Masters in Informatics Engineering
Master Thesis
Final Report

Mechanisms for resilient video transmission in wireless networks

Pedro Borges
pborges@student.dei.uc.pt

Supervisors:
Marília Curado
Roger Immich

Date: 29th of June, 2014

FCTUC DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Acknowledgements

Thank you to all my friends and colleagues who helped make this thesis a reality.

I express my gratitude to David Palma and Roger Immich which during these years in the LCT provided me my first contact to the world of research in wireless networks, protocols, network simulation and quality metrics. Both of you gave me the background that made the realization of this work possible.

I also thank Roger Immich which provided crucial mentoring during the course of this thesis. You offered me great suggestions and encouragement which helped me keep pushing when difficulties arose. Your insights were critical in achieving the significant final results. Without your invaluable help the development of these mechanisms would not be possible.

Thank you David Nunes for the productive discussions and encouragement even with this not being your work area.

A special thanks goes to Marília Curado, my advisor, which introduced to the LCT and believed in the work I could do. Your comments and suggestions were always helpful, showing themselves essential to the realization of this thesis. You were always available and supportive when there were difficulties. Thank you so much.

Thanks to my parents for encouraging me throughout my academic career and in my achievements.

Thank you Cristiana's parents, you were always supportive, encouraging and enthusiastic to my successes. Your help and support is very meaningful to me.

Thank you Cristiana for your amazing support, encouragement, cheerfulness, enthusiasm, and precious advice helping me to become better at what I do. Your research work inspired and motivated me to strive and do meaningful work like you do. You are an amazing person and my deepest love goes to you.

Thanks everyone. Without your help this work wouldn't be possible.

Abstract

Remote video viewing is largely growing, especially in wireless devices. Therefore, video transmission mechanisms such as retransmission and Forward Error Correction (FEC) are very important to ensure that the video arrives at its destination while retaining its quality. While some of these mechanisms can obtain good performance, this does not mean they provide an optimal situation. There are several factors that negatively impact Quality of Experience (QoE) in video transmission, such as network congestion, loss and type of video being transmitted. Adaptive Forward Error Correction schemes aim at providing protection against loss errors by analysing several relevant characteristics of the video and/or network, and using them to apply unequal error protection (UEP) to different parts of the video. Taking this into account, it is clear that an adaptive FEC mechanism is a viable option to achieve optimal error protection, thus increasing the video transmission's resilience against errors and losses, and therefore improving Quality of Experience for the end user.

This work presents and examines the state-of-the-art mechanisms for retransmission, FEC, QoE, video classification and algorithm optimization. The assessment of the mechanisms presents an in-depth study of each concept and individually explores their advantages and disadvantages. Three mechanisms which provide adaptive FEC are developed in and evaluated in a simulation environment, with motion classification, error correction, adaptive redundancy allocation and loss feedback components. The capabilities of the adaptive FEC mechanisms are evaluated through a wireless network environment in order to assess their performance against other methods.

The results obtained through the assessed mechanisms showed incremental improvements through each stage of development. They also showed that it is possible to accurately characterize the intensity of motion in video sequences according to their characteristics. Furthermore, the findings highlight the importance of shielding sequences with higher amounts of motion intensity with greater quantities of redundancy due to david their higher degradation when subjected to error. The mechanisms achieved a considerable reduction in the amount of redundancy used to shield the data being transmitted, particularly the loss prediction mechanism. At the same time, the mechanisms maintained video quality which combined with the overhead reduction results in an overall improvement of QoE ultimately enhancing video transmission in wireless networks.

Keywords: Adaptive FEC, Wireless Video Transmission, Loss Prediction, Quality of Experience, Motion Intensity

