Enhancement of LTE Radio Access Protocols for Efficient Video Streaming

by

Balaaji Tirouvengadam

A thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements for the degree of

MASTER OF APPLIED SCIENCE

in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Electrical Engineering and Computer Science

University of Ottawa
Ottawa, Ontario K1N 6N5, Canada
August 2012

© Balaaji Tirouvengadam, Ottawa, Canada, 2012

Abstract

A drastic increase in traffic of mobile broadband is seen in the past few years, which is further accelerated by the increase in usage of smart phones and its applications. The availability of good smart phones and better data connectivity are encouraging mobile users to use video services. This huge increase in usage will pose a lot of challenges to the wireless networks. The wireless network has to become content aware in order to offer enhanced quality of video service through efficient utilization of the wireless spectrum. This thesis focuses on improving the Quality of Experience (QoE) for video transmission over Long Term Evolution (LTE) networks by imparting the content awareness to the system and providing unequal error protection for critical video packets. Two different schemes for the improvement of video quality delivery over LTE networks are presented in this thesis. Using content awareness, the retransmission count of Hybrid Automatic Repeat reQuest (HARQ) are changed dynamically such that the most important video frame gets more number of retransmission attempts, which increases its success for delivery in-turn increasing the received video quality. Since Radio Link Control (RLC) is the link layer for radio interface, the second approach focuses on optimizing this layer for efficient video transmission. As part of this scheme, a new operation mode called Hybrid Mode (HM) for RLC is defined. This mode performs retransmission only for the critical video frames, leaving other frames to unacknowledged transmission. The simulation results of both proposed schemes provide significant improvement in achieving good video quality without affecting the system performance.

Acknowledgements

I take this opportunity to thank my thesis supervisor, Dr. Amiya Nayak, for his valuable guidance which led me in right direction for successful completion of this work. I am grateful to him for providing me the complete freedom of thought and constant motivation. I thank OPNET Inc. for providing the research license to carry out simulations needed for this thesis.

I have been blessed with good friends, whom I can depend on at any time. I would like to thank Rakesh Radhakrishnan for encouraging me to fulfill my dream of higher studies and for his invaluable suggestions and discussions that shaped me both in personal and professional front. I thank Kobi Sundar Iyakutti for his constant support and encouragement. I am glad to thank Arunkumar Pandian, who has been my inspiration all these years from my undergraduate studies.

I am extremely lucky to have wonderful parents (Tirouvengadam and Sumathy) and a brother (Suresh), who always stood by me all throughout my life. Pursuing graduate studies would have not been possible without their support and love. I would like to dedicate this thesis to my family.

Table of Contents

List of Figures

List of Tables

Acronyms

2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
ACK	Acknowledgement
AM	Acknowledged Mode
AMC	Adaptive Modulation and Coding
AMD	AM Data
AVC	Advanced Video Coding
B frame	Bidirectional predicted frame
CAAHR	Content Aware Adaptive HARQ Retransmission
CIF	Common Intermediate Format
CN	Core Network
CPT	Control PDU Type
CQI	Channel Quality Indicator
dBm	Decibel (referenced to milliwatts)
dBi	Decibel (isotropic)
DSCP	Differentiated Services Code Point
eNodeB	evolved NodeB
EPC	Evolved Packet Core
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
FDD	Frequency Division Duplex
GOP	Group of Pictures
GPRS	General Packet Radio Service
GTP	GPRS Tunneling Protocol
GUI	Graphical User Interface

GW	GateWay
HARQ	Hybrid Automatic Repeat reQuest
HM	Hybrid Mode
HSDPA	High Speed Downlink Packet Access
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
I frame	Intra coded frame
IMT	International Mobile Telecommunications
IND	Indication control PDU (HM RLC)
IP	Internet Protocol
IPv4	IP version 4
ITU	International Telecommunication Union
ITU – T	ITU - Telecommunication standardization sector
LTE	Long Term Evolution
LTE-A	LTE Advanced
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
MOS	Mean Opinion Score
MPEG	Motion Picture Experts Group
MTU	Maximum Transferable Unit
NACK	Negative Acknowledgement
NAL	Network Abstraction Layer
OFDMA	Orthogonal Frequency Division Multiple Access
OSI	Open Systems Interconnection
PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PDU	Payload Data Unit
P frame	Predictive coded frame

PHICH	Physical HARQ Indication Channel
PRB	Physical Resource Block
PSNR	Peak Signal to Noise Ratio
QAM	Quadrature Amplitude Modulation
QoE	Quality of Experience
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAN	Radio Access Network
RB	Resource Block
RLC	Radio Link Control
RRM	Radio Resource Management
SAW	Stop And Wait protocol
SC-FDMA	Single Carrier Frequency Division Multiple Access
SDU	Service Data Unit
SE	Spectral Efficiency
SN	Sequence Number
SNR	Signal to Noise Ratio
SVC	Scalable Video Coding
TCP	Transmission Control Protocol
TDD	Time Division Duplex
ToS	Type of Service
UDP	User Datagram Protocol
UE	User Equipment
UM	Unacknowledged Mode
UMTS	Universal Mobile Telecommunication System
WiMAX	Worldwide Interoperability for Microwave Access

Chapter 1

Introduction

The mobile telecommunication industry has witnessed a remarkable change in the last decade. The early generations of mobile communication targeted to provide only a good quality voice communication, while the recent 4th generation (4G) wireless technology aims at providing a high quality video as well as voice communication that can coexist. The availability of smartphones and tablets built with latest communication technology at an affordable price along with a good coverage of wireless network in the recent times has in-turn fuelled mobile broadband usage beyond one's imagination. There is a constant increase in the number of 4G users due to its ability to offer reliable and high speed wireless connectivity. It has been predicted that 6% of total world mobile connection will be of 4G users by 2016. However, the traffic generated by them will contribute to approximately 36% of the total traffic [1]. Long Term Evolution (LTE) and Worldwide Interoperability for Microwave Access (WiMAX) technologies are potential candidates for the upcoming 4G networks. Since LTE has evolved from existing 3G systems, it has been widely accepted by many service providers for 4G deployment. LTE provides various benefits in achieving the high speed data connectivity irrespective of user mobility. Many operators around the world have already started deploying LTE networks and are seeing steady increase of users day by day. It was predicted that the total number of LTE users will increase by 400% by the end of 2012 [27].

1.1 Motivation

The existing wired networks have been considered as best effort networks wherein the delay and jitter of a packet cannot be guaranteed. The wired networks have evolved due to the demanding growth of real-time applications such as video conferencing and voice over Internet Protocol (IP). As a part of this evolution, each packet in the network is identified with a Type of Service (ToS) flag, which will be used by the network elements along with the diff-server to achieve the Quality of Service (QoS) defined for the service. This has enabled the wired networks to be aware of the application service to which the packet belongs and perform an efficient network operation. This awareness in the network has made the deployment of real-time applications over the wired networks possible. The service level differentiation used in wired networks to provide the guaranteed QoS have also been adapted in wireless networks. However, content awareness within a given application traffic is still lacking in both wired and wireless networks. Each application will have application critical packets and its loss will deteriorate the user experience considerably. A major share of growing mobile broadband traffic is consumed by video related services. According to a survey done by Cisco in 2011, mobile video traffic has exceeded 50% of total global data traffic [1]. Hence, providing a reliable video service over wireless medium will attract more users, thereby increasing the total revenue generated by mobile networks. The current trend of increase in traffic over unreliable wireless networks, demands them to be content aware for an efficient utilization of physical resources without affecting the user experience. Thus, in addition to service level awareness, packet content awareness in the wireless networks is essential to offer a good Quality of Experience (QoE).

Video transmission can be considered as transmitting a sequence of images with temporal relations. This requires large bandwidth and hence, achieving reliable video transmission over bandwidth constrained wireless media is challenging. An effective codec is the key for efficient transmission especially over wireless channel. The 3GPP standard has recommended the use of H.264 also known as Advance Video Coding (AVC) for video services of LTE network [30, 43]. In video transmission, every frame is not equally important with respect to the decoding process. In a group of frames there will be a sub-set of frames that has the critical information for successful decoding of other frames in the group. These frames are referred to as critical frames, losing these frames can create more disruption in video quality than losing the non-critical frames. In order to deliver a better video quality, the network has to be content aware of the packet being transmitted to provide unequal error protection. This error protection will result in the improvement of quality perceived by the users without compromising much on the channel bandwidth. In this thesis, we address the following two problems:

Problem 1: Need for a simple adaptive retransmission scheme for HARQ

HARQ provides a better error correction mechanism in LTE which reduced the round trip delay involved in legacy ARQ. This is achieved by retransmitting the lost packet with suitable Forward Error Correction (FEC) codes at the wireless Medium Access Control (MAC) level. Though HARQ preforms an efficient retransmission, the operation is not content aware of the packet being retransmitted, creating a bottleneck in achieving good video quality. Hence, there is a need for a simple adaptive retransmission scheme for HARQ based on the packet content for video transmission.

Problem 2: Need for an optimized mode of operation for RLC

RLC is the link layer protocol for the LTE wireless interface. The existing operation modes of RLC are designed to support either a complete reliable slow channel or unreliable fast channel. However, as we mentioned earlier for efficient video transmission the demand for unequal error protection is inevitable. Hence, we need an optimal mode of operation for RLC layer which can provide fast transmission for non-critical packet and reliable transmission for application critical packets.

1.2 Objectives and Methodology

The thesis focuses on finding the best possible solutions for the problems identified in the previous section. The objectives of the thesis can be summarized as follows:

a) To provide a simple unequal error protection scheme for the video packets transmitted at the MAC level of LTE network without affecting the performance of the system. This can be achieved by varying the maximum retransmission count at the transmitting HARQ entity based on the importance of the video packet being transmitted. This can be implemented by the changing the maximum retransmission count value dynamically based on the video frame type.

b) To design a better mode of operation for the link layer of wireless access interface which is more suited for efficient video transmission thereby, enhancing the perceived video quality by the users of LTE network. The current modes of RLC support either acknowledged or unacknowledged transmission for all the packets of a channel. The acknowledgement for every transmitted packet causes a huge delay in transmission. However, in video transmission, acknowledgement is not needed for all transmitted packets. A better QoE can be achieved by providing reliable delivery only for the critical video packets. This can be achieved by ignoring the loss of non-critical packets and requesting retransmission only for the critical packets at the receiving RLC entity. The Information about critical packets can be shared between transmitter and receiver through control messages.

A systematic approach has been followed to achieve the above mentioned objectives. To study and validate the proposed schemes OPNET Modeler [26] is chosen as the simulation tool by considering the factors such as support for LTE system level simulation and wide range of configurable parameter which makes the simulation more flexible. This tool also supports simulation of video traffic based on trace files. Hence, the AVC trace file of Star Wars IV which is available online [12], is used for simulating the video traffic. The characteristics of the video can be seen in Table 3.2 and Appendix A. The OPNET Modeler code is tweaked to incorporate the changes required as per the proposed schemes and compiled. The software modules of downlink HARQ and AM RLC are modified to test the proposed schemes. As these schemes are based on unicast, single cell LTE network with a video server is used as the primary network setup. A video conference profile is configured with the trace file as input and the results of LTE system performance parameters are compared and analyzed in OPNET. However, the received video quality cannot be validated through it. Hence, a program written in MATLAB [23] is used for processing the received video trace to compute the received video quality. The accuracy of obtained results is validated by running the simulation for multiple times.

1.3 Assumptions

Some of the assumptions made during the simulation are discussed in this section. The network setup is configured in such a way that the wired links are considered to be error free and congestion free with infinite bandwidth. The mobility of wireless user terminals are not considered, hence their positions are fixed throughout the simulation duration. Since, the users are not mobile, handover scenarios are not considered. All the network elements are considered to have adequate power for processing. The configuration overheads involved in the proposed schemes are not considered. The simulation is performed only with video traffic, mixed traffic is not considered. It is assumed that the application level FEC codes for the transmitted video frames are not used. Hence, any corrupted video frame is considered to be lost as error recovery is not possible. The overhead involved between application and IP layer at video server for exchange of video frame information is not considered. All the configuration parameters of OPNET (except few mentioned in this thesis) are assigned with default values.

1.4 Contributions

The main focus of the work mentioned in this thesis is on enhancing the existing layer 2 protocols of wireless interface by introducing content awareness and providing more error protection for critical video packets to improve the received video quality. Solutions based on wireless interface protocols have been discussed in this thesis that provides unequal error protection for the video packets transmitted over LTE network. The proposed solutions and their results obtained are summarized as follows:

a) A content aware adaptive HARQ retransmission scheme has been proposed to increase the received video quality by the user. This is achieved by varying the retransmission count according to the importance of the video packet. More retransmission count is given only for the critical video frames to increase its success rate. For users under poor channel condition, the proposed scheme improves the calculated received video quality by 8%. The main advantage of this scheme is that these benefits come with negligible increase in delay and no extra processing needed at the User Equipment (UE). This work has been published in IEEE ICUFN 2012 conference [39].

b) A new mode of operation for RLC has been proposed which combines the benefits of other existing modes of the protocol. The proposed mode is called as Hybrid Mode (HM). This mode along with the cross layer interaction with video server identifies the critical video packets and provides Backward Error Correction (BEC) only to those packets by retransmitting them. The non-critical packets are transmitted without the retransmission support. This scheme provides an improvement in the calculated received video quality by 7% in comparison with RLC Unacknowledged Mode (UM). It also reduces the end-to-end delay by 99% in comparison with the Acknowledged Mode (AM) mode. Moreover, this scheme is compatible with the existing system. This work has been accepted in IEEE ICPADS 2012 conference [38].

1.5 Thesis Organization

The thesis is organized as follows. A brief overview of LTE system, its architecture, description on concepts of channels, wireless layer 2 protocols, an introduction to basics of H.264 standard and Internet protocol are discussed in Chapter 2. A new adaptive retransmission scheme has been proposed and validated for HARQ layer of LTE in Chapter 3. In Chapter 4, a new operation mode for wireless layer 2 protocol (RLC) has been suggested, and its performance and benefits are also discussed. The conclusion and possible directions for future work are discussed in Chapter 5.

Chapter 2

Background and Related Work

2.1 LTE: An Introduction

LTE is a 4G mobile broadband system whose wireless access technology is based on Frequency Division Multiple Access (FDMA). LTE network has evolved from its predecessor 3G Universal Mobile Telecommunications System (UMTS) with a major change in its wireless access technology. LTE deploys two separate access techniques for downlink and uplink transmission. It uses Orthogonal FDMA (OFDMA) in downlink and Single Carrier FDMA (SC-FDMA) in uplink [5]. The SC-FDMA is used in uplink to overcome the peak-to-average ratio problem of OFDMA thereby simplifying the transmitter part of UE [41]. Since, the eNodeB is a fixed entity with a constant power supply, the above said drawback of OFDMA has been overlooked by considering its various benefits for downlink transmission. Similar to UMTS, in LTE the uplink and downlink channel duplexing can be performed either by Time Division Duplex (TDD) or Frequency Division Duplex (FDD) operation modes. In this work, only FDD LTE is considered, as it is the mode of operation chosen to be deployed by majority of telecom service providers.

International Mobile Telecommunication-Advanced (IMT-Advanced) had put forth some requirements for 4G technologies. The initial version of LTE has met most of the requirements mentioned by IMT-Advanced. The later version of LTE called LTE-Advanced have exceeded all the requirements and qualified as a true 4G technology. Some of the requirements considered during LTE standardization phase are listed below [10]:

- A peak data rate of 100Mbps in downlink and 50Mbps in uplink
- Significant improvement in Spectral Efficiency (SE) (2-4 times of Release 6 UMTS SE)

- Reduce Radio-access network latency below 10ms
- Significant reduction in Control plane latency (e.g. state transition in less than 100ms)
- Scalable bandwidth – 1.25, 1.6, 2.5 MHz to support narrow spectral allocation and 5, 10, 15, 20MHz to support broader spectrum
- Should be backward compatible with existing 3G network
- Should be optimized for low mobile speed (less than 15km/h) at the same time it should support high mobile speed (up to 350km/h) as well

Unlike the traditional 3G system, LTE supports wide range of channel bandwidth starting from 1.4 MHz to 20 MHz which can be selected based on the deployment requirements. The wireless physical resource of LTE is defined in frequency as well as time domain. This gives a greater degree of freedom for the resource scheduler to schedule the limited wireless resource to the users with better channel condition in a most appropriate way. As per 3GPP, a downlink Physical Resource Block (PRB) in FDD corresponds to a time slot of 0.5ms with 12 subcarriers each with a bandwidth of 15 kHz (total bandwidth of 180 kHz). A depiction of PRB in time and frequency domain is shown in Figure 2.2. Hence, a LTE channel of 20MHz bandwidth can have a maximum of 110 PRBs [5]. However, in practice considering guard band, a LTE channel of 20MHz will have a maximum of 100 usable PRBs. Though a resource block is defined in slots of 0.5ms, the scheduler assigns a resource for an UE in terms of 1ms duration. The radio frame of LTE is of 10ms duration, where the Transmission Time Interval (TTI) is defined as 1ms. Two types of radio frame formats have been defined namely Type 1 and Type 2, where Type 1 is used for FDD mode and Type 2 for TDD. The Type 1 radio frame is divided into 10 sub-frames with each sub-frame of two slot duration, where a slot is of 0.5ms duration (see Figure 2.1).

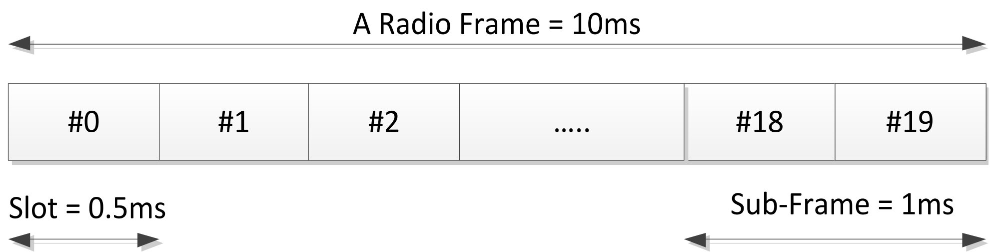


Figure 2.1 LTE radio frame [32]

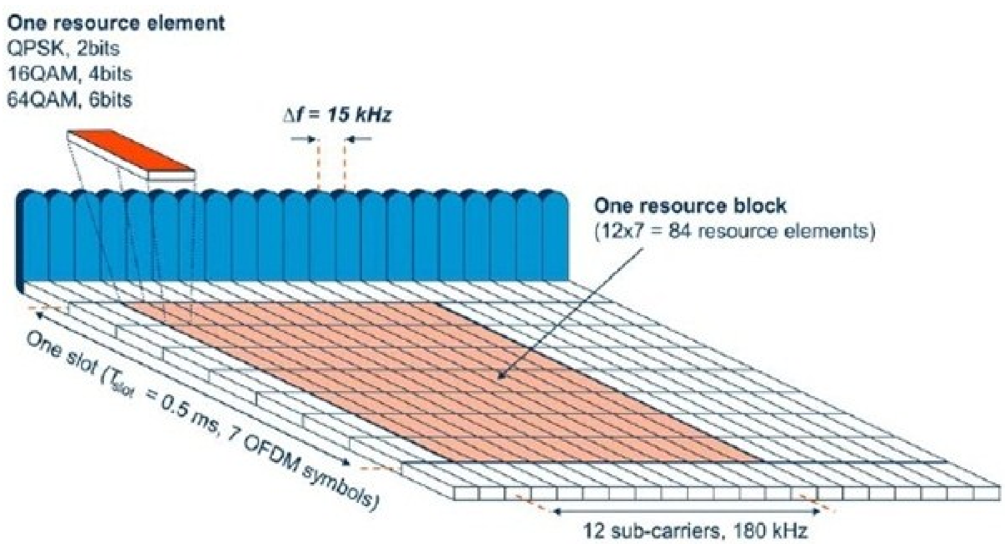


Figure 2.2 Representation of a PRB in LTE [3]

LTE supports the following three modulation schemes:

- Quadrature Phase Shift Keying (QPSK)
- 16 Quadrature Amplitude Modulation (QAM)
- 64 QAM

QPSK being the lowest modulation provides a robust channel, while 64 QAM provides a high modulation rate and can be used only when there is a good channel quality. The error correction is deployed in two steps as FEC and BEC. Coding, interleaving and puncturing operations are performed as part of FEC and rate matching, whereas BEC is deployed by retransmission at the higher layers. Proper modulation and coding methodologies are chosen by the scheduler for a transmission based on the channel quality estimation. The channel quality is estimated with a help of predetermined pattern of pilot channel and indicated as a measurement index called Channel Quality Index (CQI). This CQI will be used by the scheduler to determine the Modulation and Coding Scheme (MCS) to be employed for the transmission. Apart from robust OFDM and higher order modulation scheme, the peak user data rate is further increased by exploiting the concept of spatial multiplexing. This is deployed with the help of Multiple Input and Multiple Output (MIMO) technology. This technology uses array of transmitting and receiving antennas which are spatially separated. The signals received by these antennas are processed to improve the signal strength. 3GPP Release 8 has recommended the use of 4 transmitting and 4 receiving antennas (4x4) in downlink and 2x2 in uplink. In later release this has been extended to support 8x8 in downlink and 4x4 in uplink.

