

Form the results presented in the previous section; it is evident that the proposed hybrid congestion scheme which adapts the video transmission bit-rate with both ECN and packet loss outperforms the other schemes studied here. Figure 5-17 to 5-19 show different QoS matrices values that has been recorded with the different adaptation schemes. When comparing with the results when no congestion control mechanism was used, it has produced remarkably good results. It can also be noticed that only going down in bit-rate upon receiving TMMBR message does produce good results in terms of number of packet losses and

PSNR but it cannot fully guarantee the avoidance of further packet loss (see the results in section 5.3.1.3 for the foreman video sequence). In this case all active users will go down in bit-rate and not restore the bit-rate which will release a large amount of network resources and will eventually give better network conditions. However, running at a lower bit-rate might lead to video quality degradation. But the PSNR value in the foreman sequence for NA does not reveal this fact which lead us conclude that in order to get reliable results in terms of video quality subjective video quality tests are needed. Additionally, an ECN marked packet can also be lost which eventually will needs to be taken into account when designing the adaptation scheme. Hence, there is a need of two different back off states and the simulation results validated the two different states in the proposed adaptation scheme.

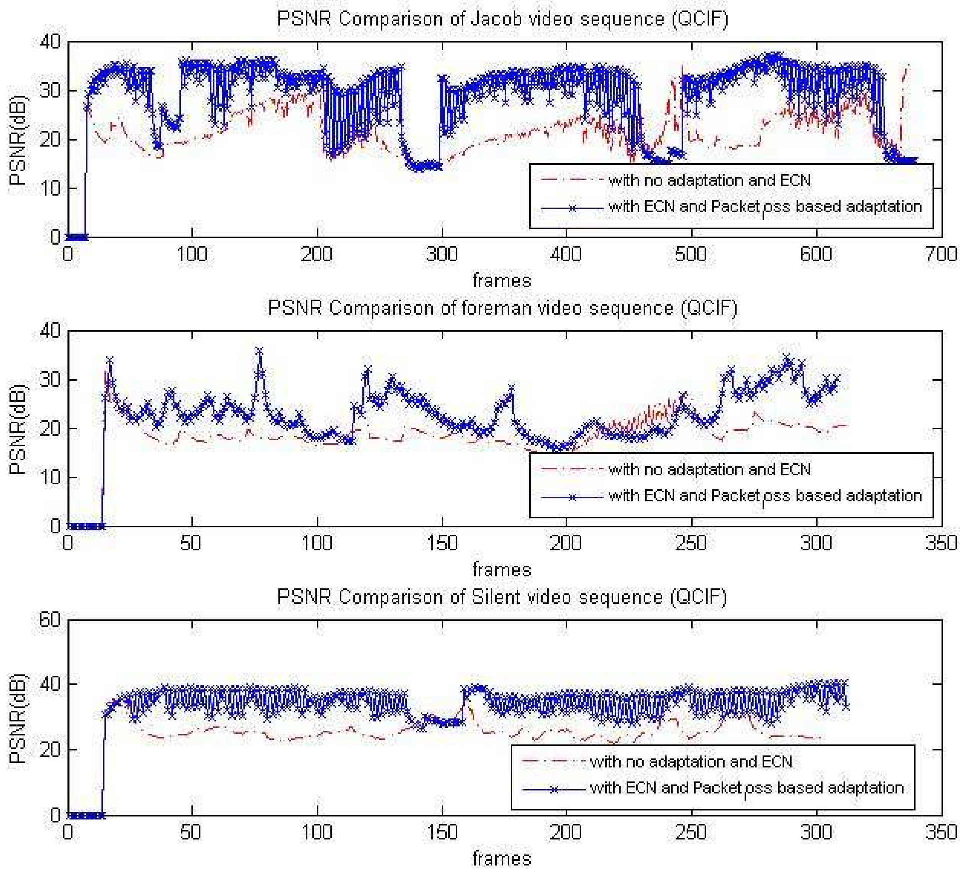
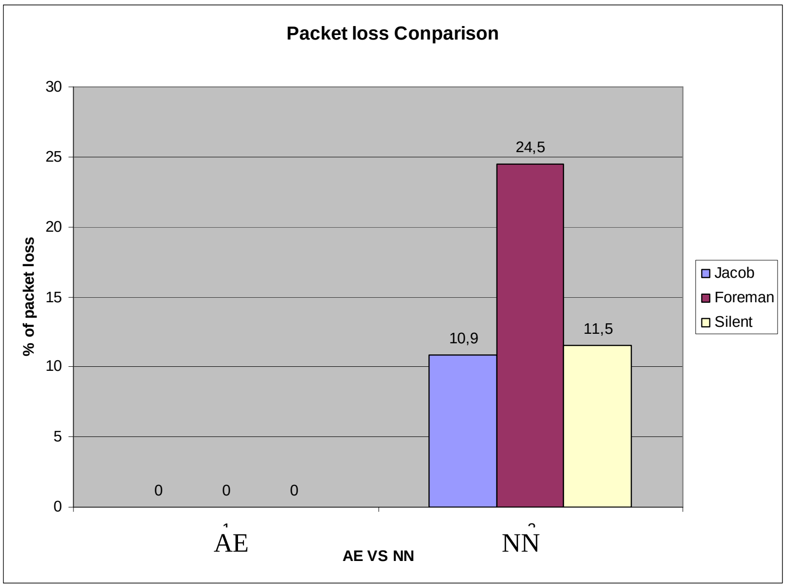


Figure 5-17: NN VS AE PSNR comparison for all video sequences



AE = With Adaptation and ECN
NN =With No Adaptation and No ECN

Figure 5-18: NN VS AE Packet loss comparison for all video sequences

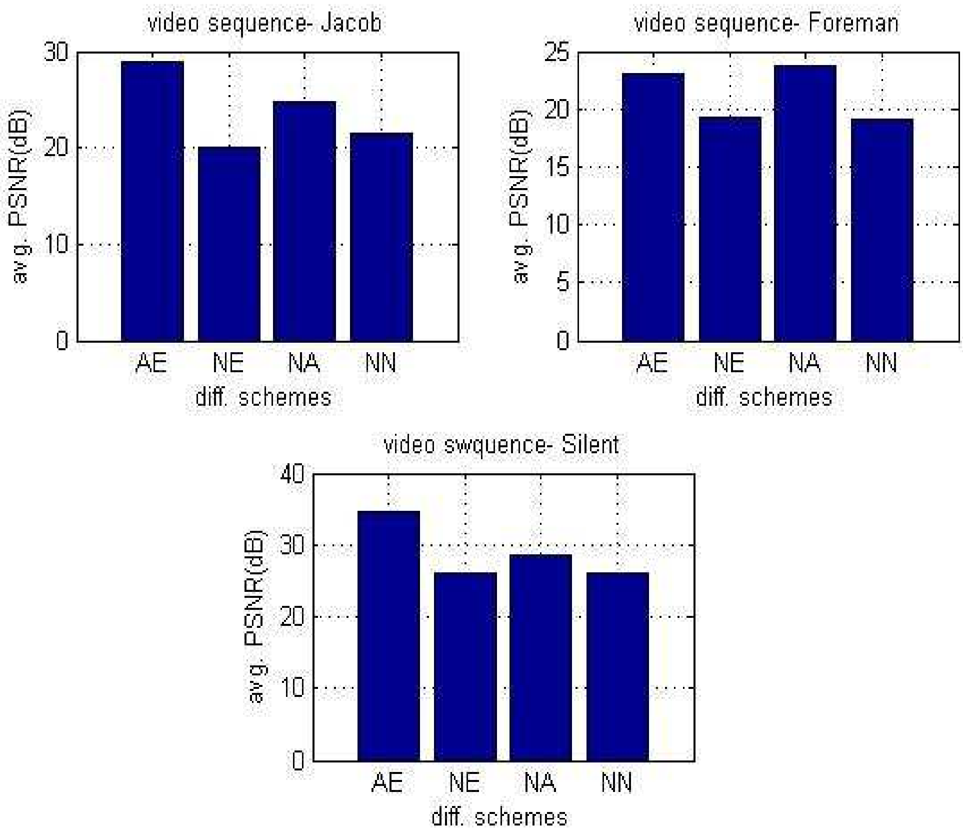


Figure 5-19: avg. PSNR comparison for all video sequences

5.3.2 Experiment 2 – Suitability of Fast decrease and Slow increase in bit-rate

This section describes the simulation results with the different bit-rate decrease and increase schemes that have been studied. Although all simulations have been performed with all three video sequences, only results for the Jacob video sequence are presented here. The results were similar for all three video sequences.

5.3.2.1 Fast bit-rate Decrease Fast bit-rate restore Up (FDFU)

In this simulation, whenever a TMMBR was received, the bit-rate was reduced to the lowest allowed bit-rate and a fast bit-rate restore method was used to restore the current bit-rate to initial bit-rate. This can be denoted as an AIMD adaptation scheme. The sender increases its bit-rate by 15 kbps each time it probes to go up in bit-rate and decreases the bit –rate to a constant floor when a reduction is needed.

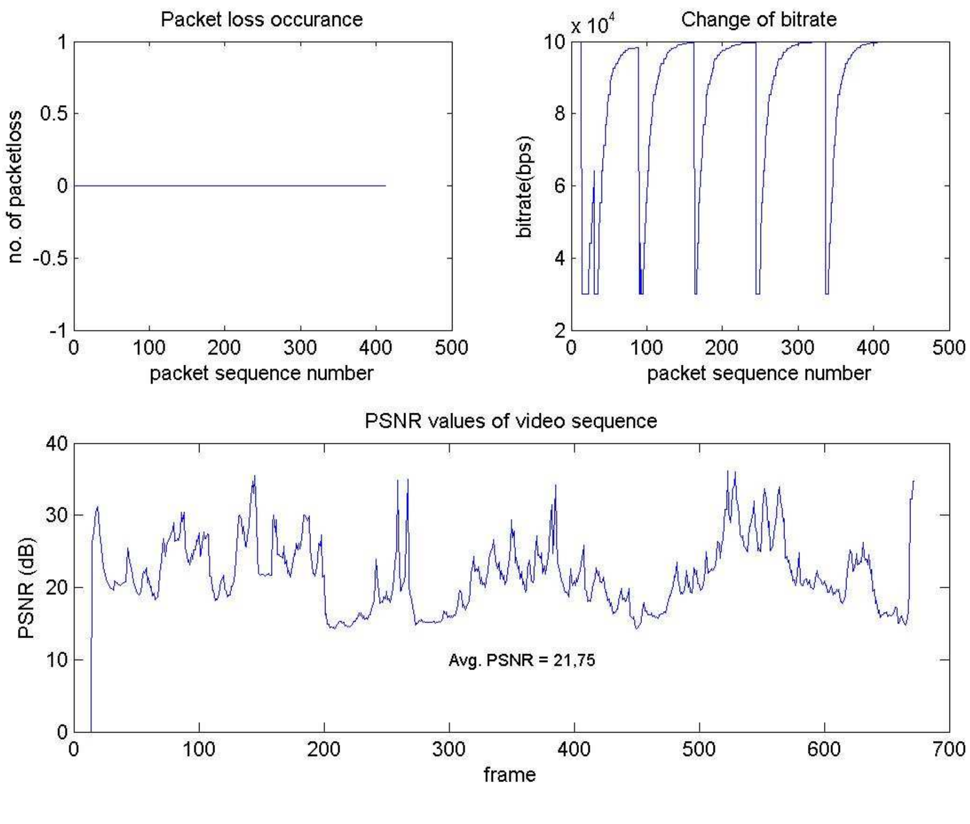


Figure 5-20: FDFU results for Jacob video sequence

In figure 5-20 we can see an oscillating behavior in the bit-rate which might not be a desirable behavior from an adaptation scheme as this will have an effect on video quality and can be annoying experience for the user. The PSNR values show this fact. The average PSNR is 21.75 which is very close to the average PSNR value (21.44) for this sequence with no adaptation and no ECN though there was no packet loss observed here. This indicates another aspect that there is a connection between fast bit-rate change and PSNR values.

5.3.2.2 Fast bit-rate Decrease Slow bit-rate restore Up (FDSU)

Here, the TMMBR message would suggest the sender to go down in bit-rate fast but the restoration process is slow. Reducing the bit-rate due to packet loss has been done according to the PLR calculation but the restoration was always slow.