2.2 LTE: Network Architecture

The LTE system from a network perspective is based on flat IP architecture, where all the network entities are connected thorough IP. The network can be split into two parts namely, Radio Access Network (RAN) and Core Network (CN). The evolved NodeB (eNodeB) which serves the UE forms the RAN whereas Evolved Packet Core (EPC) represents the CN part of the network. The Figure 2.3 shows the overall network architecture and the interfaces through which the network elements are connected. These network elements have their role either in signaling traffic (control plane), user data traffic (user plane) or both. The different entities of LTE network and their primary functionalities are listed below:

- **User Equipment** (UE) represents the mobile terminal of the system. LTE supports a wide variety of terminal ranging from handheld compact smart phones till bulky laptops.
- **Evolved NodeB** (eNodeB) is responsible for major radio related operations of the system. This has evolved from the traditional NodeB of 3G system which performed minimal functionality of Radio Network Controller (RNC). However, in LTE the functionality of RNC has been decentralized to many eNodeBs.

- **Mobility Management Entity** (MME) is the key control node for LTE access network whose primary functions include tracking user location, paging procedures, activation / deactivation of bearer channels and inter-network handover [47].
- **Serving Gateway** (S-GW) acts as a gate way for user plane traffic. It also acts as a mobility anchor for the users moving from one eNodeB to another or between 2G/3G systems. This entity is responsible for lawful interception.
- **Packet Data Network (PDN) Gateway** (P-GW) is the entry point for external IP network to the LTE network. This entity is responsible for policy enforcement and charging support. P-GW also acts as an anchor during non-3GPP inter-network mobility.
- **Home Subscriber Support** (HSS) is the central data base server for the LTE system. This server contains all identity and subscription related information for the home users. This entity replaces the Home Location Register (HLR) and Authentication Center (AuC) of 3G system.
- **Policy Charging and Rules Function** (PCRF) is a logical entity which is responsible for decision making based on the policies and rules defined. It is also responsible for enforcing them in collaboration with P-GW.

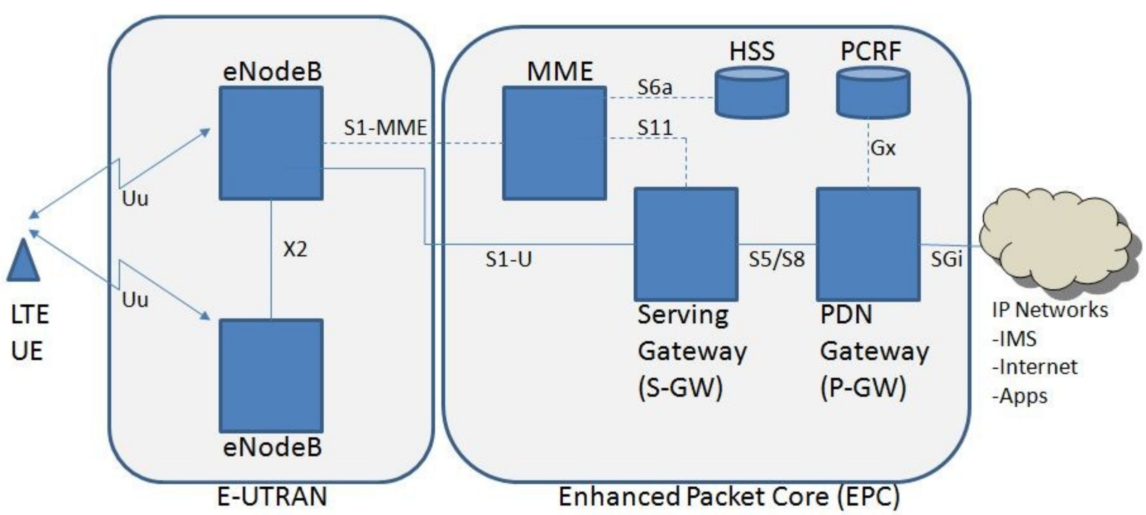


Figure 2.3 Overall network architecture of LTE [32]

The wireless interface between eNodeB and UE is named as U_U. It is the most important and the only connection for the mobile equipment with the network. The interaction between any two eNodeB is carried over the X2 interface. This interface is used during the inter eNodeB handover scenarios. A group of eNodeB is called as Evolved Universal Terrestrial Radio Access Network (E-UTRAN). The interfaces between E-UTRAN and EPC are called as S1 interface. This interface is further logically split into user plane and control plane as S1-U and S1-MME respectively. The S-GW and P-GW is connected through S5/S8 interface. MME is connected to S-GW and HSS through S11 and S6a respectively. As mentioned earlier the user plane traffic gets routed through the eNodeB, S-GW and terminates at the end point based on the type of service requested. However, the signals from control plane can either be interpreted in eNodeB or at MME. The signaling messages that terminate at eNodeB are grouped as Access Stratum (AS) and those terminate at MME are grouped as Non-Access Stratum (NAS) signals.

Figure 2.4 shows the basic LTE network pertaining to user plane traffic along with the protocol stack used at respective interfaces. Packet Data Convergence Protocol (PDCP), RLC) and MAC forms the layer 2 protocol stack of the wireless interface U_U. The tunnels between eNodeB, S-GW and P-GW are established using GPRS Tunneling Protocol User plane (GTP-U) with User Datagram Protocol (UDP) as its transport. The GTP-U acts as a transport layer for the packets transferred between eNodeB and the P-GW. The main functionality of PDCP is IP header compression and context maintenance for successful handover. RLC preforms segmentation, reassembly, reordering and flow control operations. The MAC layer is responsible for scheduling and access control of the wireless medium.

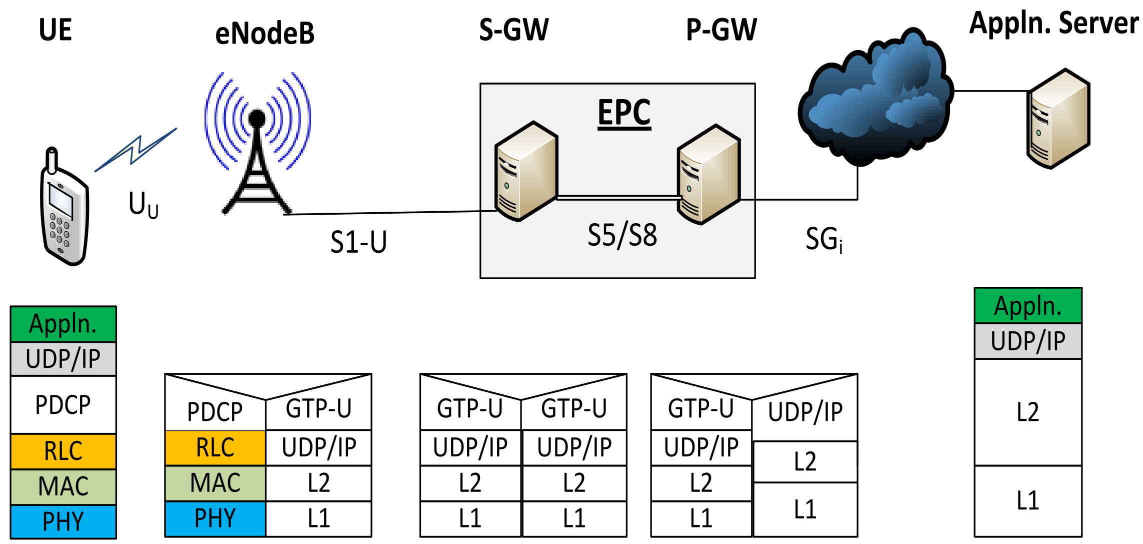


Figure 2.4 LTE user plane network architecture [32]

2.3 LTE: Channels

There are three categories of channels defined in LTE for traffic management and QoS implementation, namely physical, transport and logical channels. The physical channel specifies the modulation and coding techniques used and the way how the bits are transmitted over air. The transport channel provides the information on the format and the size of the frame being transmitted. The logical layer is for the use of upper layer to differentiate the packets based on their service types. Both physical and transport channels are unidirectional while most of the logical channels are bidirectional. Some of the important channels related to unicast are listed below:

Physical channels:

- Physical Downlink Shared Channel (PDSCH)
- Physical Downlink Control Channel (PDCCH)
- Physical Control Format Indicator Channel (PCFICH)
- Physical Hybrid ARQ Indicator Channel (PHICH)
- Physical Random Access Channel (PRACH)
- Physical Uplink Shared Channel (PUSCH)

- Physical Uplink Control Channel (PUCCH)

Transport Channels:

- Downlink Shared Channel (DSCH)
- Uplink Shared Channel (USCH)
- Radom Access Channel (RACH)

Logical Channels:

- Common Control Channel (CCCH)
- Dedicated Control Channel (DCCH)
- Dedicated Traffic Channel (DTCH)

The channels listed above are restricted to unicast transmission, because multicast, broadcast and synchronization procedures are beyond the scope of the work mentioned in this thesis. The PDSCH is a shared channel through which the downlink data transfer is carried out. The information on the users being scheduled and the type of modulation applied are communicated through a control channel called PDCCH. Similar to downlink, the uplink data transfer is carried through PUSCH. PHICH is used for positive or negative acknowledgements of the transmitted downlink data frames.

The relations between these channels are depicted in Figure 2.5. The mapping between the transport and the physical channels are well defined in [40]. The mapping between transport and logical channels may vary according to the situation (based on user idle or active state and the direction of communication). This mapping is done by MAC layer of RA. The RLC layer maps the logical channels with their corresponding Radio Bearers (RB).

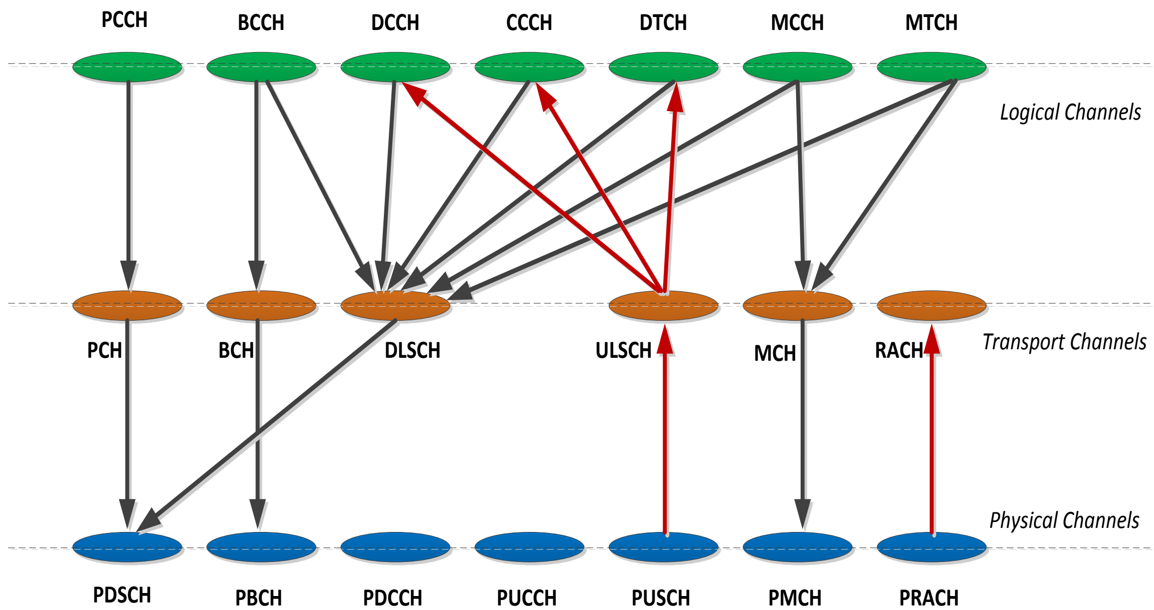


Figure 2.5 LTE channel mapping [32]

2.4 LTE: Layer 2 Protocols

As mentioned earlier, the MAC, RLC, and PDCP layers of U_U forms the second layer of radio interface. Since the work mentioned in this thesis is based on Hybrid Automatic Repeat reQuest (HARQ) and RLC, only the functionalities related to those layers are discussed in this section.

2.4.1 Hybrid ARQ

Wireless mobile networks (3G and beyond) introduced Automatic Repeat reQuest (ARQ) as a retransmission method in layer 2 RLC protocol to ensure reliability in packet transmission. To further reduce the round trip delay involved in the retransmission, a modified version of ARQ called HARQ, was introduced as a MAC functionality in UMTS Rel5 [8]. Retransmission is a type of backward error correction, whereas FEC is achieved by adding few redundant bits so that the transmitted packet has an error tolerance level. HARQ combines error detection and FEC and uses them effectively according to the channel condition. The benefit of FEC is that it reduces the number of retransmissions, thereby reducing the time and power involved in those transmissions. On the negative side, FEC comes with a cost of throughput reduction; the amount of useful data transmitted is reduced due to the addition of redundant bits. So, adding FEC in a good channel will reduce the efficiency of transmission, and at the same time it has its benefits in bad channel conditions. HARQ takes advantage of deploying the desired FEC level based on the channel condition and avoids whenever it is possible.

HARQ implements Stop And Wait (SAW) protocol [15] and allows 'n' parallel transmissions, where 'n' refers to the number of HARQ processes available. A representation of SAW operation is shown in Figure 2.6. LTE FDD supports a maximum of 8 HARQ processes [6]. Since HARQ supports parallel transmission, it requires reordering of packets before delivering it to the upper layer. However, in LTE this reordering function is performed at RLC layer. A maximum retransmission threshold is defined during channel configuration. Once the threshold is reached, the HARQ entity stops transmitting the packet and discards it, leaving it to the upper layer (RLC/application depending on the configuration) to take necessary action for error recovery. Based on the type of retransmission performed, the operation of HARQ can be classified as Type I, Type II or Type III. In Type I, the erroneous packet is discarded and all transmitted and retransmitted packets have same content. In case of Type I, a good channel condition is expected during the retransmission process to make a successful transmission. As the channel behavior is random, the error produced at one time instance may not be repeated. Hence, the same packet transmitted at different time instance can have different error patterns. A successful construction of error free packet is possible by combining the erroneous packets received at different time instances. This method is called soft combining [17] and it is used in Type II and Type III HARQ processes. The soft combining used in HARQ process can further be classified as two types, namely:

- ***Chase Combining (CC)***: The retransmission packets have the same content (data and redundancy bits). The receiver applies the maximum-ratio combining method to construct the error free packet.
- ***Incremental Redundancy (IR)***: Each re-transmitted packet differs in its content. Every retransmission is associated with a different set of error correcting codes. The receiver has to relate the previous received contents to construct an error free

packet. If the retransmitted packet contains only the redundancy information, it uses Type II HARQ method. In Type III, the retransmitted packet contains data along with a version of error correcting code. In both types, to identify the type of redundant bits transmitted, a redundancy version is attached with the transmission.

Though IR requires more buffer size than CC, it has better performance [17].

The acknowledgement procedures of HARQ can be synchronous or asynchronous, while the redundancy version used during the retransmission in the case of IR can be a predetermined sequence or it can be adaptive. 3GPP has defined certain characteristics of the HARQ process for uplink and downlink transmission [9], which are as follows:

Downlink:

- ACK/NACK sent by UE to eNodeB using either PUCCH or PUSCH
- Uses asynchronous adaptive HARQ
- Retransmission are scheduled through PDCCH

Uplink:

- ACK/NACK sent by eNodeB to UE through PHICH
- Uses synchronous HARQ

2.4.2 Radio Link Control

RLC is one of the layer 2 protocols of the RA protocol stack of LTE system whose end points are located at UE and eNodeB. This layer is sandwiched between MAC and PDCP layer (see Figure 2.4). This layer acts as a link layer for U_U interface carrying both control and data traffic. This layer is also responsible for the flow control of data transmitted over the wireless channel. In LTE, the ciphering functionality of RLC is moved to PDCP layer. The mode of operation and various parameters governing the operations of this layer are configured by Radio Resource Control (RRC), a resource control protocol of RA. RRC is responsible for establishment, maintenance and teardown of bearer links. The operation parameters of RLC are determined based on the type of service to be serviced in the created link [32]. The operations of RLC fall under ARQ category, which performs BEC by retransmitting the lost packets [15]. The transmission and retransmission mechanisms of RLC are based on sliding window protocol where the feedback (if present) involves both positive and negative acknowledgements. This layer maps application services identified through different Service Access Point (SAP) of PDCP to different logical channels (see Figure 2.5) and connects to MAC layer. The logical channels can also be called as RBs. As mentioned earlier in this section, these can be broadly classified as control and data namely, Signaling RB (SRB) and Data RB (DRB) respectively. The SRBs are mapped to CCCH and DCCH while the DRB is mapped to DTCH of logical channels.

3GPP standard has defined three mode of operation for the RLC layer [7] to support various services, they are:

- Transparent Mode (TM)
- Unacknowledged Mode (UM)
- Acknowledged Mode (AM)

The main functions of the RLC layer can be summarized as follows [4]:

- Transfer of upper layer Payload Data Unit (PDU)
- Error Correction through ARQ
- Concatenation, segmentation and reassembly of RLC Service Data Unit (SDU)
- Re-segmentation of RLC data PDUs
- In sequence delivery of upper layer PDUs
- Duplicate detection of RLC PDUs
- Protocol error detection and recovery
- RLC SDU discard

TM: This is the simplest operational mode of RLC with minimal RLC functionality and does not involve any header part to its SDUs. Since the PDU of this mode is as same as its SDU, there is no overhead to the transmitted packets. As this mode does not need any configuration, this mode is a right choice of operation mode for common control channels (CCH). Through CCH the configurations of other RLC modes can be made with the help of RRC messages. This mode is unidirectional; hence, separate entities will be created for transmission and reception.