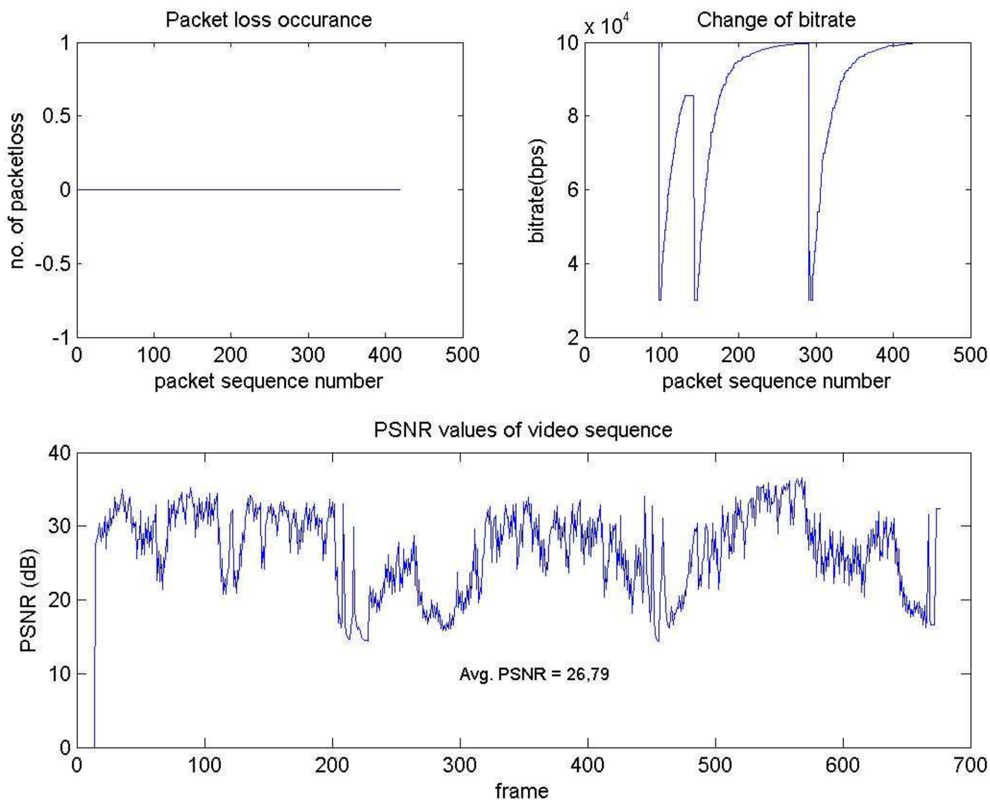


Figure 5-21: FDSU results for Jacob video sequence

We can see in figure 5-20 that the oscillation in bit-rate has reduced and there is a certain raise in PSNR values which averages 26.79. No packet loss was observed during the session.

5.3.2.3 Slow bit-rate Decrease bit-rate Fast restore Up (SDFU)

In this simulation a slow decrease and fast increase procedure has been adopted. The sender now goes down in bit-rate to 75% of current bit-rate and goes up in bit-rate with a step of 15kbps each time. So if the sender is currently running at 100kbps it will go down to 75 kbps upon receiving a TMMBR from the receiver.

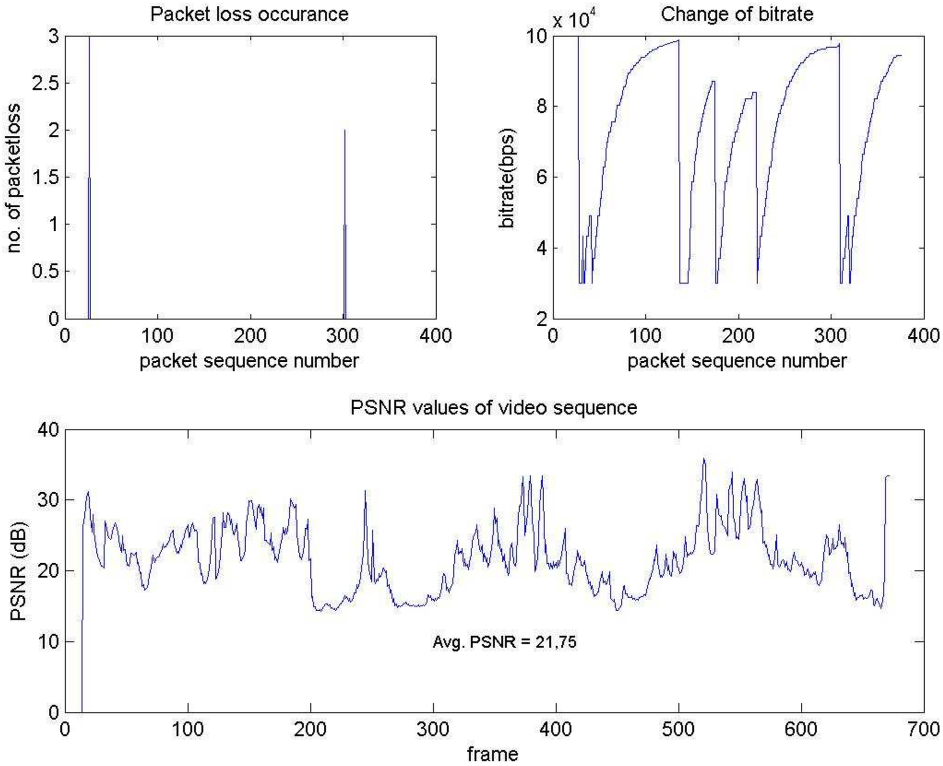


Figure 5-22: SDFU results for Jacob video sequence

The SDFU approach resulted in a number of packet losses, fast bit-rate oscillation and lower average in PSNR value. At the beginning of the session the bit-rate is going down to minimum because of the packet loss observed. Afterwards, we can see much bit-rate oscillation. In the rest of the drops in bit-rate are due to successive ECN except the last straight drop in bit-rate which was also due to packet loss. The oscillating behavior has resulted in the decrease in average PSNR. Looking at the bit-rate versus packet sequence number graph it is evident that going down slowly does not help eventually the bit-rate has gone down to minimum level before it starts to probe to restore the bit-rate.

5.3.2.4 Slow bit-rate Decrease Slow bit-rate restore Up (SDSU)

In this simulation sender is allowed to go down in bit-rate slowly and restore the bit-rate slowly. The sender now goes down in bit-rate to 75% of current bit-rate and restores the bit-rate with step of 10% of difference between current bit-rate and maximum allowable bit-rate.

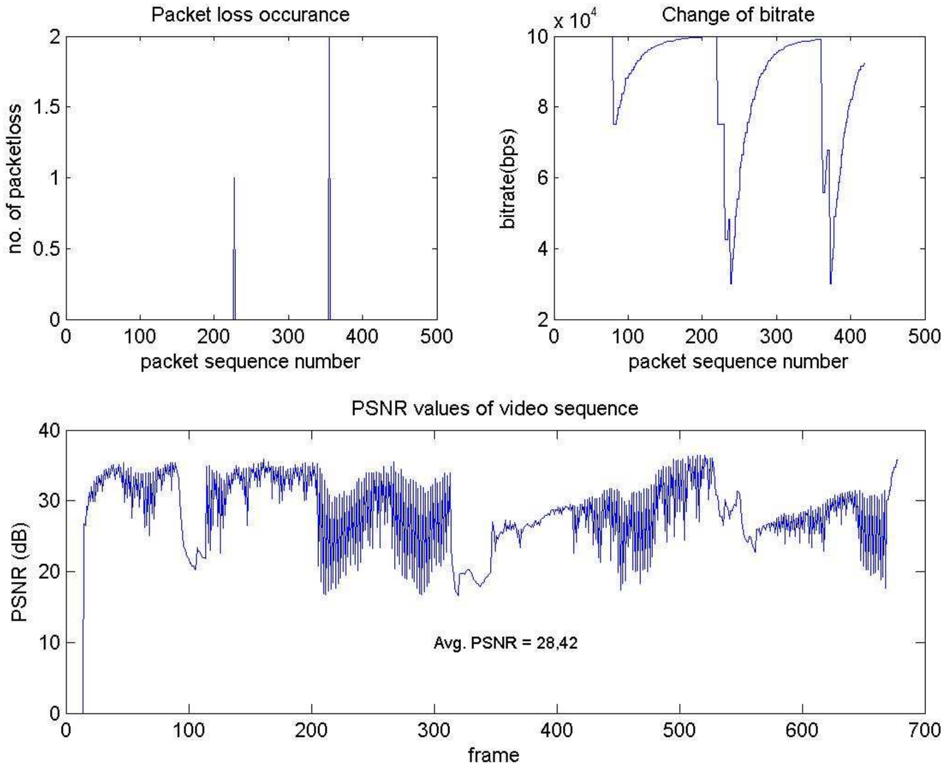


Figure 5-23: SDSU results for Jacob video sequence

Figure 5-22 shows that the bit-rate is ultimately going down to lowest possible bit-rate and we also observe packet loss though the total number of packet loss is only 3 packets for the whole session. The PSNR values are high which is quite surprising.

5.3.2.5 Discussion

The results shown in this section proposes that the FDSU method is the most suitable way to address the ECN which by turn means that the TMMBR messages need to be handled aggressively to avoid future packet loss in the session. The graphs here also disclose the fact that no matter how you handle the TMMBR messages (aggressively or non-aggressively) restoring the bit-rate slowly always have better result in terms of packet loss, PSNR and bit-rate oscillation than those of restoring it in a faster way.

5.3.3 Experiment 3 – Different bit-rate decrease steps

In the previous experiment it was established that the sender has to address the TMMBR messages aggressively and the restoration of bit-rate should be a slow process. In this simulation different ways of reducing the bit-rate relative to the current bit-rate have been observed. Also in this experiment, all the video sequences provided similar results hence only the results for the Jacob video sequence are presented here.