UM: This mode of RLC is a light weight of AM which is mainly intended for delay-sensitive and error-tolerant real-time applications [32]. This mode retains most of the functionalities of AM except for the error correcting ARQ procedures. As there is no support for error correction, this mode contains only data PDUs. Similar to TM, this mode is also unidirectional; hence, separate entities are needed for transmission and reception. This mode supports segmentation, concatenation and reordering functions. The reordering of PDUs is done with the help of Sequence Number (SN) attached as a header to every transmitted PDU. To make the overhead involved flexible, 3GPP defined two lengths of SN, 5bits or 10 bits. Based on the application the SN length will be configured by eNodeB through control plane. The format of a typical RLC UM data PDU is shown in Figure 2.6. The Frame Info (FI) gives the information on the type of SDU segment that is being carried in the PDU. The extension (E) flag indicates the presence of Length Indication (LI) field which marks the ending of a SDU in byte position within the PDU. R1 is a reserved bit field whose value is set to ‘0’ by default.

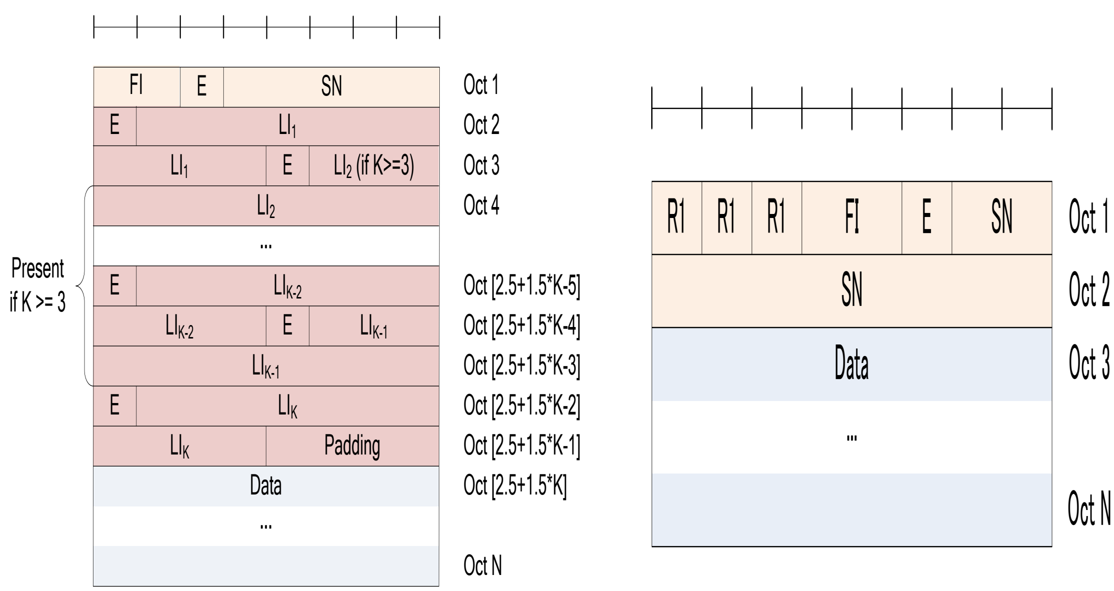


Figure 2.6 RLC UM data PDU with 5bit SN with LI (a) and 10 bit SN (b) [7]

AM: As mentioned earlier, AM is the sophisticated version of UM where the support for retransmission of PDU is available. AM supports all the functionalities that are supported by UM mode. Unlike TM and UM, AM is bidirectional in operation hence, single entity is responsible for both transmitting and receiving functionalities. AM uses the sliding window protocol to deploy ARQ operation, whose window size is configurable during channel creation. This mode does not verify the correctness of received packet content however, it checks for any missing PDUs in sequence. If missing PDUs are detected, the missed PDUs are requested for retransmission by the receiving entity through a control message. Hence, this mode provides two types of PDU formats one for data and the other for control (see Figure 2.7). The cost of retransmitting the lost PDU is the increase in end-to-end delay involved in transmission. The considerable increase in delay made this mode unsuitable for delay sensitive real-time applications. However, the retransmission reduces the error rates involved, which comes in handy for error sensitive background services like web browsing and file transfer.

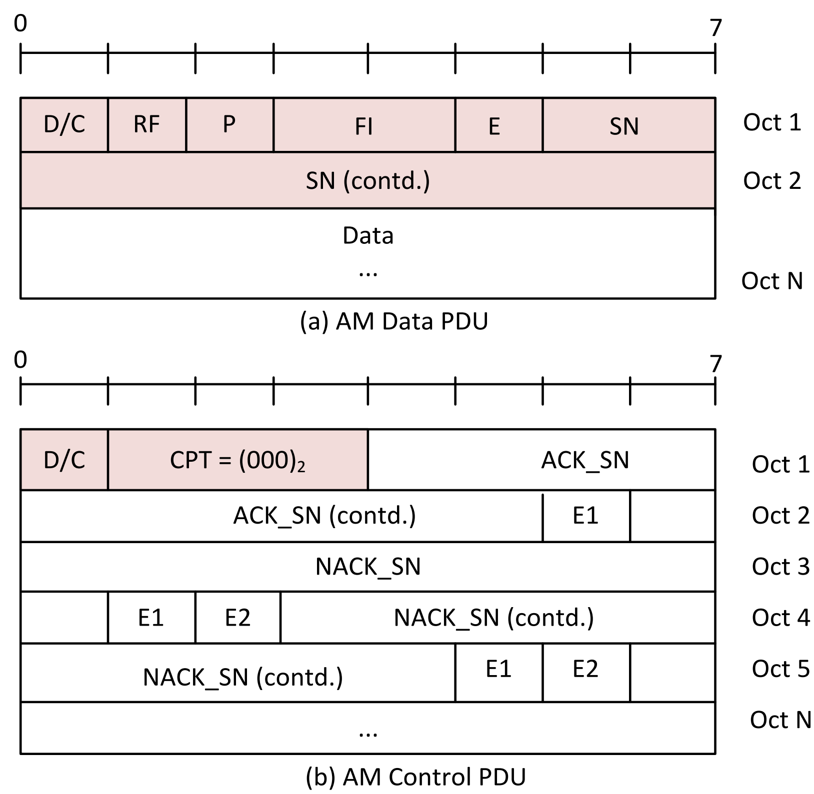


Figure 2.7 RLC AM PDUs [7]

Unlike UM, in AM the SN length is fixed to single value of 10 bits long. The control and data PDUs are identified with the help of D/C bit flag. If the D/C value is set to, the PDU represents AM data (AMD). It represents control PDU if the value is set to 0. The format of control and data PDU differs. The size of RLC PDU is not fixed and can vary from every transmission. The data PDU can be split into three sections, mandatory header part, optional / extension header part and the data part. The extension of header fields are marked by extension flags (E, E1 and E2) based on the requirements. The poll bit (P) is set if the transmitting entity requires a status update on the PDUs received by the receiving entity. The FI and E flags are used for the same purposes as used in UM. The re-segmentation flag (RF) indicates whether the packet is an AMD PDU or an AMD PDU segment. If the RF flag is set then it indicates that the packet is an AMD PDU segment. In that case the header field will contain one more field called segment offset (SO). This SO field is of 15 bits in length and indicates the position of the AMD PDU segment in bytes within the original AMD PDU [7].

The AM RLC end points communicate with a request and response method. The transmission entity maintains a transmission window which stores the transmitted PDU for retransmission. The transmission window is cleared based on the acknowledgement received from the receiving entity through a type of control PDU called STATUS PDU. The process of request for acknowledgement is called polling event which can be triggered based on many reasons (interval, transmission window half filled, etc.). The polling event is carried out by setting the polling bit (P) in the data PDU by the sender. The frequency at which the STATUS PDU is sent depends on the frequency at which the poll bit is set by the transmitter. If the acknowledgement is very slow the transmission window gets full and the transmitting entity stops the transmission leading to a phase called window stalling. The control PDU is identified as a STATUS PDU (see Figure 2.7.b) with the help of a field called Control PDU Type (CPT). This is 3 bit field, where the value $(000)_2$ identifies the PDU as status message [7], the remaining values are reserved for future use (see Table 2.1). The STATUS PDU will have an acknowledgement SN which informs the transmitter that the receiver has received all the PDUs till one number less than the mentioned SN. The STATUS PDU may also have a list of NACK SNs whose presence is indicated by the extension bit (E1). If NACK has to be made for an AMD PDU segment, then the extension bit E2 is used to indicate the presence of SO in the STATUS PDU to indicate the exact lost segment.

Table 2.1 CPT value description [7]

CPT value	Description
$(000)_2$	STATUS PDU
$(001)_2$ to $(111)_2$	Reserved, will be discarded by receiving entity

The transmitting entity on receiving the STATUS PDU will advance the transmission window based on the ACK SN and initiate the retransmission for those PDUs with SN listed in the NACK list. There is an upper limit for number retransmissions that can be tried for a given PDU. If the maximum tries exceed, the PDU will be discarded and informed to the upper layer. The receiving entity on the expiration of reordering timer detects a missing PDU in sequence and initiates the SDU discard operation for the corresponding PDU, assuming that the PDU is lost forever. An example of SDU discard is shown in Figure 2.8. In this example, the PDU7 contains the last segment of SDU22 and first segment of SDU23. The PDU is lost during transmission and after maximum retransmission attempts, the HARQ transmission entity discards the packet. During reassembly at RLC, missing of PDU7 is detected and if the error is found to be non-recoverable, then SDU discard procedure will be initiated. Since, PDU7 contains the segments of SDU22 and SDU23, both SDUs will be discarded.

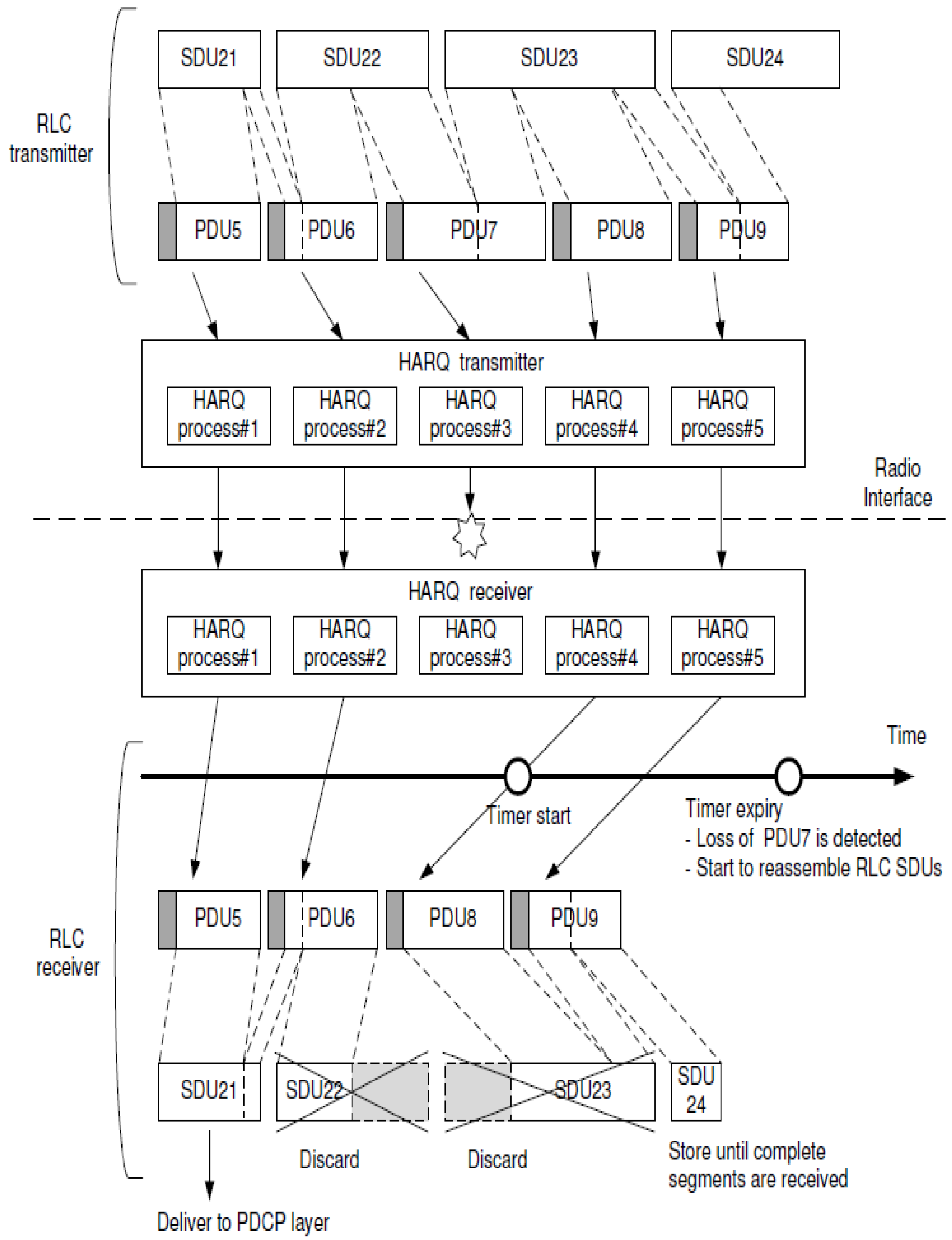


Figure 2.8 Example of PDU lost and SDU discard operation [32]

Table 2.2 summarizes the important functions of RLC layer with respect to its operation modes.

Table 2.2 Functions of RLC layer summarized

Functions	RLC TM	RLC UM	RLC AM
Data PDU	✓	✓	✓
Control PDU	×	×	✓
Error Correction (ARQ)	×	×	✓
Bi-Directional Entity	×	×	✓
Concatenation, segmentation and reassembly	×	✓	✓
Duplicate detection	×	✓	✓
SDU discard	×	✓	✓
PDU re-segmentation	×	×	✓
Mandatory Header Size	0 Bytes (No header)	1 Byte (5 bit SN) 2 Bytes (10 bit SN)	2 Bytes
Applications	CCH	Delay sensitive	Error sensitive

2.5 Advanced Video Coding (AVC) / H.264

AVC also known as H.264 is a video compression standard, which is used as one of the most common codec for video streaming and High Definition (HD) video applications [44]. AVC is a block-oriented motion-compensation-based codec standard developed by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG). The partnership between VCEG and MPEG for this standardization is called as Joint Video Team (JVT). Some of the major applications that use AVC as their codec are Blu-ray disc, YouTube videos, Digital video Broadcasting (DVB) and video conferencing [37, 44]. As it can be seen, AVC supports variety of applications which demands various levels of requirements starting from video resolution to decoding and playback complications. Hence, AVC provides different video profiles wherein the above mentioned factors are modified to suit the applications. For example, HD videos need more bitrate and video resolution where the video clarity is given more importance than the complications involved in the decoder design. However, in case of mobile applications, considering the limited channel bandwidth and computational power of mobile terminals, the video clarity and resolution are compromised for low power and bandwidth consumption. H.264/AVC is the successor of H.263 (of ITU-T standard) and MPEG-4 Part 2 (of ISO/IEC standard), where the video bit rate has been reduced less than half of its predecessors without compromising the video quality and design complexity of encoder and decoder.

Video can be considered as a sequence of pictures screened with temporal relation. Each pictures of this sequence is called as video frame. The compression of video can be applied in two levels, one at the picture level and the other on the temporal relation between the successive frames. The picture compression is called as spatial compression which is performed by applying the Discrete Cosine Transform (DCT) on the picture and the resulting bit stream is further encoded using entropy coding [43]. The method applied during spatial compression is similar to the compression method of Joint Photographic Experts Group (JPEG). Though video contains sequence of frames, most of the contents of successive frames remain unchanged. Hence, transmitting the whole content of each frame will be inefficient as the redundant information will be transmitted at the cost of bandwidth. The video transmission will be more efficient if the successive frames carry only the delta of previous frame. In order to maintain synchronization and for error recovery, one frame without delta is sent on a regular interval. The frames which carry the complete information are called as key-frames and the frames which carry only the differential values can be called as delta-frames [2]. Figure 2.9 gives a basic analogy on the contents carried in key-frames and delta-frames. In Figure 2.9, the frame 1 and 5 are the key-frames as they convey some useful information on their own. However, the frames 2, 3, and 4 are the delta-frames as they carry only the difference of frame 1. The delta-frames convey some useful information only if their corresponding key-frame is available.

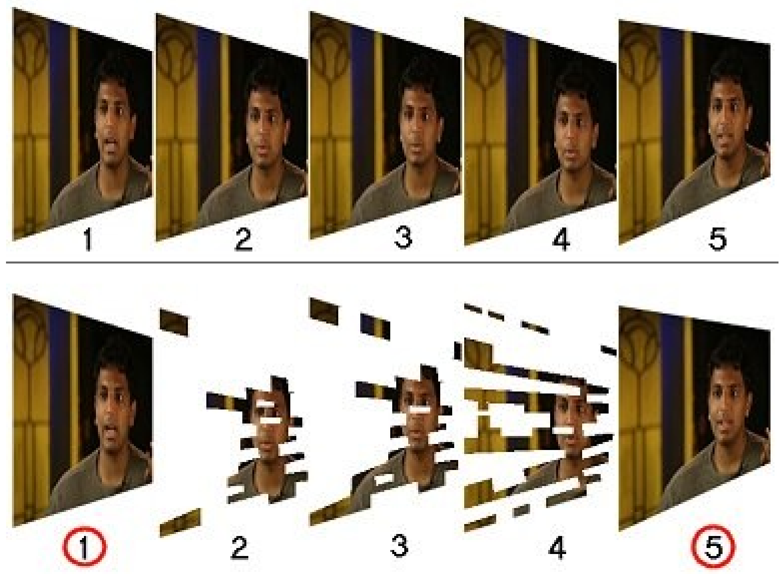


Figure 2.9 Key-frames and delta frames in video transmission [2]

In video, other than a major scene change the successive frame shares most of its content. However, the contents or objects may not be remaining in the same position as the previous frame. This information of positional changes for an object is represented using the motion vector. The movements of the objects in the video are tracked by segmenting the video frame into small blocks called as macroblocks. A motion vector is defined for each macroblock. The motion vector is calculated by matching the macroblock of previous frame with the current and calculating its positional deviation in terms of pixels. The motion vector is transmitted along with the delta frames. However, the key-frames does not require motion vector.

H.264/AVC has defined three types of frames [31] for its video encoding and decoding process. They are as follows:

I-Frame: Intra-coded (I) frame is the key-frame. This frame contains the complete information for an image construction therefore this frame is not dependent on any other frames. Since this frame contains independent data, it does not contain motion vector [37]. This frame is the least compressed frame of all three. This frame is treated as a standalone picture where the spatial compression and prediction is applied to compress its data. Figure 2.10 shows an example of I-frame where the frame contains all the information needed to construct a complete picture of a Pac-Man.

P-Frame: Predictively-coded (P) frame is a type of delta frame. This frame applies temporal prediction and carries only the differential content of previous I or P-frames. This frame contains the reduced image data and motion vector of all macroblocks [37]. The compression rate is higher than I-frame and the frame size will be less than I-frame. Since this frame contains only the differential content, the decoding process of this frame depends on both current and the previous frame. In Figure 2.10, the P-frame is represented where it carries only the information on the positional change of pellets (Pac-Man is shown in grey to indicate that the information is not being transmitted).

B-Frame: Bi-directionally predictive (B) frame is also a type of delta frame. Unlike P-frame, the temporal prediction is made based on two frames. The B-frame is derived based on one previous I or P-frame and one successive I or P-frame. Hence, both the source frames has to be successfully received to decode this frame [37]. This frame is the most compressed frame and can contain both image data and motion vector. Since, this is the most compressed frame, more the number of this frame smaller will be the bandwidth consumed for transmission. As it can be seen from Figure 2.10, the B-frame carries only the differential information of its previous P-frame and successive I-frame.