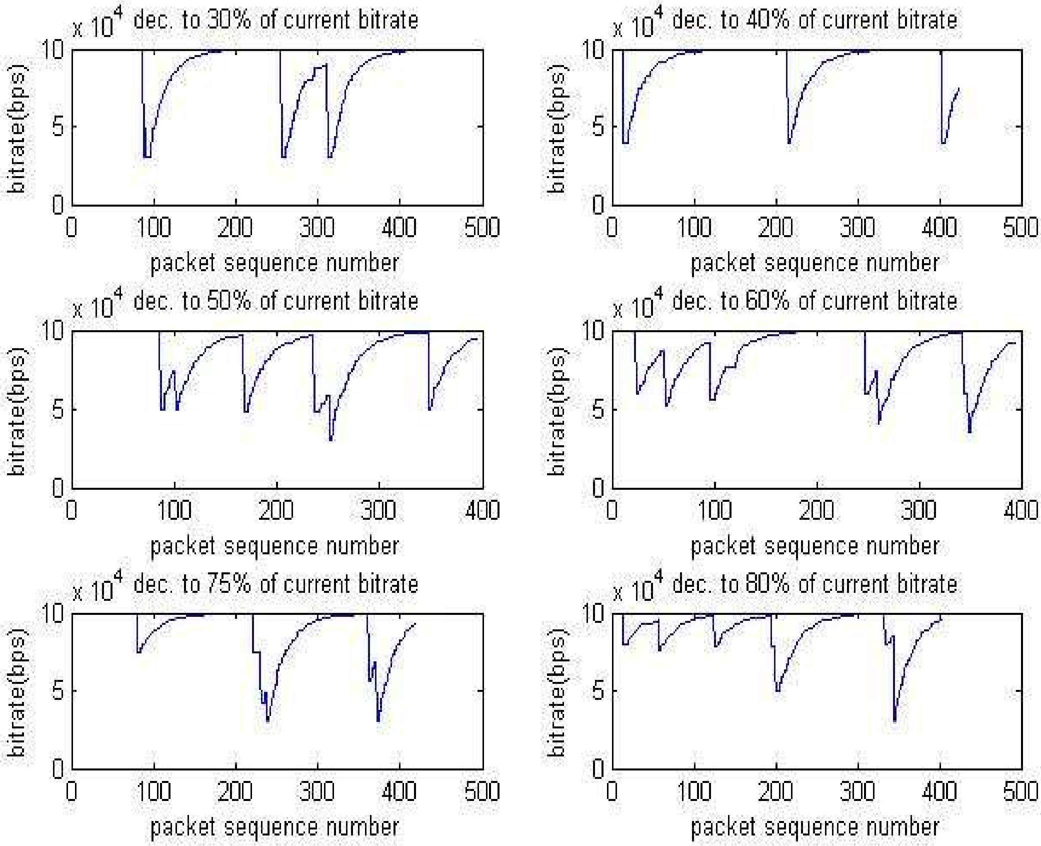


Figure 5-24: Bit-rate oscillations with different decreasing steps

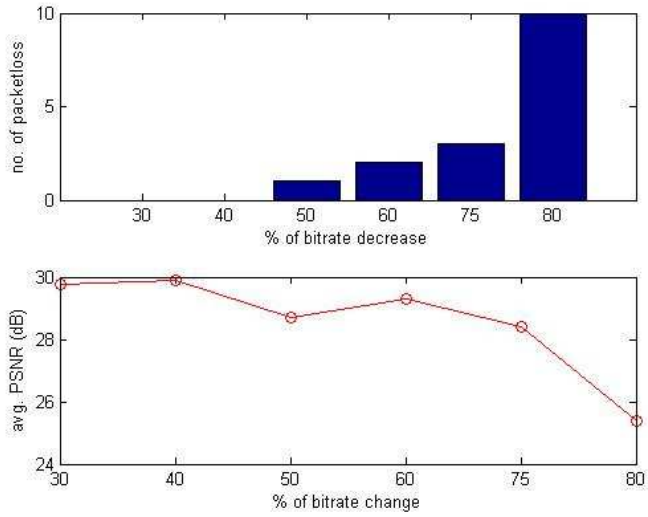


Figure 5-25: Packet loss and avg. PSNR with different decreasing steps

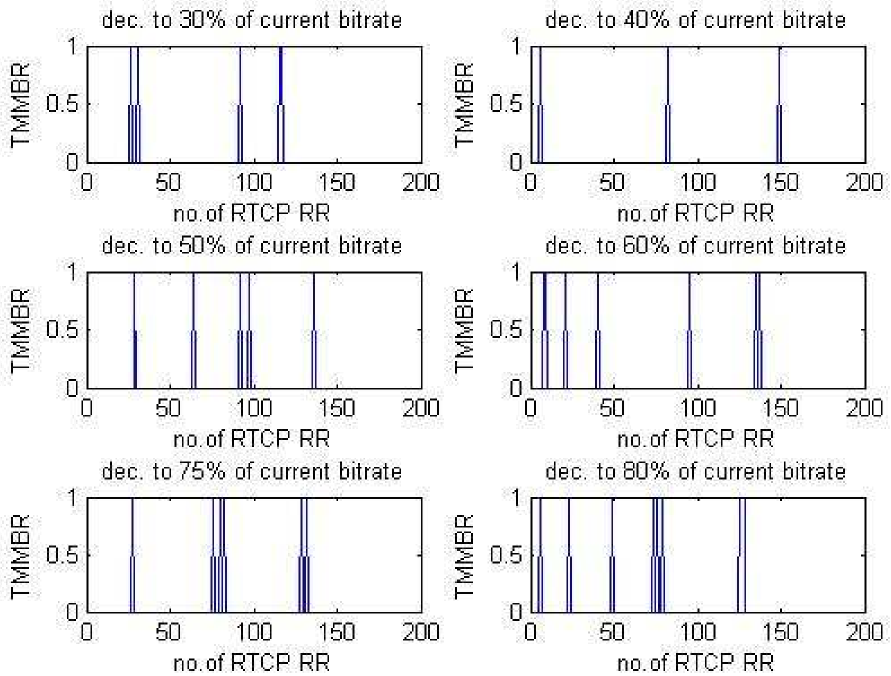


Figure 5-26: no. of TMMBR messages with different decreasing steps

Figure 5-23 shows the bit-rate oscillation with different steps of going down in bit-rate. It can be observed that the faster we reduce the bit-rate upon receiving a TMMBR message, the less bit-rate oscillations occur. Again, we can see that if we go down to a value above 50% of current bit-rate we might observe more packet losses and the amount of packet loss increases with lower decreasing steps (see figure 5-24). The highest PSNR value is achieved when we go down to 40% of the current bit-rate. As the packet loss increases we can also see the decrease in PSNR values as expected. The number of TMMBR messages in figure 5-25 also shows that going down in bit-rate more aggressively is also network friendly as it will release the resources faster, hence there will be less necessity of marking more packets in the e_node_B.

5.3.4 Subjective Quality Evaluation

To complement the results that we have found by comparing the PSNR values, an informal subjective video test has been performed. The test used the processed video clips of the three different video sequences which were recorded with the different adaptation schemes described in section

5.3.1. Due to lack of time it was not possible to design a full subjective video test. Instead, an informal method was used where all the video sequences were sent to different persons working in Ericsson, most of them are experts' in subjective video and audio testing. The test persons were asked to rate the video clips using a five-grade scale (see table 5-3). To break the chain of expectation the file clips ware randomly changed. The MOS scores for the different video sequences are shown in the figures 5-26 through 5-28.

MOS	Quality
5	Excellent
4	Good
3	Fair
2	Bad
1	Poor

Table 5-3: MOS scores

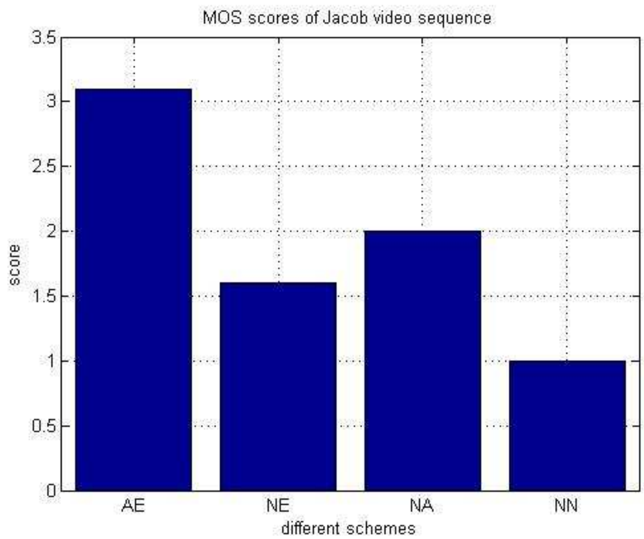


Figure 5-27: MOS scores with different schemes for the Jacob video sequence

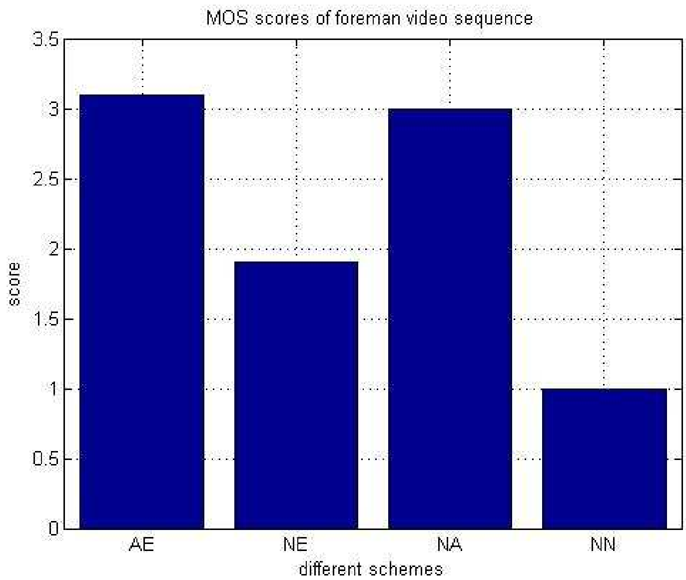


Figure 5-28: MOS scores with different schemes for the Foreman video sequence

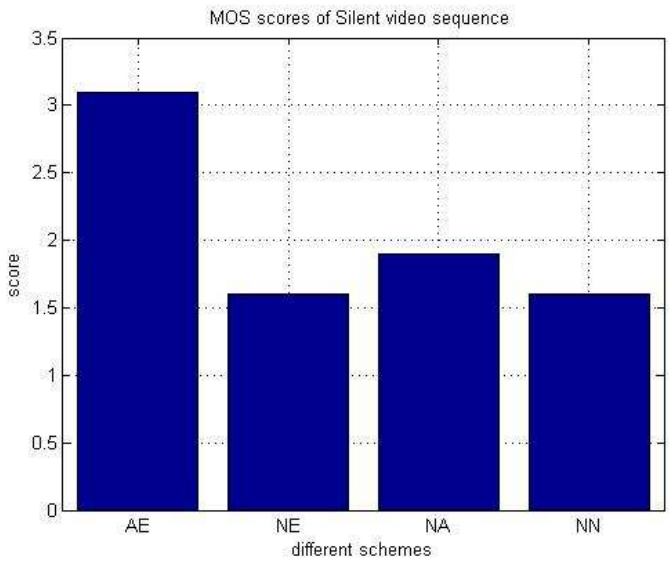


Figure 5-29: MOS scores with different schemes for the Silent video sequence

From the figures we can see that in all the three cases the MOS scores with the proposed adaptation scheme are higher than that of the other schemes. In the case of the Jacob and Silent video sequences the results are same as the objective test (PSNR values). But in case of Foreman video sequence, the MOS score of AE is higher than MOS score of NA which indicates that both the video clips have roughly same quality. This makes a statement that for all video sequences only looking at PSNR values in case of rate adaptation is not a good practice. This can be further explained in this way. In objective tests like PSNR calculation the values calculated are heavily dependent on correlation between the frames of the reference video sequence and the video sequence to be tested. A significant change of bit-rate typically changes the frame rate hence breaking the synchronization of the correlation which can lead to a misleading conclusion.

Chapter 06

6. Conclusion and Future Work

This chapter describes different findings and conclusion made from this study and finishes by stating different future work those can be performed related to this study.

6.1 Conclusion

Congestion control is very important for all kinds of traffic to secure high media quality and efficient resource utilization in all-IP networks. This is especially important in IMS multimedia telephony and next generation mobile network such as LTE.

For real-time traffic transported via RTP/UDP/IP, RTCP feedback can be utilized in various ways to shape the traffic in response to congestion. In this thesis it has been shown that the best results can be achieved by using an aggressive RTCP TMMBR feedback together with ECN. Although, there is no standard to use ECN for UDP traffic it has been observed that other rate adaptation proposals are not always compatible with other standards available which makes the use of rate adaptation hard to deploy universally. The only solution on how to guarantee proper use and open up congestion control mechanisms to rate adaptation is through standardization in organizations such as 3GPP and/or IETF. Looking at the previous work done on ECN for UDP traffic it can be said that this topic has not been well investigated yet. Hence, introducing support for ECN for UDP traffic marking in the LTE environment should be further investigated.

In this thesis the above discussed facts are taken into consideration and emphasis has been given on real-time video transmission over 3GPP LTE network with an aim to see the effect of pre-warning of congestion on the perceived video quality. The goal was to see that whether going down in bit-rate depending on early congestion notification improves the video quality rather allowing packet losses to occur and then initiate counter measures to reduce future packet losses. In the process of achieving this goal, a new adaptation state machine was designed and developed with different strategies to address the early congestion notification and packet losses. Simulations have been carried out with different adaptation schemes to compare results and benefits of the newly proposed scheme. Three different kind of video sequences are used in all experiments. These video sequences have been carefully selected to span typical use cases for a real-time conversational video service. The simulation results have been shown that the early congestion notification coming from the network need to be addressed aggressively otherwise even an early notification will not be able to guarantee a packet loss free transmission. Further, going up in bit-rate to restore the bit-rate to its initial value should be a slow process. It has been seen in the simulation results that a fast restoration process does not improve the congested scenario in terms of bit-rate oscillation, packet loss and PSNR. Moreover, the results also justifies that only ECN based adaptation or only packet loss based adaptation is not enough to experience a packet loss free video transmission over LTE network; both are needed. Both informal subjective video analysis and objective video analysis has shown that the proposed scheme performs better than the other solutions studied in this thesis.

This study shows the effectiveness of early congestion notification for real-time video transmission and reveals some important aspects of adaptation procedure which need to be taken into account when designing adaptation for video traffic to avoid the congestion in the LTE network.

6.2 Contributions

This thesis has following contributions-

- It proposes a new adaptation scheme with 5 states to handle the congestion in IP networks.
- It devices a new bit-rate reduction and restoration approach named "Distance Dependent Adaptive Increase and Adaptive Decrease (DD-AI/AD)".
- A new PLR (Packet Loss Rate) calculation formula has been implemented.
- It shows the effect of pre-warning of congestion in terms of subjective and objective quality of video in an LTE network.
- It argues that only PSNR is not enough to make a good conclusion about video quality in the case of rate adaptation.

6.3 Future works

The following future work can be done related to this thesis to make the adaptation more effective and robust to the sudden changes in the network.

- Bandwidth monitoring can be done at the receiver to suggest an appropriate bit-rate in TMMBR. This is important as far the Codec Control Message draft is concerned. Besides going down in bit-rate according to the bandwidth monitoring at receiver, it also might help better utilize network resources.
- Currently in this study every end node goes up in bit-rate as soon as a good network condition is detected which might lead to a situation where many of the nodes in the cell tries to go up in bit-rate simultaneously which can make the network congested again within a short period of time. A randomized approach in going up in bit-rate where each node waits a random period of time before it tries to go up in bit-rate should be studied to see the effect on the network.
- It has been seen in the simulation results that the current PLR calculation formula gives different PLR values for same amount of packet loss depending on current network condition. This is the effect of the history window which is currently fixed in size. Here, a single packet loss will give higher PLR value when there is no other packet loss in the network and will give a lower PLR value when there are distributed packet losses observed in the history. This result is not unexpected as we need to address the first packet loss aggressively to reduce further packet loss. It should also be noted that the packet losses are reported in regular RTCP RR where it is normal that the number of packet losses will be reported quite regular. However, a PLR calculation with variable history window size where the size of the window depends on network condition might be able to reduce the difference between the PLR values in the described scenarios.
- In this study no other network traffic is present except video and VoIP traffic for virtual users and only video traffic for the real video client. A further study with other types of traffic in the network should be done to see how adapting a traffic flow which uses higher bandwidth affect the performance of other traffic and on the other hand to see how other traffic

with different QoS and different quality models and bandwidth requirements affect the adaptation procedure studied in this study.

- In the studied unicast session in this thesis the sender is connected to the Internet with a wired connection and the receiver is connected to the network through a radio interface. To see the radio utilization in both uplink and downlink both the sender and the receiver need to be connected with the network through radio connections. A scenario where both the sender and the receiver are connected through a radio connection need to be studied.
- The results in this thesis show that adaptation with ECN improved the quality of video with a big margin but the same may be true for audio and speech transmission. Since this study emphasized video transmission no experiment with audio and speech transmission was done. Future studies needs to include adaptation with ECN for speech and audio to see the full impact on the service quality.