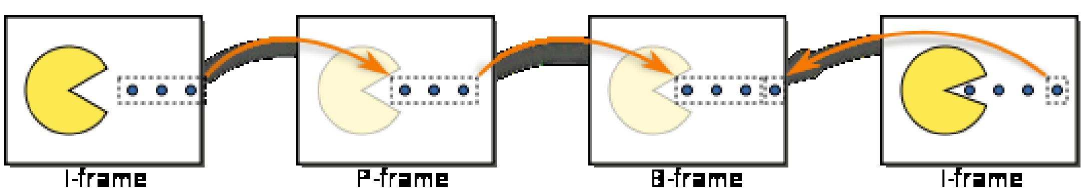


Figure 2.10 Depiction of I,P,B-frames [48]

In addition to the above mentioned types, H.264 also defines Switching I (SI) and Switching P (SP) frames which facilitates the switching between coded streams. However, these frame types are defined under extended profile which is used in high end and more sophisticated decoders. Hence, these frames are not considered for mobile communication [48].

I, B and P- frames are arranged in a sequence to form a pattern and this is repeated throughout the transmission. This pattern is called as Group of Pictures (GOP) [31]. Figure 2.11 shows the dependencies of frames for a GOP pattern of size 16 frames with 3 B-frames between successive I or P frames. The figure shows that the frame B_1 is derived from I_0 and P_4, whereas P_4 itself is derived from I_0. Similarly, P_8 is derived from P_4. So to decode B_5, we must successfully receive I_0, P_4 and P_8. The derivations of P frames are implementation specific. The AVC frame creation implementation considered for this thesis work is as shown in Figure 2.11.

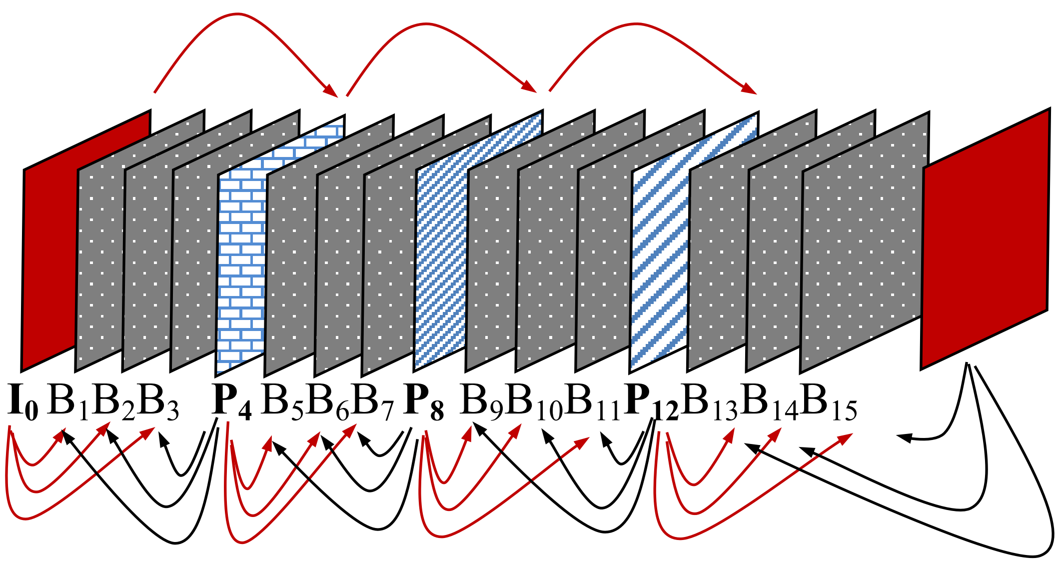


Figure 2.11 GOP pattern and frame dependencies [31]

2.6 Internet Protocol

The most widely used Internet Protocol plays a vital role in computer communication. This protocol maps to the network layer of legacy OSI model and represents the layer 2 of TCP/IP protocol suite. The standardization of IP is governed by Internet Engineering Task Force (IETF) [45]. As a function of network layer, IP ensures the successful delivery of data packet to the right system on right time. This is achieved by defining a logical address to each entity of the network system called IP address. The most successful and widely deployed version of IP is version 4 (IPv4), which is being succeeded by version 6 (IPv6) [45]. The specifications of IPv4 are defined in [28]. The major difference between IPv4 and IPv6 is in the address space. Due to the huge increase in number of devices connected to the Internet, the address space provided by IPv4 was not enough. This has been overcome in IPv6 where the IP address length has been increased from 32bits to 128bits. IPv4 was not designed with security features which made this more susceptible to various attacks [46]. However, this has been addressed in IPv6. Though IPv6 is the future of internet communication, IPv4 still remains to dominate the internet video traffic [16]. Hence, IPv4 is considered for validating the work presented in this thesis.

IPv4 supports packet fragmentation and reassembly for which a parameter called Maximum Transferrable Unit (MTU) is defined. This parameter value has to be agreed between the transmitting and receiving entities. The packets with size more than the configured MTU will be segmented by the transmitting entity and indicated to the receiver as segmented packet using one of its header fields. At the receiver, the packets of size more than the configured MTU will be dropped. IPv4 header has a minimum size of 20 bytes with 13 mandatory fields and one optional field. A brief description of the header fields of IPv4 are as follows [46]:

- ***Version:*** Provides the version of IP used
- ***Internet Header Length (IHL):*** This field gives the size of IP header in multiples of 32bit. This field is of 4 bit in length, whose minimum value is 5 (5*32bits = 160bits = 20 bytes) and maximum header length that can be represented is 60 bytes.
- ***DSCP & ECN:*** These fields are modified from the old header field of IPv4, ToS. These fields play a major role in QoS and congestion management. These fields and their importance are further discussed in the next subsection.
- ***Total Length:*** This field gives the total size of the IP packet including both header and payload in bytes.

- ***Flags & Fragment Offset:*** These fields are used for IP packet fragmentation control and communication.
- ***Identification:*** This is used for unique identification of the transmitted IP packet in the network.
- ***Time To Live:*** An important field of IP header which defines the life time of an IP packet in the network, thereby preventing the IP packets to prevail in the network forever.
- ***Protocol:*** This identifies the protocol used in the payload.
- ***Header checksum:*** This is used for error detection for IP header. A separate header checksum is added at the IP header level in addition to the data checksum of the transport layer, to make the router function easier and faster in dropping the corrupted packets.
- ***Source & Destination IP address***: Each field is of 32bits in length and indicates the source and destination nodes. These addresses may be changed in transit based on the network configuration.

2.6.1 Differentiated Services in IP

The IP provides a way to differentiate the application packets even at the IP header level for effective routing decisions which help in achieving the QoS requirement of a service. The field in the IP header (IPv4) that provides this information is Differentiated Service Code Point (DSCP). DSCP field is a modification of old field called ToS, and it is backward compatible with ToS. The first 6 bits of ToS are used for DSCP representation while the last 2 bits are used as Explicit Congestion Notification (ECN) (see Figure 2.12). The ECN field is used in conjunction with the TCP to reduce the packet drop during congestion. If the network entities of the data path support ECN, then during congestion this flag is set to Congestion Encountered (CE) value and informed to the receiver. The receiver echoes it to the transmitter to reduce its transmission rate. These fields are also valid for IPv6. DSCP field is of 6 bits length and its values are classified into 3 groups namely Pool 1, Pool 2 and Pool 3 [25]. IETF has defined a set of standard actions for some of the code points of Pool 1. The Pool 2 is reserved for experimental and local use, whereas Pool 3 is kept for future use when Pool 1 gets exhausted. The range of DSCP values used for each pool is given in Table 2.3, where X represents the possible value as either ‘0’ or ‘1’.

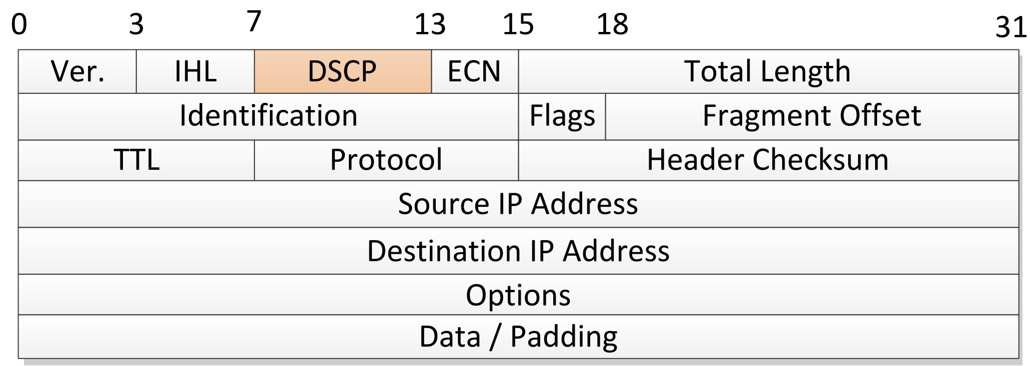


Figure 2.12 IPv4 header showing DSCP field [46]

Table 2.3 DSCP pool categories [25]

Pool	DSCP value in binary	Assignment policy
1	XXXXX0	Standard action as per defined as per the identified service.
2	XXXX11	Experimental or local use.
3	XXXX01	Experimental or local use. This can also be used for standard action in future if the Pool 1 gets exhausted.

2.7 Related Work

The characteristics of wireless medium differ totally from wired medium, where in wireless medium the channel condition fluctuates with respect to time. This demands the wireless system to be channel aware, which made the highly efficient layered approach of wired medium less efficient for wireless. Hence, cross layer design approach has been applied for wireless medium to adapt and optimize the system according to the situation demands. Lots of studies and research work have been made towards applying cross layer design in mobile communication systems to optimize its performance. The author of [36] discusses various options for deploying cross layer design for mobile video transmission. The author categorizes the possible approaches into two namely, top-down and bottom-up. A good example for top-down approach would be application content based adaption where the information from top layer is passed on to the bottom layer for action. Channel quality based adaption where, the control flows from bottom layer to top is an example for bottom-up.

Wang et al. have presented an efficient end-to-end solution for video broadcast over mobile WiMAX networks in [42]. In this paper, various cross layer optimizations are applied to achieve robust video quality with better coverage and energy efficiency. In [29], the author uses cross layer design approach to drop few video frames at the video server based on the channel quality estimation which in-turn increases the utilization factor of backhaul resource. Similarly in [22], Luo et al. have proposed a QoS aware channel condition based adaptation scheme, where the video encoding is adapted in such a way to maintain the committed QoS. The wireless physical resource scheduler takes necessary actions to make sure that the delay constraints are met and the video encoding process uses channel quality feedback to reduce video distortion. The authors of [34] use Mean Opinion Score (MOS) to measure the QoE of the received video and uses application rate adaptation scheme to optimize the resource utilization of the LTE system. The authors of [24] have applied application level FEC codes to improve the QoE of received video. However, this approach will introduce a decoding delay at the receiver side.

Many research works have been published with an emphasis on delay sensitivity of the real time applications and several schemes have been proposed based on it. Bae et al. have proposed a delay aware packet scheduler in [13]. In their work, each service is defined with a delay weight based on their delay requirement. The packet scheduler schedules the channel with more weight to a set of PRB with better Signal to Noise Ratio (SNR). This ensures the packets to be transmitted with lesser number of retransmission thereby achieving the delay requirements. A new approach has been put forth by Lu et al. for wireless video transmission in [21]. In this paper, the authors have proposed a scheme called cooperative HARQ, where the receiver is assisted by nearby nodes. The nearby nodes echo the packets received by them originally sent by the transmitter with addition of parity bits. This reduces the delay involved in transmission. In [35], the HARQ process is made delay aware of the packets being transmitted. As per the proposed scheme each packet is attached with a time value. The HARQ process before initiating the (re)transmission process, it calculates the propagation delay of the packet. If the delay involved exceeds the timestamp of the packet, the packet will be discarded without being transmitted thereby saving the physical resource.

Several efforts have been undertaken for optimizing LTE networks for real-time applications such as voice and video. Most of them concentrate on efficient scheduling of physical resources based on user positions and services used. There are only few approaches that focus on user experience and content awareness. Content aware networks have become an active area of research in recent times. The authors of [19] have proposed content aware video transmission over wired networks by providing unequal error protection using DiffServ networks. They have introduced a packet marker system which color codes the video packet based on its content. During congestion scenarios the packet coded with least important color code will be dropped to restore the situation to normal. Karachontzitis et al. have designed a LTE based MAC scheduler that uses content awareness of the packet being scheduled in [18]. The eNodeB uses the Network Abstraction Layer (NAL) of H.264 video frame to identify its category as either I or P. As I frame have more importance in decoding process, it will be scheduled using the PRBs with good SNR for the concerned user. However, the author did not mention about the priority and processing details of B frames. In [33], the authors have proposed a packetization scheme that constructs the application layer packet in such a way that it is decomposed exactly into an integer number of equal-sized radio link protocol (RLP) packets. This reduced the decoding delay involved at the receiver end as it need not wait for segment of SDU. However, this mechanism does not hold good for varying channel condition, where the packets might be re-segmented. Benayoune et al., the authors of [14] have reduced the loss of critical packets by increasing SDU discard timer value of the critical packets at the RLC layer of WiMAX system. This timer is responsible for allowing the RLC entity to retransmit the lost segment of a SDU. In addition, the authors have also used an error robustness tool called Flexible Macro-block Ordering (FMO) which further increased the reliability of the video frames. In most of the above mentioned work, the video frame importance level is defined only based on the type of video frame. However, it is possible to have different importance level within same type of frame. To the best of our knowledge, there is no existing approach that emphasizes and utilizes the packet importance level within the same type of frame. Literature survey reveals that majority of the research publications focus on downlink transmission and do not provide a simple and efficient end-to-end solution by making the whole system content aware.

Chapter 3

Content Aware Adaptive HARQ Retransmission Scheme

3.1 Introduction

In this chapter, a proposal is made for a new scheme called Content Aware Adaptive HARQ Retransmission (CAAHR in short) which reduces the unnecessary retransmission of least important packets and exploits the true benefits of HARQ for most important packets, thereby increasing the QoE of the received video. The proposed scheme is explained for downlink video transmission scenario in a LTE FDD system. However, this approach can be applied to uplink transmission and TDD variant of LTE system. The importance of an application packet (video frame) is identified at the sender (video server), and it is notified to the eNodeB which is serving the destined UE, using the DSCP field of IP header that carries the application packet. The maximum retransmission count of HARQ process is determined based on the packet importance identified at the eNodeB.

3.2 Proposed Scheme

As discussed in the earlier Section 2.5, all video frames do not carry equal level of importance, and decoding one type of frame depends on the successful reception of other types of frames. In H.264, I frame is considered to be the most important frame in the GOP. The P frame has more importance than the B frame. Even for successful decoding of a P frame, there should have been a successful reception of the previous P or I frame of the same GOP. This means that the importance of P frame is not equal throughout the GOP sequence. The first occurring P frame has more importance than the last P fame in a GOP. Though the importance is stressed in application level, this information is lost while reaching the physical transmission. The loss ratio of important packets can be decreased considerably by adapting the HARQ retransmission count. Hence, in this scheme the importance of the video packet is identified at the MAC layer and the retransmission count is changed accordingly.

According to the proposed scheme (CAAHR), the packet importance is defined at the video server based on its content and it is used at the HARQ level. In this method, the packet loss ratio for important frames is decreased by increasing the maximum number retransmission attempts made by the HARQ entity for packet transmission. Increasing the HARQ maximum retransmission count for all packets will affect the throughput and latency of the system as the retransmission packet has more priority than the new transmission. To overcome this issue, this approach increases the maximum count only for selective frames and keeps the system default count for other frames.

In the existing system, the information on the importance of a video frame is only available at the video server, and this is not propagated till eNodeB. However, to provide a differentiated service based on the packet content at HARQ, eNodeB should be aware of its content. This can be achieved by using the IPv4 header field DSCP, which is used for differentiated services in the IP network. In this approach, we consider a GOP of size 16 which has 3 P frames, 1 I frame and remaining as B frames. Hence, we choose 5 values of DSCP from the Pool 2 $(XXXX11)_2$ [25] to represent I, P, and B frames at the IP header level (see Table 3.1).

The steps to be performed at the video server are explained in a flow chart shown in Figure 3.1. Based on the GOP pattern used by the video stream, the type of next frame (I/P/B) to be transmitted is determined, and its corresponding DSCP value mentioned in Table 3.1 is filled in the IPv4 header. The values of DSCP are categorized as I_{CP}, P_{CP} and B_{CP} for easy representation in the flow chart. Since the IPv4 supports IP fragmentation, the fragmented packets should also carry the same DSCP value to make sure that the same priority is given during the HARQ retransmission. Using DSCP field for differentiating the video frames also ensures differentiated service for I, P and B frames in wired network elements like backhaul routers.

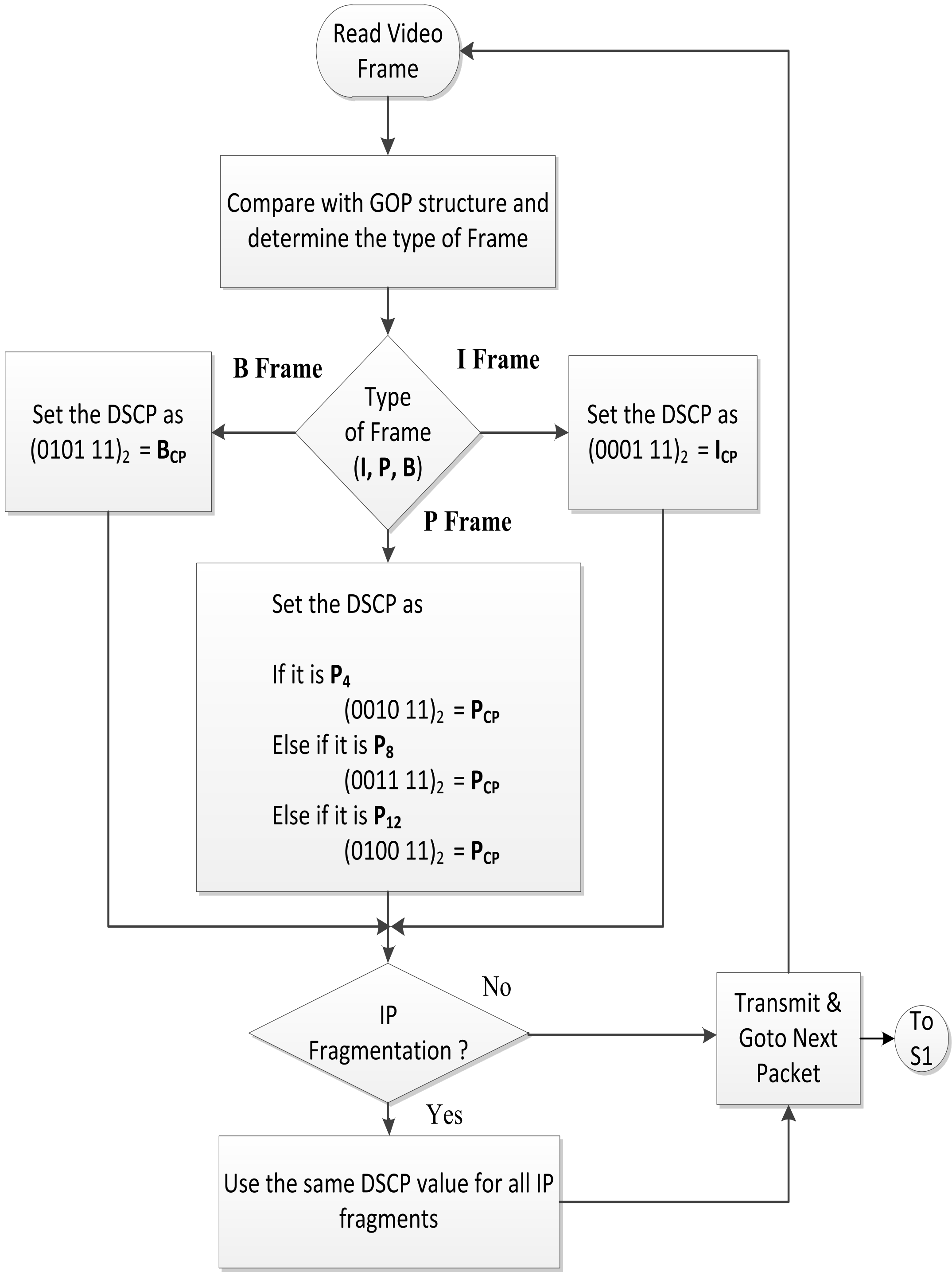


Figure 3.1 Adaptation algorithm at video server

The steps to be followed at the eNodeB are shown as a flow chart in Figure 3.2. When a GTP-U packet is received over S1 interface at eNodeB from S-GW, the encapsulated IP packet is retrieved and forwarded to the PDCP layer for processing. As per our proposed scheme, the PDCP layer labels the packets as I, P_4, P_8, P_{12} or B frames (see Figure 2.11 for GOP pattern) according to the DSCP values present in the IP header.