7. References

1. J. Postel. *User datagram protocol.* ftp://ftp.rfc-editor.org/in-notes/rfc768.txt, 1980.
2. Information Sciences Institute. *Transmission control protocol.* ftp://ftp.rfceditor.org/in-notes/rfc793.txt, 1981.
3. H. Schulzrinne et al., *RTP: A Transport Protocol for Real-Time Applications,* http://tools.ietf.org/html/rfc3550, 2003
4. Ramakrishnan K., S. Floyd, D. Black, *The addition Explicit Congestion Notification (ECN) to IP*, http://tools.ietf.org/html/rfc3168 , January 2001.
5. U. Olsson, *Toward the All-IP Vision*, Ericsson review, vol.1, 2005 [online].Available:http://www.ericsson.com/ericsson/corpinfo/publications/review/2005_01/files/200511.pdf.
6. 3rd Generation Partnership Project (3GPP). http://www.3gpp.org/
7. 3GPP, *IP Multimedia Subsystem Stage 2*, 3rd Generation Partnership Project, Tech.Rep. TS 23.228, 2006.
8. S, Chakraborty, J.Peisa, T. Frankkila and P. Synnegren, *IMS Multimedia Telephony over Cellular Systems*. John Willey and Sons, Ltd, England, 2007.
9. 3GPP, *IMS Multimedia Telephony- Media handling and interaction*. TS 26.114, 3rd Generation Partnership Project (3GPP), September 2006.
10. A.Toskala, H. Holma, K. Pajukoski, E. Tiirola, *Utran long term evolution in 3GPP*, In proceedings of 17th Annual IEEE international symposium on Personal, Indoor and Mobile Radio Communication (PRIMRC'06),2006, page(s):1 - 5
11. Wang Hai; Wan Lei; Liao Min, *HSDPA link adaptation based on novel quality model*, In proceedings of Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st Volume 1, Issue , 30 May-1 June 2005, page(s): 334 - 338 Vol. 1
12. 3GPP, *Requirements for Evolved UTRA and UTRAN*, TR 25.913 version 2.1.03rd Generation Partnership Project, June 1-3, 2005, Tdoc RP-050384
13. R. Ludwig, H. Ekström, P. Willars and N. Lundin,*An Evolved 3GPP QoS Concept,* IEEE 1006, available :
14. 3GPP, *Service requirements for the Internet Protocol (IP) multimedia core network subsystem (IMS); Stage 1*, 3rd Generation Partnership Project TS 22.228
15. G. Camarillo, A. Garcia-Martin, *The 3G IP Multimedia Subsytem (IMS)*, 2nd edition, John Wiley & Sons LTD. Sussex, England, May 2006
16. 3GPP, *IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3*, 3rd Generation Partnership Project TS 24.229
17. 3GPP, *Multimedia telephony over IP Multimedia Subsystem (IMS); Optimization opportunities*, 3rd Generation Partnership Project TR 24.229
18. 3GPP, *IMS Multimedia Telephony service; and supplementary services*, 3rd Generation Partnership Project TR 22.973
19. M.Allman et al., *TCP Congestion Control,* http://tools.ietf.org/html/rfc2581, 1999.
20. H. Schulzrinne, S. Casner, *RTP profile for Audio and Video conferences with Minimal Control,* http://tools.ietf.org/html/rfc3551, 2003
21. J. Ott et al., *Extended RTP Profile for Real-Time Transport Control Protocol (RTCP)- Based Feedback,* http://tools.ietf.org/html/rfc4585, 2006

22. S. Wenger, U. Chandra, M. Westerlund, B. Burman, *Codec Control Messages in the RTP Audio-Visual Profile with Feedback (AVPF),* IETF Internet draft, http://tools.ietf.org/html/draft-ietf-avt-avpf-ccm-05, March 2007
23. Ramakrishnan K. and S. Floyd, *A Proposal to add Explicit Congestion Notification (ECN) to IP*, http://tools.ietf.org/html/rfc2481, January 1999.
24. V. Jacobson, *Congestion Avoidance and Control*, in the proceedings of ACM SIGCOMM '88, page(s): 314-329.
25. Braden, B., et al., *Recommendations on Queue Management and Congestion Avoidance in the Internet*, http://tools.ietf.org/html/rfc2309 , April 1998.
26. Postel, J., *Transmission Control Protocol*, STD 7, http://tools.ietf.org/html/rfc793 , September 1981.
27. Floyd, S., *TCP and Explicit Congestion Notification*, ACM Computer Communication Review, V. 24 N. 5, October 1994, page(s): 10-23.
28. Fredrik Hultin, *Congestion notification and rate-adaptation for real-time services in ALL_IP radio networks*, Masters Thesis, Luleå University of Technology.
29. ITU-T, *Advanced video coding for generic audiovisual services,* ITU-T Recommendation H.264 , June 2006
30. Wenger S: *H.264/AVC over IP*, In proceedings of IEEE Transactions on Circuits and Systems for Video Technology, 13, No 7 (July 2003) — http://ieeexplore.ieee.org/iel5/76/27384/01218197.pdf
31. T. Wiegrand, et al. *Overview of the H.264/AVC video coding standard,* In proceedings of IEEE Transaction on Circuits and Systems for Video Telephony, V.13, issue 1, july 2003 pages :560-576
32. J. Sullivan et al., *The H.264/AVC Advanced Video Coding Standartd: Overview and Introduction to the Fidelity Range Extensions,* In the proceedings of SPIE conference on Applications on Digital Image Processing XXVII, August, 2004.
33. ITU-T, *Methodology for the subjective assessment of the quality of television pictures*, ITU-T recommendation BT.500.
34. Reeves, B. & Nass, C. *The media equation*: How people treat computers, television, and new media like real people and places. Cambridge University Press. 1996
35. J Nagle, *Congestion Control in IP/TCP Internetworks*, http://tools.ietf.org/html/rfc896 , January, 1984
36. Lawrence S. Brakmo and Larry L. Peterson. *Tcp vegas: End to end congestion avoidance on a global internet*. IEEE Journal on Selected Areas in Communications, page(s):1465–1480, 1995
37. Saverio Mascolo, Claudio Casetti, Mario Gerla, M. Y. Sanadidi, and Ren Wang. *Tcp westwood: Bandwidth estimation for enhanced transport over wireless links*. In Mobi-Com ’01: Proceedings of the 7th annual international conference on Mobile computing and networking, pages 287–297, New York, NY, USA, 2001. ACM Press
38. Sally Floyd and Kevin Fall. *Promoting the use of end-to-end congestion control in the internet*. IEEE ACM Transactions on Networking, vol 7(4) page(s):458–472, 1999.
39. D. Katabi, M. Handley, and C. Rohrs. *Internet congestion control for future high bandwidth-delay product environments*, SIGCOMM’02*,* August, 2002. available : http://acm.org/sigs/sigcomm/sigcomm2002/papers/xcp.pdf

40. R. Stewart et al. *Stream control transmission protocol*. ftp://ftp.rfc-editor.org/in-notes/rfc2960.txt.
41. E. Kohler, M.handley, S.Floyd. *Datagram Congestion Control Protocol (DCCP)*, http://www.ietf.org/rfc/rfc4340.txt, March 2006.
42. J. Kurose, O*pen issues and challenges in providing quality of service guarantees in high-speed networks*, ACM computer Communication Review 23 (1) page(s) :6-15, 1993.
43. David D. Clark, Scott Shenker, and Lixia Zhang. *Supporting real-time applications in an integrated services packet network: Architecture and mechanism*. In SIGCOMM, page(s) 14–26, 1992.
44. Chip Elliot. *High-quality multimedia conferencing through a long-haul packet network*. In MULTIMEDIA '93: Proceedings of the first ACM international conference on Multimedia, page(s) 91–98, New York, NY, USA, 1993.
45. K. Kim, K. Nahrstedt, *A resource Broker Model with Integrated Reservation Scheme*, In proceedings of IEEE International Conference on Multimedia and Expo 2000 (ICME 2000), August 2000, Newyork , page (s) 859-862
46. D. Ivan –Rosu et al. *On Adaptive resource allocation for Complex real time application*, Real- time systems symposium, December 1997, IEEE, page(s) 249-262.
47. B. Li et al. *End-to-End QoS support for Adaptive applications over the Internet,* SPIE International Symposium of Voice, Video and Data Communication, November 1998, page(s) 147-161
48. G. Bianchi, A. Capone, and C. Petrioli, *Throughput analysis of end-to-end measurement-based admission control in IP*, in Proc. IEEE INFOCOM, vol. 3, Mar. 2000, page(s) 1461–1470.
49. R. Rejaie, M. Handley, and D. Estrin, *RAP: An end-to-end rate-based congestion control mechanism for realtime streams in the internet*, in *Proc. IEEE INFOCOM*, vol. 3, Mar. 1999, page(s). 1337–1345.
50. S. floyd et al. *Equation-Based Congestion Control for Unicast Applications*, SIGCOMM2000, May 2000
51. Fan Yang, Qian Zhang et al., *End-to-End TCP-Friendly Streaming Protocol and Bit Allocation for Scalable Video Over Wireless Internet*, IEEE journal on selected areas in communications, vol. 22, no. 4, may 2004, page(s) 777-790
52. Minghua Chen Zakhor, A. ,*AIO-TFRC: a light-weight rate control scheme for streaming over wireless,*in proceedings of International Conference on Wireless Networks, Communications and Mobile Computing, 2005, vol. 2, June 2005 page(s) 1124 – 1129.
53. Qi Li Yuncai Liu L. Zheng Lina Zheng,*Jitter Ratio Based TFRC Scheme In Wireless-Wired Hybrid Network,* in the proceedings of International Conference on Digital Telecommunications, 2006. ICDT '06, page(s): 38 - 38
54. Deepak Bansal and Hari Balakrishnan, *Binomial Congestion Control Algorithms*, IEEE INFOCOM 2001
55. D. Chiu and R. Jain, *Analysis of the increase and decrease algorithm for congestion avoidance in computer networks*, *Comput. Netw. ISDN Syst. J.*, vol. 17, no. 1, page(s). 1–14, Jun. 1989.
56. Zhiheng Wang, Sujata Banerjee, and Sugih Jamin. *Media-friendliness of a slowlyresponsive congestion control protocol*. In Proceedings of the 14th

international workshop on Network and operating systems support for digital audio and video NOSSDAV '04, Cork, Ireland 2004, page(s) 82–87.

57. Kang-Won Lee; Tae-eun Kim; Bharghavan, V., *A comparison of two popular end-to-end congestion controlalgorithms: the case of AIMD and AIPD*, Global Telecommunications Conference, 2001. GLOBECOM IEEE Volume 3, 2001 page(s):1580 – 1584
58. Tae-eun Kim; Songwu Lu; Bharghavan, V. *improving congestion control performance through loss differentiation,* in Proceedings of Eight International Conference on Computer Communications and Networks, 1999. page(s):412 – 418
59. Minghua Chen; Zakhor, A. *Rate control for streaming video over wireless,* INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies Volume 2, Issue , 7-11 March 2004 page(s): 1181 - 1190
60. Pantelis Balaouras and Ioannis Stavrakakis, *A multi-state Congestion Control Scheme for Rate Adaptable unicast Continuous Media Flows*, Lecture notes in commuter science, Springer Berlin Volume 2170/2001, February 2004.
61. Pantelis Balaouras and Ioannis Stavrakakis, *A congestion Control Scheme for Continuous Media Streaming Applications.* Lecture notes in commuter science, Springer Berlin Volume 2511/2002, February 2004.
62. H. Kanakia, P. Mishra, A. Reibman, *An Adaptive Congestion Control Scheme for Real time Packet Video Transport*, in SIGCOMM93, Ithaca , NY, USA, September 1993. page(s) 20-31
63. Schierl, T.; Wiegand, T.; Kampmann, M *3GPP compliant adaptive wireless video streaming using H.264/AVC.* In the proceedings of IEEE International Conference on Image Processing, 2005. ICIP 2005.Volume 3, 11-14 Sept. 2005 page(s): 696-699
64. Cooperative Association for Internet Data Analysis Path and Round-Trip Time Measurements (2004, Apr.). [Online]. Available: http://www.caida.org/outreach/presentations/IEPG.9808/
65. Seong-jun Bae; Song Chong, *TCP-friendly wireless multimedia flow control using ECN marking,* Global Telecommunications Conference, 2002. IEEE GLOBECOM Volume 2, 17-21 Nov. 2002 page(s): 1794 – 1799
66. Chaudhary, R.; Jacob, L, *ECN based TCP-friendly rate control for wireless multimedia streaming,* in the proceedings The 12th International Conference on Computer Communications and Networks, ICCCN 2003, 20-22 Oct. 2003 page(s): 599 – 602
67. O. Karimi, M. Fathy, S. yousefi, *Application level Wireless Multi-level ECN for Video and Real-time Data,* In the Proceedings of the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies (ICNICONSMCL'06), 2006
68. Dracinschi, A.; Fdida, S, *Efficient congestion avoidance mechanism*. In the proceedings. 25th Annual IEEE Conference on Local Computer Networks, 2000. LCN 2000,page(s):228 - 237
69. Gharavi, H. Ban, K. Cambiotis, J., *RTCP-based frame-synchronized feedback control for IP-video communications over multipath fading channels,*in the proceedings of IEEE International Conference on Communications, 2004 June 2004, Volume: 3, page(s): 1512- 1516
70. Baldo, N.; Horn, U.; Kampmann, M.; Hartung, F., *RTCP feedback based transmission rate control for 3G wireless multimedia streaming*, in the

proceedings of 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004. Volume 3, 5-8 Sept. 2004 page(s): 1817 – 1821.

71. P Mulroy, *Application layer QoS for video telephony,* BT Technological Journal, vol. 24 no 2, April 2006.
72. N. Cranley, L. Murphy, P. Perry. *User perceived Quality Aware adaptive delivery of MPEG-4 content,* in the proceedings of *NOSSDAV'03*, June 1-3, 2003, Monterey, California, USA. page(s): 42-49
73. M. Talaat et al. *Content-aware Adaptive Video Streaming system,* in the proceedings of 3rd International Conference on Information and Communications Technology, 2005. Enabling Technologies for the New Knowledge Society: ITI , December 2005, page(s): 265-276
74. www.netchico.com/support/glossary/c.html date : 2007-08-28
75. D.-M. Chiu and R. Jain, *Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks*, *Journal of Computer Networks and ISDN Systems*, 1989.