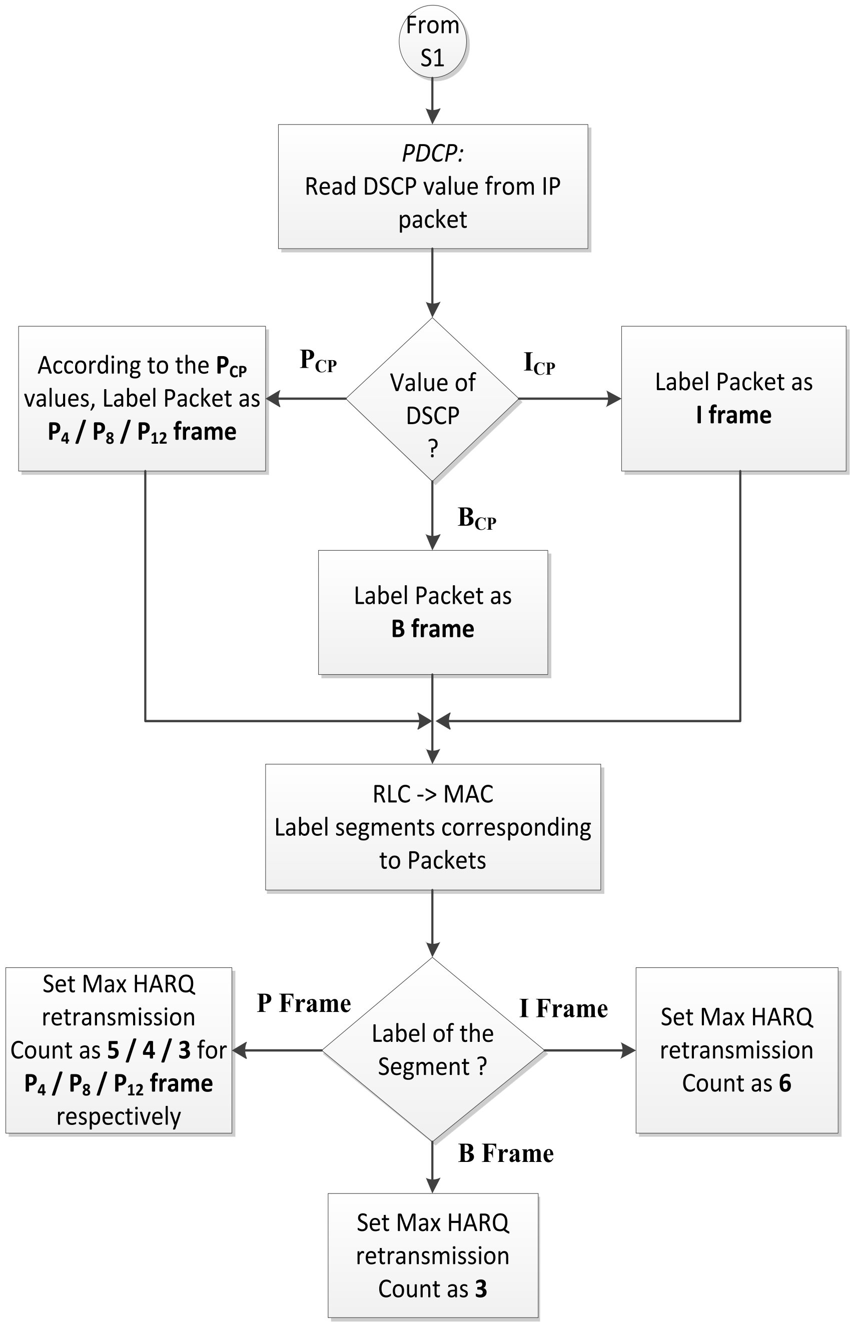


Figure 3.2 Adaptation algorithm at eNodeB

These DSCP values used will be agreed upon by both eNodeB and video server during the connection setup process. The label is attached to the packet till it is either successfully transmitted or discarded after exceeding maximum number of retransmissions. The packet along with its label is propagated through the layer 2 protocol stack (RLC and MAC) till it reaches the HARQ entity for transmission. The HARQ entity utilizes this label to decide the retransmission count that can be allowed for the scheduled packet. Since I frame is the most important frame to decode the complete GOP and considering its less frequency of occurrence, the maximum retransmission count for this frame will be set as a high value. The importance of P frame decreases as the position index of P frame increases within a GOP, and all P frames are less important than I frames, and B frames have least importance of all with respect to the decoding process. As there are different levels of importance for P frames, different retransmission counts have been used (see Table 3.1). The least value of retransmission count is given to the B frame.

3.3 Simulation Model

Validation of the proposed scheme has been done using simulation in OPNET network modeler [26]. For this simulation, a single cell LTE network with two users in different coverage regions and a general purpose video server have been used. The video is streamed from the video server to the users (downlink). The H.264/AVC video trace of Star Wars IV [12] is used for simulation, and the properties of the video trace used are given in Table 3.2. The network model used for simulation is shown in Figure 3.3.

To validate the scheme under practical wireless channel condition, Suburban Macrocell path loss model is used. Out of the two users, one is placed in better channel condition where it will be assigned with a higher MCS value; another user is placed in a relatively bad channel condition where it will be assigned a lower MCS value. The wired links between network entities are assumed to be error free. Important LTE system parameters used in the simulation are summarized in the Table 3.3 (extended list of parameters can be found in Appendix B).

To simulate the proposed scheme, code modifications have been done at the video server node and the eNodeB node of the modeler. The DSCP value of the IPv4 header which carries the video frame from the video server is modified according to the type of video frame ($I/P_4/P_8/P_{12}/B$) being carried.

Table 3.1 Values used in simulation for proposed scheme

Packet Type	HARQ Max Retransmission Value	IPv4 DSCP value
I Frame	6	$(000111)_2$
P_4 Frame	5	$(001011)_2$
P_8 Frame	4	$(001111)_2$
P_{12} Frame	3	$(010011)_2$
All B Frames	3	$(010111)_2$

Table 3.2 Star War IV video trace details

Parameters	Value
Codec	H.264 / AVC
Quantization Parameter	10
Resolution	CIF (352 X 288)
Frame Rate	30 frames per second
GOP Size	16
Number of B Frames between I or P frames	3
GOP Structure	$I_0\ B_1\ B_2\ B_3\ P_4\ B_5\ B_6\ B_7\ P_8\ B_9\ B_{10}\ B_{11}\ P_{12}\ B_{13}\ B_{14}\ B_{15}$

The DSCP values used are chosen from the pool 2 ranges [25] and assigned to the respective frames based on the values shown in Table 3.1. Based on the received DSCP values, the type and importance of a video frame is identified at the PDCP layer of eNodeB, and the information is propagated to the MAC layer. The HARQ retransmission limit is changed dynamically based on the importance of a frame that the packet is carrying. If a MAC PDU contains even a segment of I frame, the complete PDU will be treated as I frame, and the retransmission count will be set accordingly. The value of HARQ retransmission count used for different frames in our simulation is given in Table 3.1.

Table 3.3 LTE system parameters

Parameters	Value
LTE Mode	FDD
eNodeB Operating Power	46 dBm
UE Antenna Gain	-1 dBi
Number of Antennas	1 (Tx / Rx)
LTE Bandwidth	20 MHz
Path Loss Model	Suburban Macrocell
Multipath Channel Model	ITU Vehicular B
RLC Mode	UM
Max HARQ retransmission (For existing scheme)	Uplink – 3, Downlink - 3
Max HARQ retransmission (For proposed scheme)	As per Table 3.1
Internet Protocol version	IPv4

As discussed earlier, loss of a P frame will affect all the B and P frames derived from it, whereas the loss of an I frame will affect the complete GOP and make it non-recoverable. A MATLAB program has been written to compute the total number of non-decodable frames based on the type of frame lost during the transmission. To measure the quality of the video received (Q) quantitatively, the following formula adapted from [20] is used,

$$Q = \frac{N_{dec}}{N_{rcvd}} = \frac{N_{rcvd} - N_{non\_dec}}{\mathrm{N}_{I-rcvd} + \mathrm{N}_{P-rcvd} + \mathrm{N}_{B-rcvd}} \quad (3.1)$$

where, $0 \leq Q \leq 1$ (1 being good quality and 0 being bad). N_{dec} is the total number of frames that can be decoded and played successfully; this is derived from the difference of total received frames without error (N_{rcvd}) and total frames that is received yet cannot be decoded (N_{non_dec}). $N_{I\text{-}rcvd}$, $N_{P\text{-}rcvd}$, $N_{B\text{-}rcvd}$, is the total number of I, P, and B frames received without error, respectively.

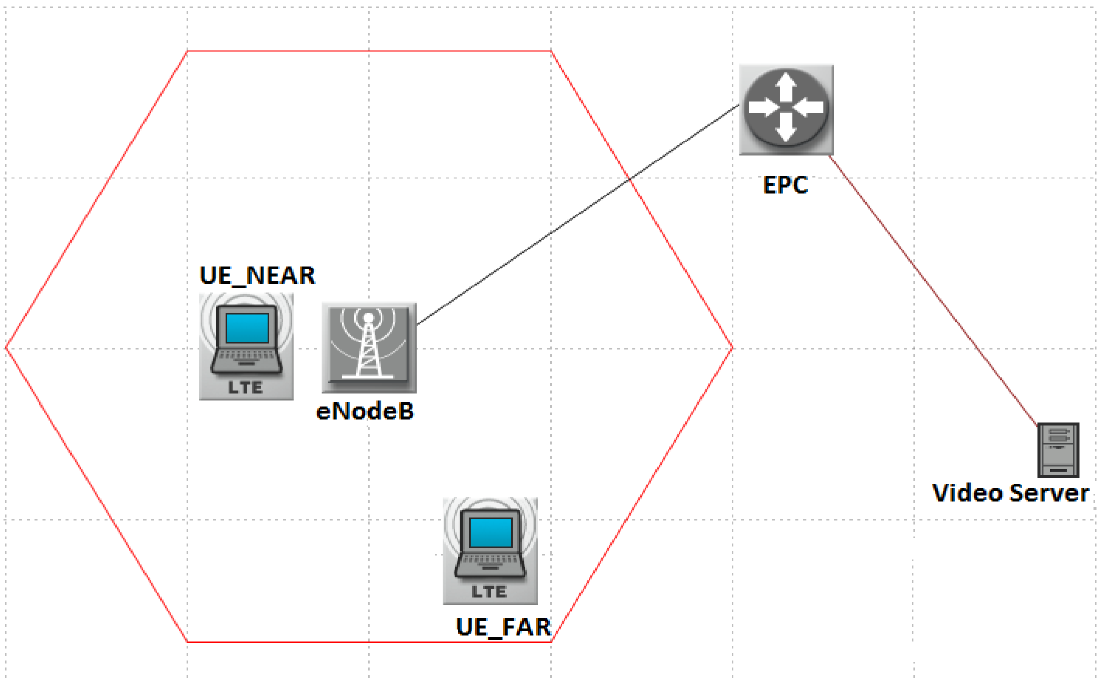


Figure 3.3 OPNET simulation model

3.4 Results and Analysis

The simulation has been performed for both the existing system and the proposed scheme under the same network conditions. The obtained results are compared and analyzed in this section. Due to suburban path loss and multipath channel model, the channel condition varies frequently; hence, most of the graphs are presented as time average for better comparison. The two users UE_Far and UE_Near are kept in two regions of coverage. UE_Near is in a good coverage region and hence it has been allocated with an average MCS index of 25 and UE_Far which is in relatively poor coverage region has been allocated with an average MCS index of 18 (as shown in Table 3.4).

Figures 3.4 and 3.5, show the HARQ downlink drop rate for users at far and near position from the eNodeB respectively. Both graphs compare the proposed scheme (CAAHR) with the existing approach. It can be noticed that there has been a reduction in drop rate by as much as 30% with the proposed scheme which is due to the increase in maximum retransmission count for I and P frames.

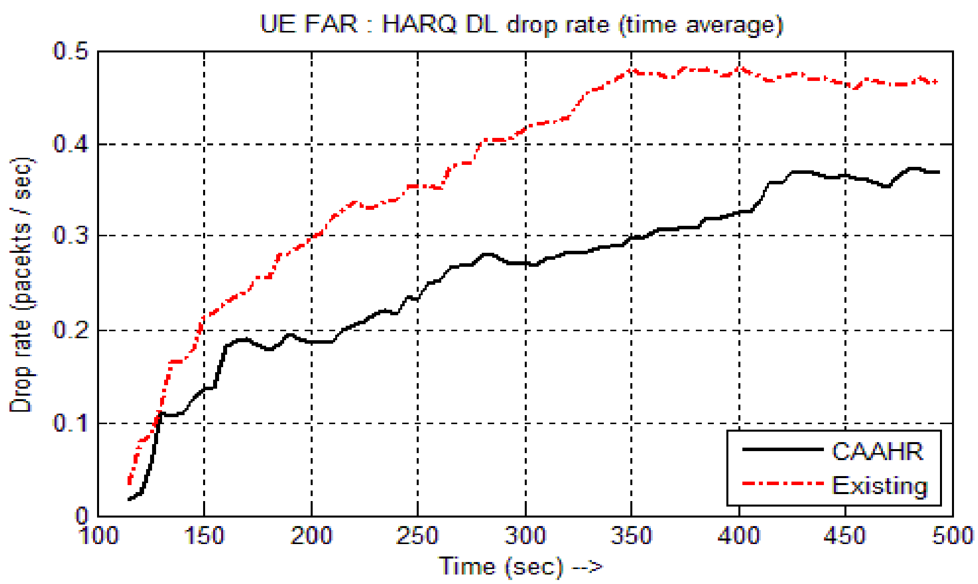


Figure 3.4 HARQ downlink drop rate for UE_Far

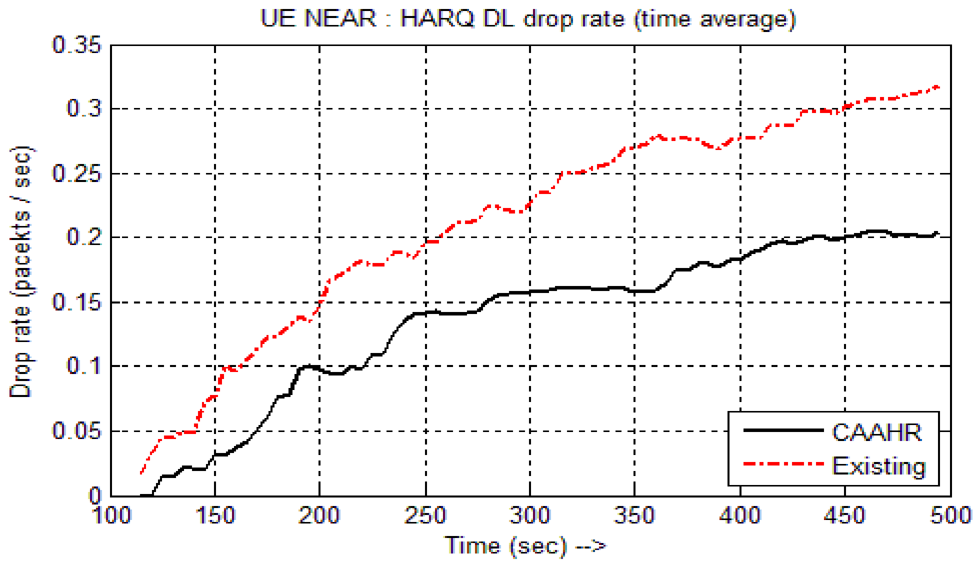


Figure 3.5 HARQ downlink drop rate for UE_Near

Figure 3.6 compares the received packet delay of UE_Near for proposed and existing approach. The difference in delay for UE_Far is similar to UE_Near; hence, it is not shown here. In the proposed scheme there is an increase in delay, and this is expected as the retransmission count increases. The delay increase happens only when the channel condition is bad and important packets are being scheduled. The major percentage of GOP consists of B frames (75% for GOP 16 B 3 pattern). In our approach the B-frame retransmission count is not changed; hence, the delay is not much affected.

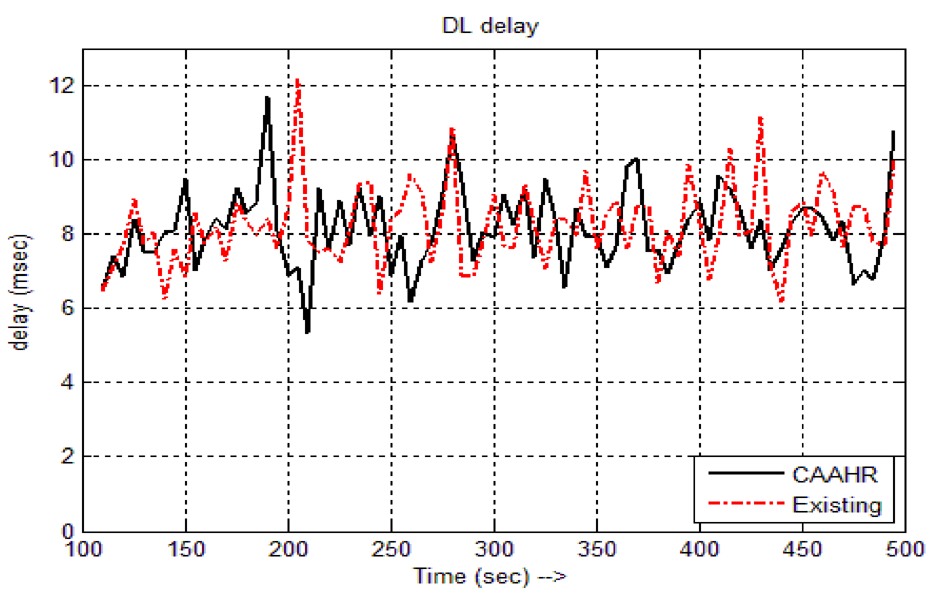


Figure 3.6 Delay for UE_Near

Figure 3.7 shows the PDSCH utilization factor for the existing and the proposed approach. The graph shows a slight increase (0.2%) in channel utilization which is expected; however, the increase in utilization is compensated by the increase in received video quality.

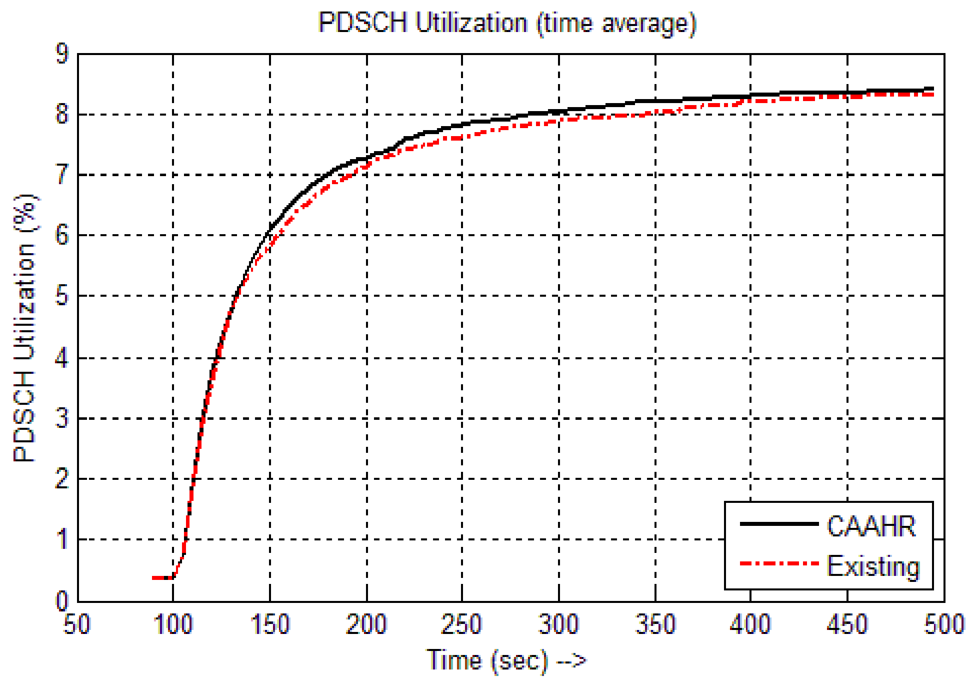


Figure 3.7 PDSCH utilization factor

Figures 3.8 and 3.9 show the received video quality (Q) computed with formula given in Equation (3.1) for both UE_Far and UE_Near, respectively. It is evident from these two graphs that the proposed scheme gives a consistent and noticeable improvement to the video quality over the existing approach. The dips in the proposed scheme are due to bad channel condition and yet it still outperforms the existing approach.

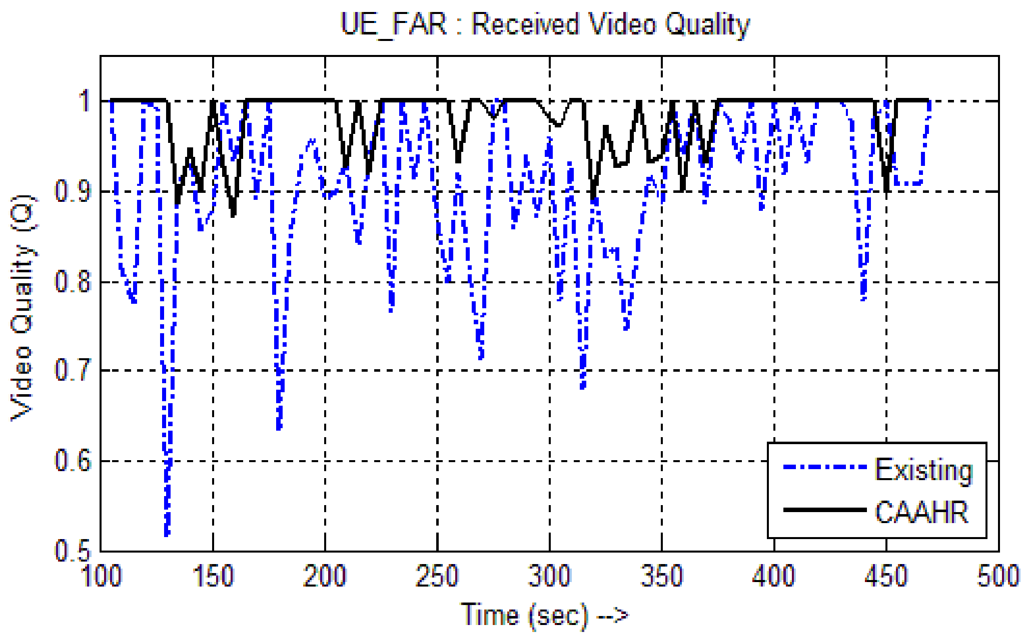


Figure 3.8 Calculated received video quality for UE_Far

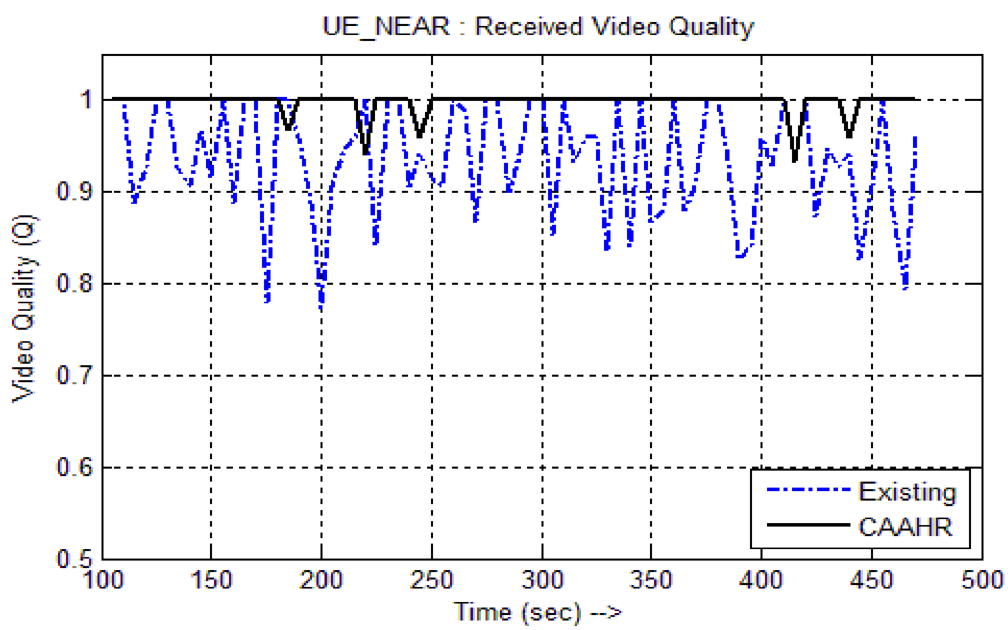


Figure 3.9 Calculated received video quality for UE_Near

Table 3.4 shown below summarizes the packets transmitted and received during the simulation, and we can notice that there is a significant reduction in I frame loss and considerable reduction in P frame loss. Since the loss ratio of important frames such as I,

P_4 and P_8 are drastically reduced; we can see good improvement in the number decodable frames, thereby improving the overall video quality.

Table 3.4 Received packet summary

Metrics		UE Near (MCSAvg = 26)		UE Far (MCSAvg = 18)	
		Legacy	CAAHR	Legacy	CAAHR
Received packet		11457	11531	11216	11265
Number of frames lost	**I**	20	0	32	5
	P_4	18	0	23	8
	P_8	12	2	27	10
	P_{12}	12	4	24	11
	B (all)	97	89	238	245
Calculated non-decodable frames		761	42	1041	166
Calculated received video quality (Q) in percentage		93.4 %	99.6%	90.7%	98.5%

3.5 Summary

This chapter introduced a method to make the RAN content aware for video transmission and a new adaptive HARQ retransmission scheme. This scheme makes use of the content awareness to improve the received video quality by providing unequal error protection for the critical video frames. The simulation strategy and its obtained results have also been discussed. This scheme in particular provides an improvement in the video quality received in the order of 6 to 8 % depending on the user position. The main advantage of this scheme is that these benefits come with negligible increase in delay and no extra processing needed at the UE.

Chapter 4

Hybrid Mode Radio Link Control for Efficient Video Transmission

4.1 Introduction

This chapter introduces a new scheme for an end-to-end video transmission that allows the network to be content aware, which is also independent of transmission direction. The network elements are made content aware with help of DSCP field in the IP header. The identified critical video frames are error protected with a novel idea of creating a new operating mode for RLC layer called Hybrid Mode (HM). In this mode we combine the benefits of both existing AM and UM of RLC for an effective video transmission. The proposed HM RLC layer is content aware of the video transmission, thereby identifying the critical frame of a transmission sequence. Only the identified critical frames are given error protection with help of ARQ mechanism, thereby avoiding the unnecessary retransmission for non-critical packets. This approach gives an edge over the existing legacy operation modes of RLC by achieving good video quality at the receiving end without compromising on the transmission delay. This scheme is explained for downlink transmission, yet it can also be applied to uplink without any modifications.

4.2 Challenges in Existing System

The importance of a video frame with respect to the decoding process varies with type and position of the frame in the GOP. The highest level of importance is given to the I-frame in the GOP. If the I-frame is lost then the information carried by subsequent P frames are useless because they carry only the differential data of I-frame. If the P-frames are non-decodable then the B-frames derived from it will also be non-decodable. Hence, loss of one I frame will affect the complete GOP. The importance of a frame for the given GOP (see Figure 2.11) decreases in the following order: I, P_4, P_8, P_{12} and the remaining

B-frames have least importance. Hence the system should be content aware to provide an unequal error protection for video frames throughout the LTE network. In order to make the RAN content aware, the content information has to be shared with the RAN entities by the video server. In most of the research papers, the content awareness in network is achieved with the help of NAL. Retrieving the packet information from NAL by eNodeB will be an expensive task.

As mentioned in the Section 2.4.2, RLC supports three modes of which the UM is the preferred mode for video transmission. UM is better option for delay sensitive services such as video streaming. However, it does not support packet retransmission hence, this mode does not guarantee low packet-loss rate. AM mode instead guarantees a lower packet loss rate but this is achieved by expensive retransmission of all the lost packets. This retransmission causes a huge delay and jitter at the receiver. This huge delay involved in AM transmission makes this mode unsuitable for video services. UM mode of RLC does not provide error protection which results in loss of video quality, while AM provides equal error protection for all the transmitted packets causing large delay in transmission.

4.3 Proposed Scheme

In our proposed scheme, the eNodeB is made content aware of the video frame transmitted. A new operation mode (HM) has been proposed for RLC layer that uses content awareness to provide unequal error protection. Since LTE uses flat IP architecture, the DSCP field of IP header can be used by the video server to convey the video frame type transmitted to the eNodeB. The advantage of using DSCP field for identifying the packet content is that the processing time needed for packet identification will be less as we have to process only the header fields. The video server and eNodeB should agree upon a set of DSCP values to be used to identify the video frame types (see Table 4.1). Once the eNodeB is aware of the type of video frame scheduled, the unequal error protection (retransmission only for critical video frames) has to be applied on the video frames transmitted. This is done with new mode of operation in RLC called HM.

Table 4.1 DSCP values used in simulation for proposed scheme

Packet Type	IPv4 : DSCP value
I Frame	$(000111)_2$
P_4 Frame	$(001011)_2$
P_8 Frame	$(001111)_2$
P_{12} or all B Frames	$(010011)_2$

This RLC mode combines the benefits of both UM and AM. The main idea of this mode is to avoid unnecessary retransmission for least important frames and to ensure successful delivery of critical frames. The major problem in AM is that it treats all the frames in a fair way (which is not beneficial) leading to window stalling which results in further packet drop. In HM, the retransmission request is made only for the selected critical packets whose sequence number is shared with the receiving entity by the transmitting entity through a special control indication (IND) PDU. Similar to AM, HM PDU can carry either data or control information. The data and control status PDUs of HM use the same PDU format of AM (as shown in Figure 2.8). In addition to exiting STATUS PDU, a new control PDU called IND is defined. The format of this PDU is shown in Figure 4.1 and it is identified with a CPT value of $(001)_2$. The IND PDU informs the receiver about the sequence numbers of RLC PDUs that carry critical video frames. The receiving entity sends an explicit status PDU with ACK/NACK for the mentioned SN based on its reception status. The explicit STATUS PDU is sent by the receiver to inform the transmitter that it has received the IND PDU and has either received or lost the RLC data PDU. The ACK ensures the delivery of critical packets. If the transmission entity does not receive an explicit status message for the SN, it retransmits only the IND control PDU which is smaller in size compared to the actual data PDU until a maximum retransmission count is reached. Once the maximum retransmission count is reached, both data and control PDUs are discarded and transmission proceeds with the next SDU.

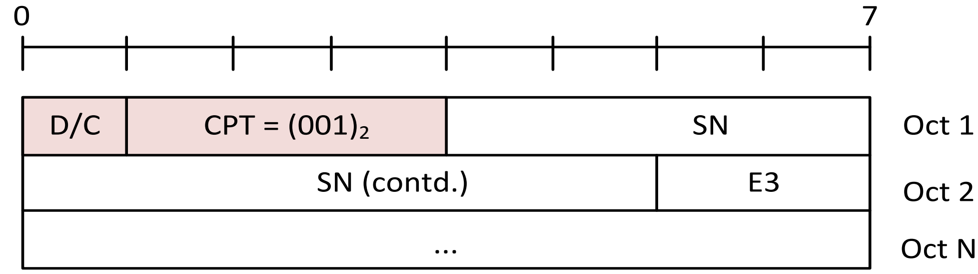


Figure 4.1 Proposed HM control IND PDU

The control IND PDU contains two fields, sequence number (SN) and an extension flag (E3). The SN is 10 bit long and gives the sequence number of the RLC data PDU that will carry critical content which needs explicit acknowledgement. The E3 flag is 2 bit long, out of four possible values, the value $(00)_2$ represents end of the control PDU, $(01)_2$ identifies the following field as SN. The E3 flag with value $(10)_2$ has a special meaning, this value indicates that the control PDU carries a range of sequence number. The SN before this flag marks the start of the range and the SN followed by this flag marks the end. This range option is defined to address the case where a sequence of RLC PDUs may carry critical packets. The value $(11)_2$ of E3 flag is reserved for future use. The description of E3 flag values are summarized in Table 4.2.

Table 4.2 E3 flag value description

Binary values of E3 Flag	Description
00	Marks end of IND PDU
01	Indicates following bytes as SN and E3 pair
10	Indicates the PDU contains range of SN. The SN before this flag is the starting SN and the bytes followed (SN & E3 pair) contains the ending SN of the range.
11	Reserved

The video server uses the algorithm VideoSrv mentioned below to identify video frames and fill the corresponding value of DSCP in the header field of IP packet transmitted to P-GW. The IP packet received at P-GW is transmitted to S-GW through the GPRS tunnel (see Figure 2.4). This tunnel further extends till the encapsulated IP packet reaches eNodeB. The GTP-U layer of eNodeB removes the GTP header and passes on the IP packet to PDCP layer for header compression. The PDCP layer identifies the packet type while parsing the IP header and labels the packet as either "IMP" or "OTH", where "IMP" represents I-frame, P_4 frame or P_8 frame and "OTH" (others) represent P_{12}-frame or B-frames. Since, the loss of I, P_4 or P_8 frames create more disruption in video quality, these frames are considered as critical frames.

Algorithm VideoSrv: To set DSCP value based on type of video frame

Identify the video frame type as I or P or B
Identify the position of the video frame in the GOP
if Frame type is I
    set DSCP value as $(000111)_2$
else if Frame type is P_4
    set DSCP value as $(001011)_2$
else if Frame type is P_8
    set DSCP value as $(001111)_2$
else if Frame type is P_{12} or B
    set DSCP value as $(010011)_2$
end if
The new DSCP value is used in the IPv4 header

The flow chart shown in Figure 4.2 shows the steps involved at eNodeB for downlink transmission using the proposed HM RLC. The SDUs of RLC carry the label set by the PDCP layer. The RLC transmission entity checks the label of each SDU. If the SDU is labeled as "OTH", the RLC layer creates the data PDU and transmits without having a copy of it in the transmission buffer as this PDU will not be retransmitted. If the SDU is identified as "IMP", then the RLC entity calculates the SN of RLC data PDU that will carry this SDU. The calculated SN is filled in the control IND PDU and transmitted before transmitting the RLC data PDU. The RLC data PDU containing the critical frame is copied to the transmission buffer and a timer for receiving status PDU is started. If the timer expires, the control IND PDU is retransmitted provided the retransmission count is within the maximum count. If RLC entity receives a NACK status PDU for the critical RLC PDU, the corresponding RLC data PDU is retransmitted. The transmission proceeds with next queued SDU if the RLC entity receives an ACK or if the maximum retransmission count exceeds.

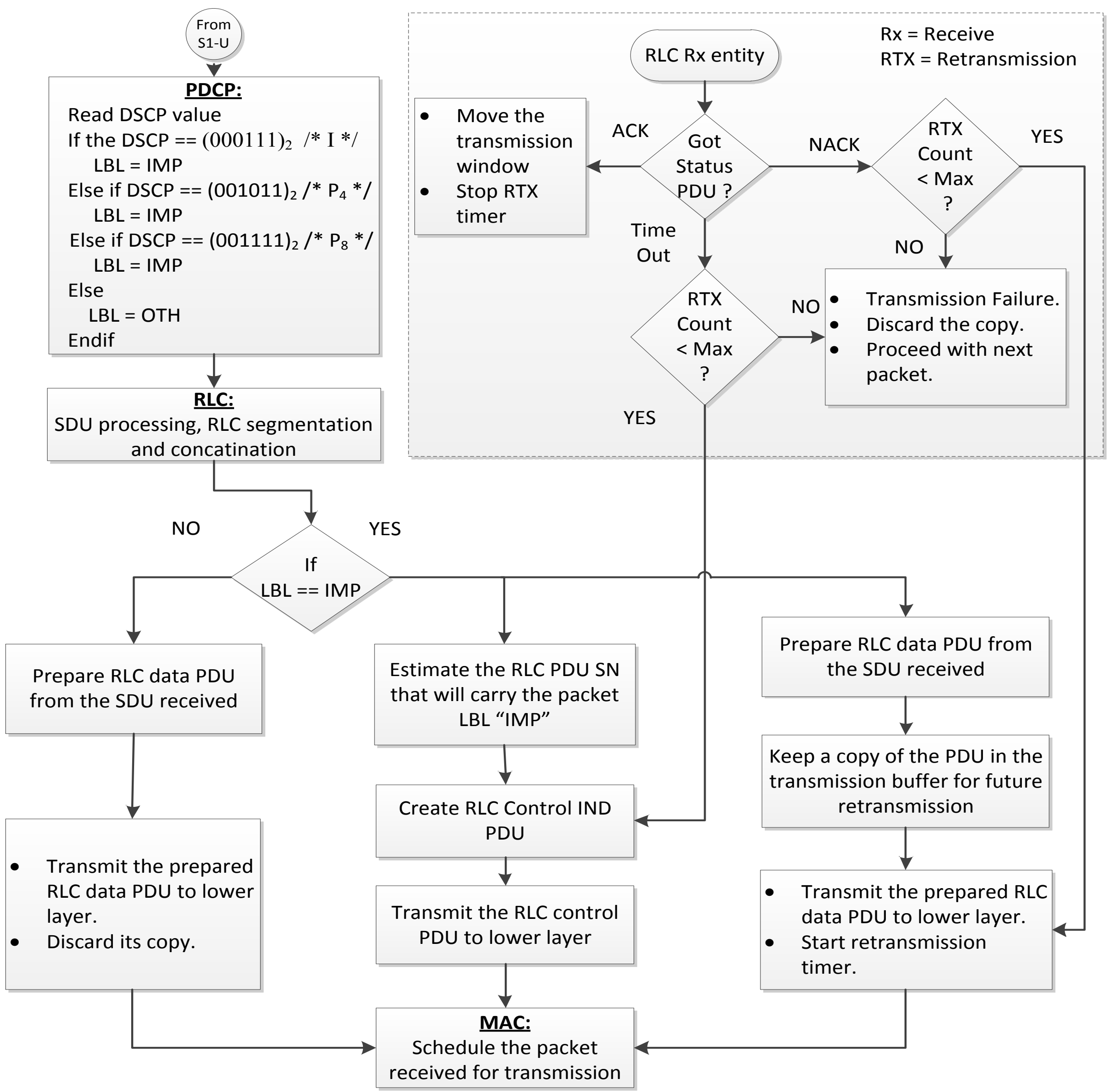


Figure 4.2 Operations of HM RLC at eNodeB

The flow chart shown in Figure 4.3 shows the steps involved at RLC receiving entity of UE which operates in the proposed HM mode. The RLC entity on receiving the control IND PDU parses and stores the SN transmitted by the sender. Later the RLC entity compares the SN of received RLC data PDU with the SN received in the control IND PDU. If the RLC receives the expected SN, it creates an explicit status PDU with an ACK to the SN received. The RLC entity ignores the missing PDU if the missing PDU SN is not the SN received in control IND PDU. If the missing PDU SN matches with the SN of control IND PDU, then the RLC entity creates an explicit status PDU with NACK. In case of non-continuous SN reception, the RLC PDUs are kept in reordering queue till the reordering timer expires. After the timer expires, the PDUs are reassembled and delivered to the upper layer.

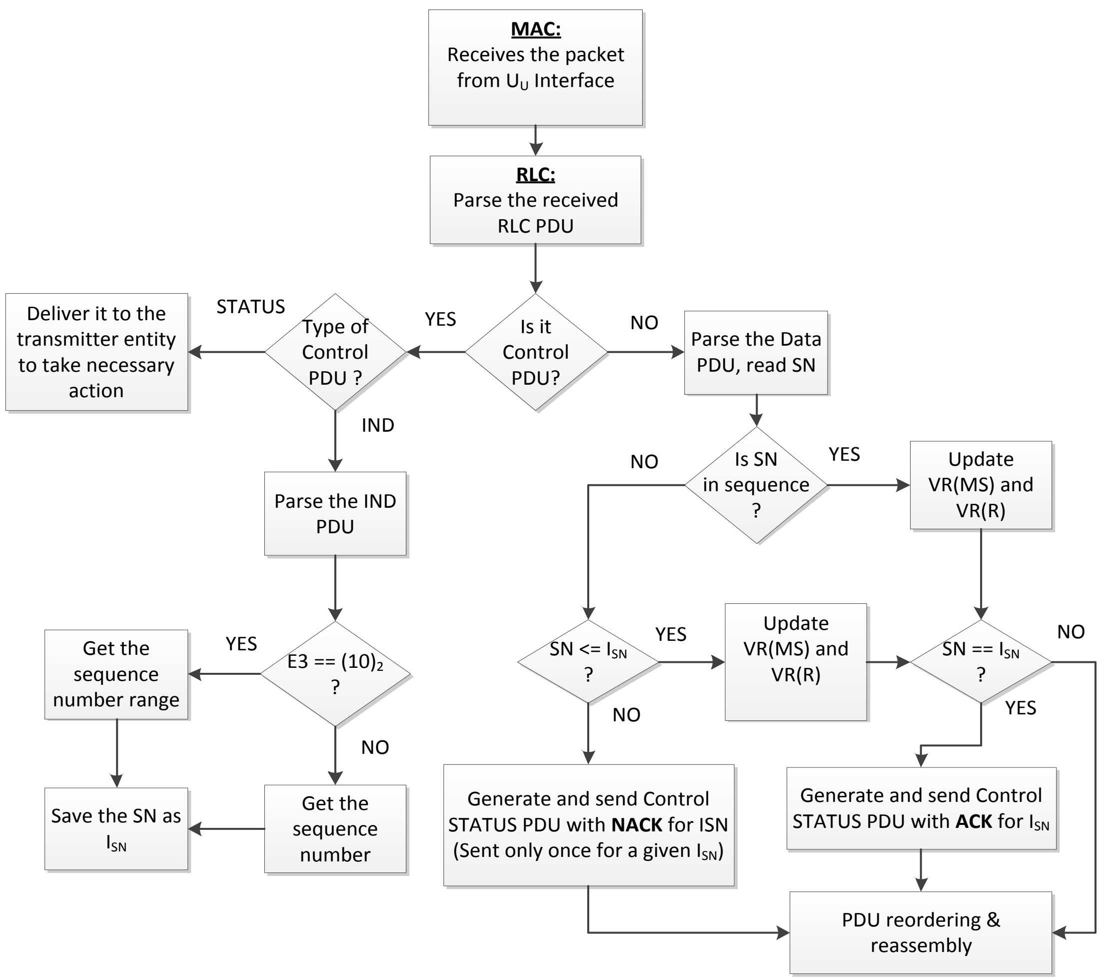


Figure 4.3 Operations of HM RLC at UE

4.4 Simulation Model

Validation of the proposed scheme has been done using simulation in OPNET network modeler [26]. For this simulation, a single cell LTE network with one user located at the cell edge region and a general purpose video server have been used. The video is streamed from the video server to the users (downlink). The H.264/AVC video trace of Star Wars IV [12] is used for simulation, and the properties of the video trace used are given in Table 3.2. The network model used for simulation is shown in Figure 4.4. To validate the scheme under practical wireless channel condition, the suburban Macrocell path loss model is used. For the worst case scenario, the user is placed at the cell edge where it will have bad channel condition and hence, it will be assigned a lower MCS value. The wired links between network entities are assumed to be error free. Important LTE system parameters used in the simulation are summarized in the Table 4.3 (extended list of parameters can be found in Appendix B).

Table 4.3 LTE system parameters

Parameters	Value
LTE Mode	FDD
eNodeB Operating Power	46 dBm
UE Antenna Gain	-1 dBi
Number of Antennas	1 (Tx / Rx)
LTE Bandwidth	20 MHz
Path Loss Model	Suburban Macrocell
Multipath Channel Model	ITU Vehicular B
Max. HARQ retransmission	Uplink – 3, Downlink - 3
RLC Mode	UM / AM / HM
Max. RLC retransmission	4
RLC reordering timer	35ms
Explicit status timer (HM)	20ms
Internet Protocol version	IPv4

To simulate the proposed scheme, code modifications have been done at video server, eNodeB and UE nodes of the modeler. The DSCP value of the IPv4 header which carries video frame from the video server is modified according to the type of video frame (I/P/B) being carried. The DSCP values used are chosen from the Pool 2 ranges [25] and assigned to the respective frames based on the values shown in Table 4.1. Based on the received DSCP values, the type and importance of a video frame is identified at the PDCP layer of eNodeB, and this information is propagated to the RLC layer.

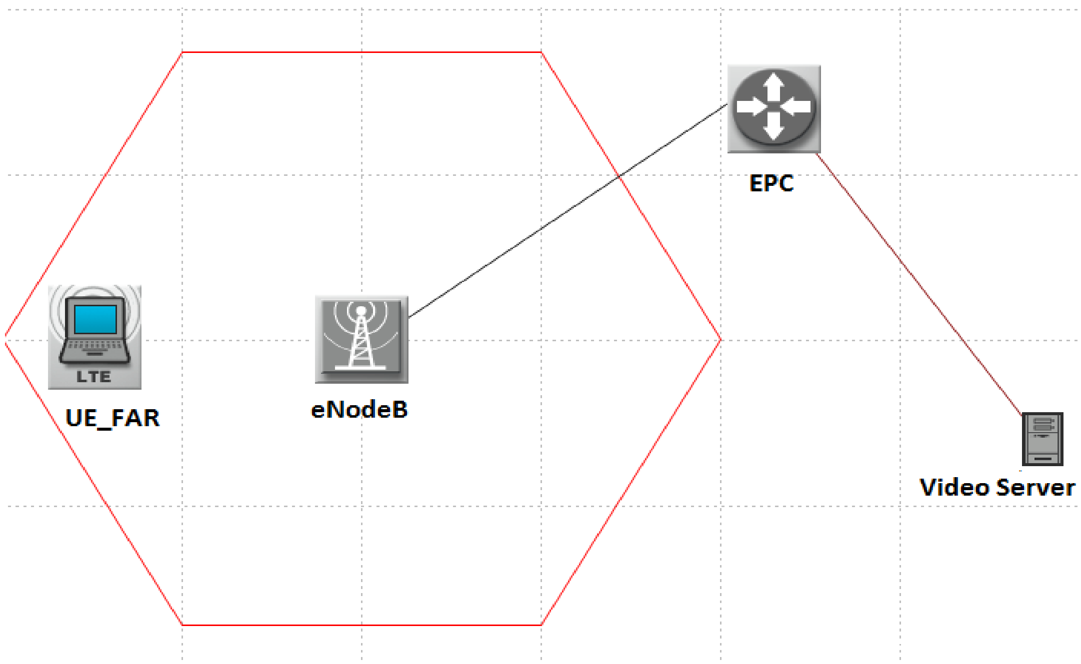


Figure 4.4 OPNET simulation model

As discussed in earlier chapter, loss of a P frame will affect all the B and P frames derived from it, whereas the loss of an I frame will affect the complete GOP and make it non-recoverable. The same MATLAB program used in Section 3.3 is used here as well to compute the total number of non-decodable frames from the received video frame trace. The quality of the received video (Q), is calculated using the formula (Equation 3.1) which is presented in Section 3.3.

4.5 Results and Analysis

The simulation has been conducted for video transmission over downlink channel for two operation modes of RLC (UM and AM) and with proposed mode of operation

(HM) under the same network conditions to make a better comparison. The results are compared and analyzed in this section. Due to suburban path loss and multipath channel model, the channel condition varies frequently; hence, some of the graphs are presented as time average for better analysis.

The main goal of the proposed mode is to reduce unnecessary retransmission for least important packets. As per the GOP pattern considered in this paper, the least important B frames whose loss will not affect other frames during the decoding process constitute 75% of the GOP sequence. The most important I frame constitute only 6% to GOP size. Since the legacy RLC AM does a fair retransmission for all packets, the probability of retransmission of the least important B frames whose loss can be tolerated is more. Hence, by avoiding the B-frames for retransmission we can save a huge bandwidth and time in transmission. This point is evident from the graph shown in Figure 4.5 which shows the retransmission rate of HM has drastically reduced by 95% in comparison with AM.

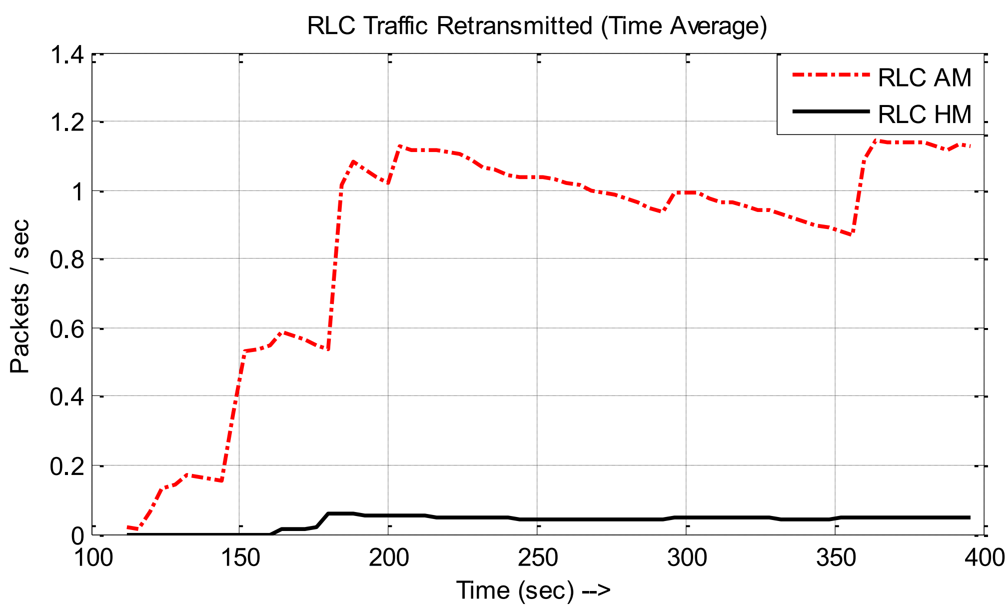


Figure 4.5 RLC downlink retransmission rate (eNodeB)

Figure 4.6 compares the delay involved in end-to-end transmission of video packets from videos server to UE for all three RLC operation modes. As AM performs retransmission for all the lost packets, the delay involved has increased quite large. The delay involved for AM is in the range of seconds which is not suitable for any real time streaming application. Hence, this mode can only be used for services that fall under background class of service, like file transfer. From the graph it is evident that the delay involved for HM has reduced at par to UM, this is because this mode does the retransmission only for the critical packet which are transmitted less frequent. There are few slight increases in delay when the critical packets are retransmitted for a successful delivery.

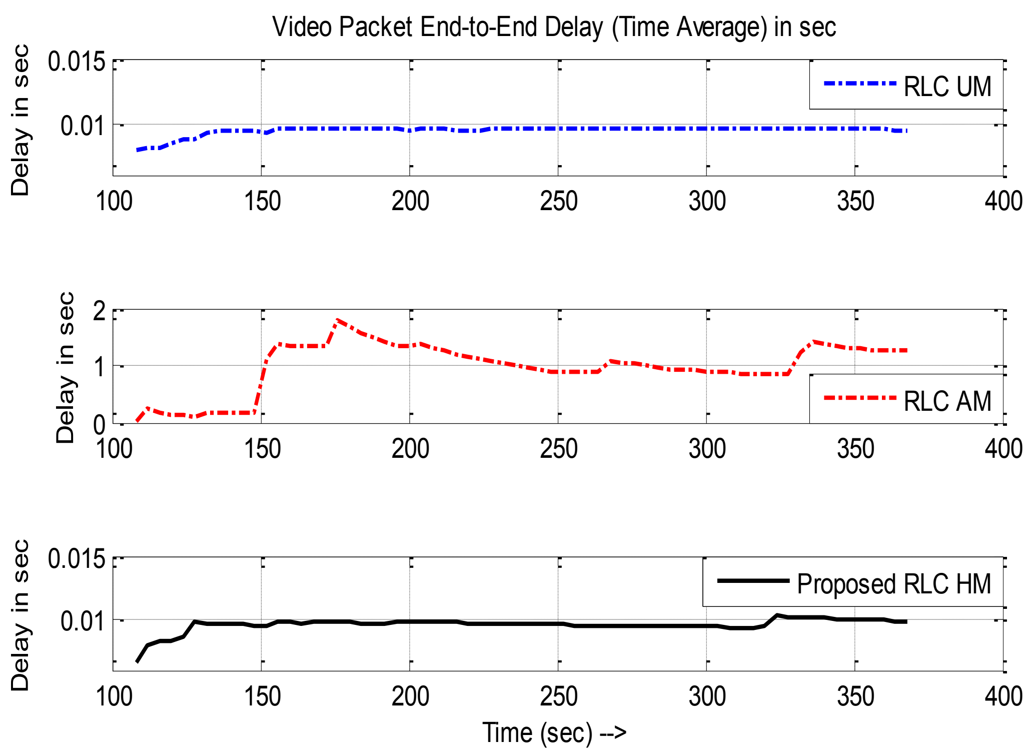


Figure 4.6 Downlink video packet end-to-end delay in sec

Figure 4.7 shows the graph of video traffic received at the UE shown in packets per second. It is evident from the graph that the UM and HM has similar characteristics whereas, AM shows huge variation in the traffic rate. Since AM involves huge delay in transmission, the video reception was not at constant rate. This variation in traffic rate of AM creates a big challenge in using this mode for streaming applications. However, by performing the retransmission only for the critical packet as in case of HM removes the huge variation in traffic and provides a constant video traffic.

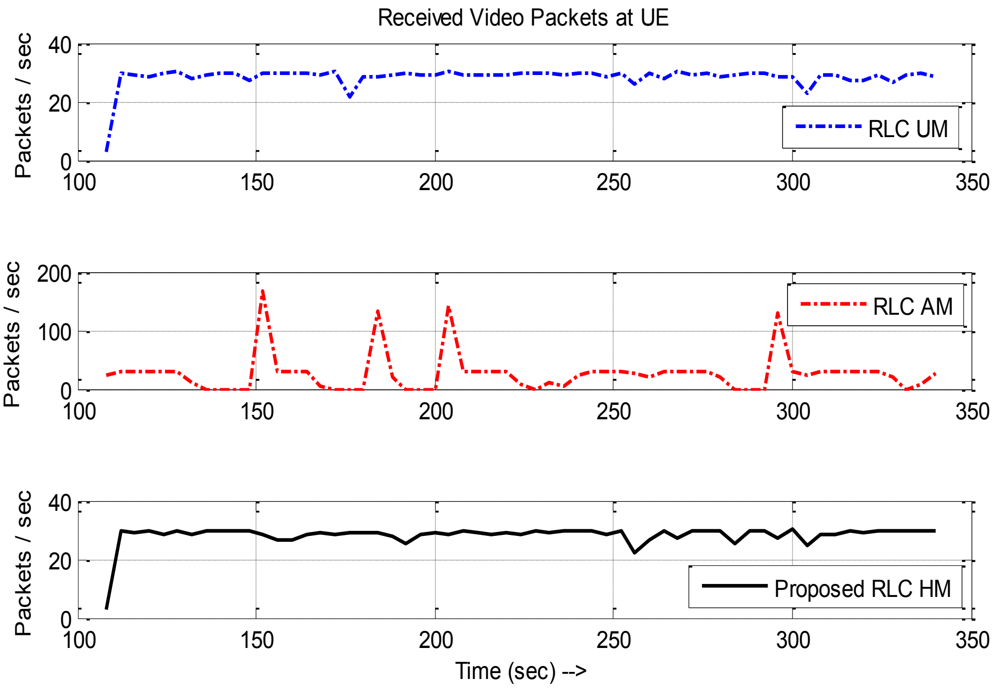


Figure 4.7 Downlink received video traffic

Figure 4.8 shows the quality of video decoded at the UE side based on the received video frames. The graph is drawn based on the value computed using Eqn. (3.1) at different time instances. Since AM provides the least packet loss rate when compared to the other two, the video quality of AM channel was the best of all. Since UM does not have the intelligent on the type of frame, some of the critical frames are lost thereby affecting the video quality resulting in reduction of received video quality. The HM strikes a balance between delay and quality, this mode provides a good quality video without compromising on the delay involved.

The overall summary of the video frame statistics is given in Table 4.4. There is a negligible increase in average end-to-end delay for HM in comparison to UM. However, it is compensated by an appreciable increase in received video quality by 7%. There is a huge reduction in end-to-end delay by 99% in HM in comparison with AM. Though AM video quality was good, the cost of delay involved was really high. Also the rate at which the packets received in AM is not constant, to overcome this problem a huge jitter buffer is needed at the receiving end which makes the AM mode impractical for video streaming in mobile devices.

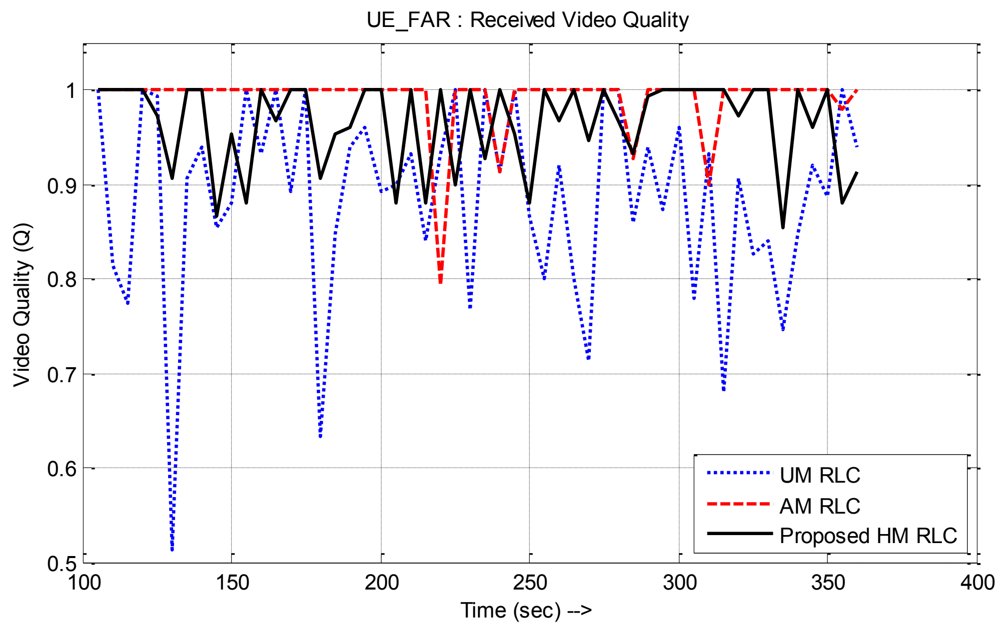


Figure 4.8 Calculated received video quality at UE

Table 4.4 Received packets summary

Metric		**UE Far (MCSAvg = 18)**		
		AM	**UM**	**HM**
Number of Received packets		7993	8350	7877
Number of lost frames	**I**	5	30	8
	P_4	6	18	6
	P_8	2	23	8
	P_{12}	3	17	15
	B (all)	52	183	170
Non decodable frames		73	905	278
Average end-to-end delay (in millisecond)		1254.2	9.4	9.8
Calculated video quality (Q) in percentage		99.07%	89.16 %	96.47%

4.6 Summary

This chapter discussed about introducing the content awareness in LTE system and using it at the RLC layer to provide unequal error protection for critical video packet thereby, improving the video quality. A new operation mode of RLC called HM has been proposed which combines the benefits of both AM and UM RLC. The simulation setup and the results obtained were also discussed in this chapter. The proposed mode reduced the end-end delay drastically in comparison with AM and gives a video quality improvement by 7% when compared with UM RLC. The benefit of this scheme is that it is compatible with 3GPP and IETF standards and provides a simple approach to achieve good quality video reception.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This thesis work is mainly focused on optimizing the radio access user plane protocols for efficient video transmission. Many research works have been published which concentrated on optimizing the network resource utilization for video transmission mostly in schedulers. Few researches have been carried out on content aware networks, where some adaptation schemes have been proposed by differentiating the types of video frames. However, the schemes proposed as part of this thesis work exploits the varying importance level within a frame type based on its position. The simulation results show that the existing system could not provide good video quality reception even with good channel conditions (only 93.4% video quality for UE_NEAR) due to the lack of content awareness. Hence, two schemes for improving quality of video reception in LTE network based on content awareness have been presented and discussed.

In CAAHR scheme (Chapter 3), the HARQ maximum retransmission count is dynamically varied in accordance with the importance level of the video frame transmitted. From the simulation results, it is noticeable that there is an increase by 6 to 8% in received video quality depending on the user position in the cell. The proposed scheme offers a noticeable improvement in the end user experience without compromising on the network performance. Moreover, with the proposed usage of DSCP for differentiating the video frames, wired networks can also prioritize and minimize packet loss for application critical packets.

In Chapter 4, we have proposed a new hybrid mode of operation (HM) for RLC. The simulation results indicate that the proposed mode is well suited for video transmission using RLC layer. This mode ensures that the delay caused by acknowledgement and retransmission for every packet in the legacy RLC AM mode is drastically reduced. This is achieved by acknowledging and retransmitting only the PDU which carries the critical video frame. Other PDUs are transmitted without any retransmission. Unlike UM mode, HM guarantees the delivery of the critical frames thereby increasing the received video quality by at least 7%. Both schemes are designed carefully in such a way that it can coexist with existing system without causing any performance degradation. Though the explanations of the schemes are made only for downlink transmission, they are also applicable to uplink transmission and TDD variant of LTE system.

5.2 Future Work

The schemes proposed and discussed in this thesis work are based on unicast video transmission over LTE network. Some of the recent services like mobile TV use multicast / broadcast for video transmission. Due to its huge number of users and demanding traffic, it becomes very important to impart content awareness to multicasting system as well for efficient transmission. Hence, as future work, adaptation of the proposed schemes for multicast / broadcast can be studied. Moreover, the proposed schemes are designed with respect to AVC video transmission. Since AVC is the base profile for its extended video codecs such as Scalable Video Coding (SVC) and Multi-view Video Coding (MVC), the proposed schemes can also be extended to support them.

References

[1] "Cisco Visual Networking Index: Global mobile Data Traffic Forecast Update, 2011-2016", Cisco White Paper, [Online], Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html, [Jun. 21, 2012].

[2] "Keyframes & Delta Frames Explained", [Online], Available: http://nickyguides.digital-digest.com/keyframes.htm, [Jul. 15, 2012].

[3] "LTE Technical Merits", [Online], Available: http://ambes-lte.blogspot.ca/2011/05/architecture-of-lte.html, [Jul. 15, 2012].

[4] "Radio Link Control (RLC)", [Online], Available: http://lteworld.org/specification/radio-link-control-rlc, [Jul. 18, 2012].

[5] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); LTE Physical layer; General description (Rel. 10)", 3GPP TS 36.201 V10.0, Dec. 2010.

[6] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Rel. 10)", 3GPP TS 36.213 V10.4.0, Dec. 2011.

[7] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Control (RLC) protocol specification (Release 10)", 3GPP TS 36.322 V10.0.0, Dec. 2010.

[8] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Medium Access Control (MAC) Protocol Specification (Rel. 5)", 3GPP TS 25.321 V5.10.0, Dec. 2004.

[9] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Rel. 10)", 3GPP TS 36.300 V10.6, Dec. 2011.

[10] 3GPP, “Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN) (Rel. 9)”, 3GPP TS 25.913 V9.0.0, Dec. 2009.

[11] I. Adhicandra, “Using AMC and HARQ to Optimize System Capacity and Application Delays in WiMAX Networks”, Journal of Telecommunications, 2(2): 15-20, 2010.

[12] AVC video traces: http://trace.eas.asu.edu/h264/index.html

[13] S. J. Bae, B. Choi, M. Y. Chung, "Delay-aware packet scheduling algorithm for multiple traffic classes in 3GPP LTE system", 17th Asia-Pacific Conference on Communications (APCC), pp. 33-37, 2011.

[14] S. Benayoune, N. Achir, K. Boussetta, K. Chen, "Content-Aware ARQ for H.264 Streaming in UTRAN," IEEE Wireless Communications and Networking Conf. (WCNC), pp. 1956-1961, 2008.

[15] A.M. Cipriano, P. Gagneur, G. Vivier, S. Sezginer, "Overview of ARQ and HARQ in beyond 3G systems," IEEE 21st Int. Symposium on Personal, Indoor and Mobile Radio Communications, pp. 424-429, 2010.

[16] S. Corner, “Ipv6 traffic volumes going backwards”, [Online], Available: http://www.itwire.com/business-it-news/networking/46689-ipv6-traffic-volumes-going-backwards [Jan. 15, 2012].

[17] P. Frenger, S. Parkvall, E. Dahlman, "Performance comparison of HARQ with Chase combining and incremental redundancy for HSDPA", IEEE 54th Vehicular Technology Conference, 3: 1829-1833, 2001.

[18] S. Karachontzitis, T. Dagiuklas, L. Dounis, "Novel cross-layer scheme for video transmission over LTE-based wireless systems", IEEE Int. Conf. on Multimedia and Expo (ICME), pp. 1-6, 2011.

[19] F. Li, G. Liu, "A Novel Marker System for Real-Time H.264 Video Delivery Over Diffserv Networks", IEEE Int. Conf. on Multimedia and Expo, pp. 2142-2145, 2007.

[20] C.H. Lin, C.H. Ke, C.K. Shieh, N.K. Chilamkurti, “The Packet Loss Effect on MPEG Video Transmission in Wireless Networks”, Int. Conf. on Advanced Information Networking and Applications, pp. 565-572, 2006.

[21] M. Lu, P. Steenkiste, T. Chen, "CHARQ: Cooperative Hybrid ARQ for wireless video streaming", IEEE Int. Conf. on Multimedia and Expo, pp. 750-753, 2009.

[22] H. Luo, S. Ci, D. Wu, J. Wu, H. Tang, "Quality-driven cross-layer optimized video delivery over LTE", IEEE Communications Magazine, 48(2): 102-109, 2010.

[23] MATLAB: http://www.mathworks.com/products/matlab

[24] M. Mazzotti, E. Paolini, M. Chiani, B. Gadat, C. Bergeron, R. Fracchia, "Analysis of Packet-Level Forward Error Correction for Video Transmission", 73rd IEEE Vehicular Technology Conference, pp. 1-5, 2011.

[25] K. Nichols, S. Blake, F. Backer and D. Black, "Definition of the differentiated services field (DS field) in the IPv4 and IPv6 Headers", RFC 2474, Dec. 1998.

[26] OPNET LTE Specialized Model: http://www.opnet.com/LTE

[27] M. Plumb, "Fantastic 4G - Hundreds of telecoms will invest in 4G LTE networks in 2012", IEEE Spectrum Magazine, Jan. 2012.

[28] J. Postel, Editor, "Internet Protocol", STD 5, RFC 791, Sep. 1981.

[29] R. Radhakrishnan, A. Nayak, "Cross Layer Design for Efficient Video Streaming over LTE Using Scalable Video Coding", IEEE ICC Workshop on Smart Communication Protocols and Algorithms, 2012.

[30] R. Radhakrishnan, B. Tirouvengadam, A. Nayak, "Channel Quality-Based AMC and Smart Scheduling Scheme for SVC Video Transmission in LTE MBSFN Networks", 2nd IEEE Int. Workshop on Smart Communication Protocols and Algorithms, 2012.

[31] P. Seeling, M. Reisslein, "Video Transport Evaluation with H.264 Video Traces", IEEE Communications Surveys & Tutorials, 99: 1-24, 2011.

[32] S. Sesia, I. Toufik, M. Baker, "LTE – The UMTS Long Term Evolution From Theory to Practice", 2nd ed., United Kingdom: Wiley, 2011.

[33] Y. Shan, "Cross-layer techniques for adaptive video streaming over wireless networks", EURASIP J. Appl. Signal Process, 1: 220-228, 2005.

[34] M. Shehada, S. Thakolsri, Z. Despotovic, W. Kellerer, "QoE-based Cross-Layer Optimization for video delivery in Long Term Evolution mobile networks", 14th

Int. Symposium on Wireless Personal Multimedia Communications (WPMC), pp. 1-5, 2011.

[35] S. Soundararajan, P. Agrawal, L. Yihan, "An efficient HARQ retransmission algorithm in OFDMA based wireless networks", 41st Southeastern Symposium on System Theory, pp. 88-93, 2009.

[36] T. Stockhammer, “System and Cross-Layer Design for Mobile Video Transmission”, Ph.D. thesis, Munich University of Technology, Munich, 2008, Available: http://mediatum2.ub.tum.de/node?id=652900, [Dec. 2011].

[37] G. J. Sullivan, P. Topiwala, A. Luthra, “The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions,” SPIE Annual Conf. Apps. of Digital Image Processing XXVII, pp. 454–74, 2004.

[38] B. Tirouvengadam, A. Nayak, “Hybrid Mode Radio Link Control for Efficient Video Transmission Over 4G/LTE Network”, IEEE 18th Int. Conf. on Parallel and Distributed Systems, 2012 (to appear).

[39] B. Tirouvengadam, R. Radhakrishnan, A. Nayak, “CAAHR: Content Aware Adaptive HARQ Retransmission Scheme for 4G/LTE Network”, Int. Conf. on Ubiquitous and Future Networks (ICUFN), 2012.

[40] S. Tripathi, V. Kulkarni, A. Kumar, “LTE E-UTRAN and its Access Side Protocols’, White paper, RadiSys [Online], Available: http://www.ccpu.com/articles/2010/lte-e-utran-access-side-protocols, [Jul. 3, 2012].

[41] S. Tomazic, G. Jakus, "Long term evolution: Towards 4th generation of mobile telephony and beyond", 9th Int. Conf. on Telecommunication in Modern Satellite, Cable, and Broadcasting Services, pp. 91-96, 2009.

[42] J. Wang, M. Venkatachalam, Y. Fang, “System Architecture and Cross-Layer Optimization of Video Broadcast over WiMAX”, IEEE Journal on Selected Areas in Communications, 25 (4):712 - 721, 2007.

[43] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Trans. on Circuits Systems for Video Technology, 13(7): 560–576, 2003.

[44] Wikipedia, "H.264/MPEG-4 AVC", [Online], Available: http://en.wikipedia.org/wiki/H.264/MPEG-4_AVC, [Jul. 27, 2012]

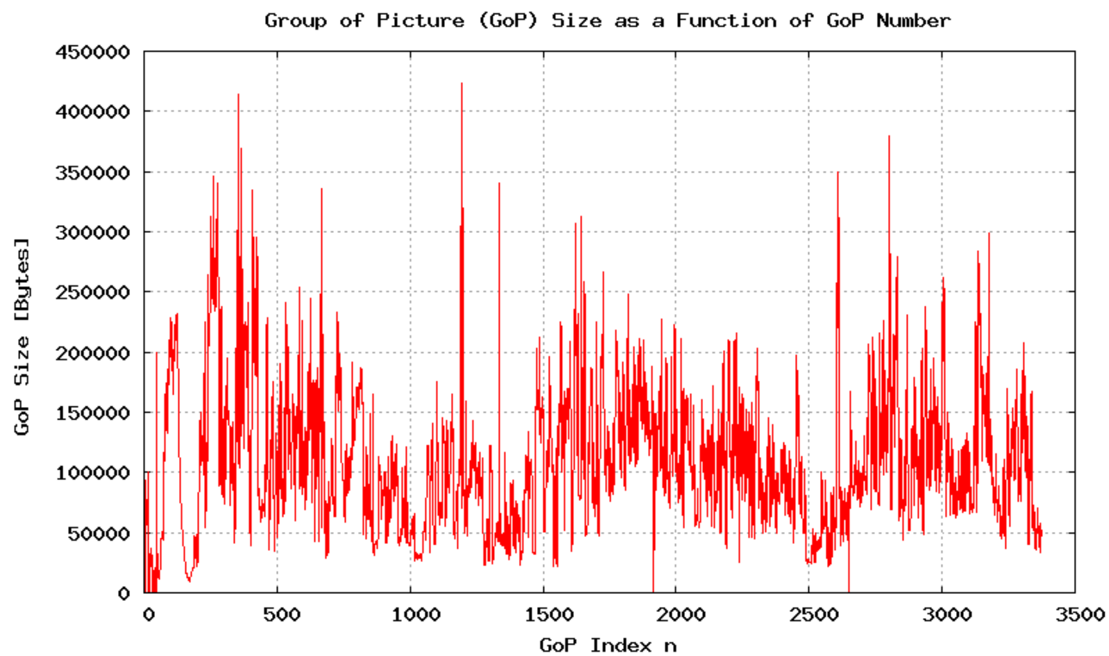
[45] Wikipedia, "Internet Protocol", [Online], Available: http://en.wikipedia.org/wiki/Internet_Protocol, [Jul. 23, 2012]

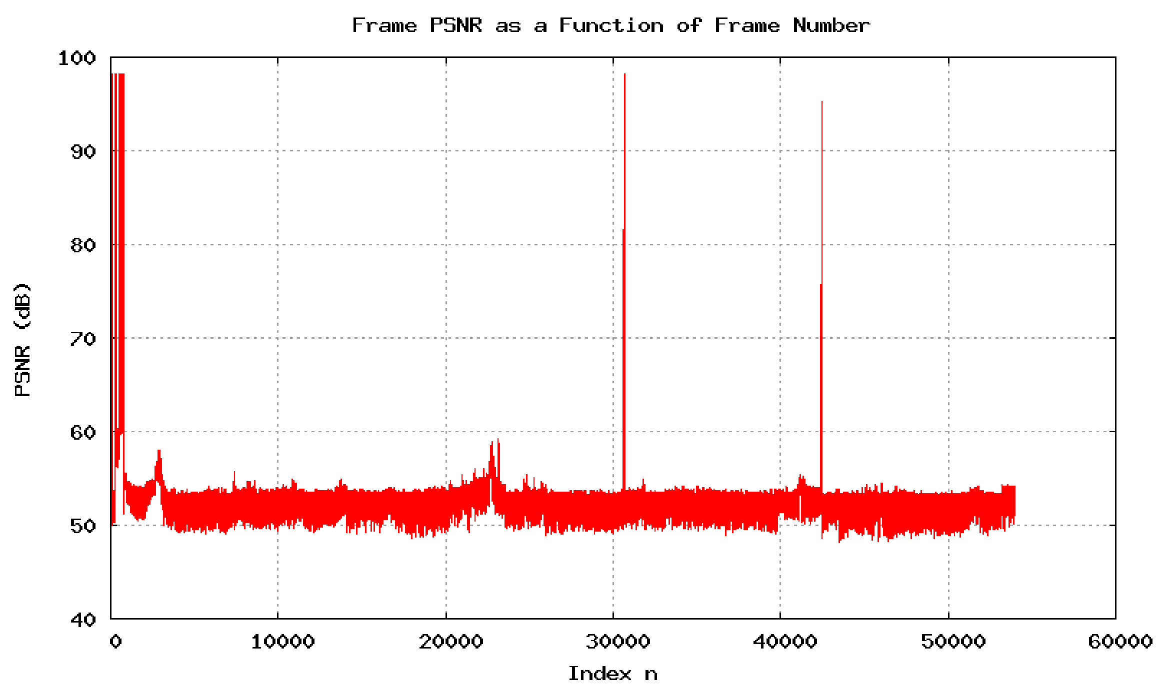
[46] Wikipedia, "IPv4", [Online], Available: http://en.wikipedia.org/wiki/IPv4, [Jul. 23, 2012]

[47] Wikipedia, "System Architecture Evolution", [Online], Available: http://en.wikipedia.org/wiki/System_Architecture_Evolution, [Jul. 23, 2012]

[48] Wikipedia, "Video compression picture types", [Online], Available: http://en.wikipedia.org/wiki/Video_compression_picture_types, [Jul. 27, 2012]

Appendix A: Star Wars IV [12] - AVC Video Characteristics





Appendix B: OPNET Modeler Simulation Parameters

Parameters	Value
LTE Mode	FDD
LTE Bandwidth	20 MHz
Base Frequency	2110 MHz
Cyclic prefix type	Normal (7 symbols per slot)
eNodeB failure / recovery modeling	Disabled
CQI – periodic configuration index	40
PHICH error probability	Disabled
Scheduling Mode	Link Adaptation and Channel Dependent Scheduling
Path Loss Model	Suburban Macrocell
Multipath Channel Model	ITU Vehicular B
HARQ retransmission improvement factor	2
PDCP compression	Disabled
RLC – poll retransmit timer	45ms
GTP processing delay	IP datagram forwarding rate based
Internet Protocol version	IPv4
MTU	1500
Application type	Video conferencing
Frame interval time information	30 frames/sec