Contents

List of Figures

List of Tables

Abbreviations

ACFEC	**A**daptive **C**ross-Layer **F**orward **E**rror **C**orrection
ACO	**A**nt **C**olony **O**ptimization
AHECM	**A**daptive **H**ybrid **E**rror **C**orrection **M**odel
ANN	**A**rtificial **N**eural **N**etwork
APB-FEC	**A**daptive **P**acket and **B**lock length **F**orward **E**rror **C**orrection
ARQ	**A**utomatic **R**epeat **R**e**Q**uest
AVC	**A**dvanced **V**ideo **C**oding
BLER	**B**lock **E**rror **R**ate
BOP	**B**lock **O**f **P**ackets
CAKE	**C**luster **A**analysis **K**nowledge Bas**E**
CD	**C**ompact **D**isc
CLAM	**C**ross-**LA**yer infor**M**ation
CMSE	**C**omulative **M**ean **S**quared **E**rror
CPU	**C**entral **P**rocessing **U**nit
CRC	**C**yclic **R**edundancy **C**heck
CSF	**C**ontrast **S**ensitivity **F**unction
DSIS	**D**ouble-**S**timulus **I**mpairment**s**cale
DSCQS	**D**ouble-**S**timulus **C**ontinuous**Q**uality-**S**cale
DVD	**D**igital **V**ersatile **D**isc
DVB-S	**D**igital **V**ideo **B**roadcasting -**S**atellite
DVQ	**D**igital **V**ideo **Q**uality
EBU	**E**uropean **B**roadcasting **U**nion
ECC	**E**rror **C**orrecting **C**ode
EEP	**E**qual **E**rror **P**rotection

FEC	**F**orward **E**rror **C**orrection
FR	**F**ull-**R**eference
GoP	**G**roup of **P**ictures
GT	**G**ame of **T**heory
HARQ	**H**ybrid **A**utomatic **R**epeat **R**e**Q**uest
HSPA	**H**igh **S**peed **P**acket **A**ccess
HT	**H**igh **T**roughput
HVS	**H**uman **V**isual **S**ystem
HyQoE	**Hy**brid **Q**uality **o**f **E**xperience
IEEE	**I**nstitute of **E**lectrical and **E**lectronics **E**ngineers
ITU-T	**I**nternational **T**elegraph **U**nion **T**elecommunication **S**tandardization **S**ector
JND	**J**ust-**N**oticeable **D**istortion
LC	**L**ocal **C**ontrast
LDPC	**L**ow **D**ensity **P**arity **C**heck
LT	**L**uby **T**ransform
LTE	**L**ong **T**erm **E**volution
MAC	**M**edia **A**ccess **C**ontrol
MBMS	**M**ultimedia **B**roadcast **M**ulticast **S**ervices
MOS	**M**ean **O**pinion **S**core
MPEG	**M**oving **P**ictures **E**xpert **G**roup
MSE	**M**ean **S**quared **E**rror
NE	**N**ash **E**quilibrium
NR	**N**o-**R**eference
NS-3	**N**etwork **S**imulator **3**
OLSR	**O**ptimised **L**ink **S**tate **R**outing
PCA	**P**rincipal **C**omponent **A**nalysis
PET	**P**riority **E**ncoding **T**ransmission
PHY	**P**hysical **L**ayer of the OSI model
PSNR	**P**eak **S**ignal to **N**oise **R**atio
PSQA	**P**seudo **S**ubjective to **Q**uality **A**ssessment
PVQM	**P**erceptual **V**ideo to **Q**uality **M**etric

QoE	**Q**uality of **E**xperience
QoS	**Q**uality of **S**ervice
RBF	**R**adial **B**asis **F**unction
RCPR	**R**ate-**C**ompatible **P**unctured **C**onvolutional
RF	**R**andom **F**ield
RNN	**R**andom **N**eural **N**etwork
RR	**R**educed-**R**eference
RS	**R**eed **S**olomon
RSC	**R**ecursive **S**ystematic **C**onvolutionary Code
RTP	**R**eal-time **T**ransport **P**rotocol
RTSP	**R**eal **T**ime **S**treaming **P**rotocol
RTT	**R**ound-**T**rip **T**ime
SCSF	**S**patial **C**ontrast **S**ensitivity **F**unction
SDSCE	**S**imultaneous **D**ouble **S**timulus for **C**ontinuous **E**valuation
SNR	**S**ignal-to-**N**oise **R**atio
SS	**S**ingle-**S**timulus
SSCQE	**S**ingle **S**timulus **C**ontinuous **Q**uality **E**valuation
SSIM	**S**tructural **SIM**ilarity
UDP	**U**ser **D**atagram **P**rotocol
UEP	**U**nequal **E**rror **P**rotection
VA	**V**isual **A**ttention
VaEEP	**V**ideo-**a**ware **E**qual **E**rror **P**rotection
VaUEP	**V**ideo-**a**ware **U**nequal **E**rror **P**rotection
VIF	**V**ideo **I**nformation **F**idelity
ViewFEC	**V**id**E**o a**W**are **F**orward **E**rror **C**orrection
VQM	**V**ideo **Q**uality **M**etric
VSNR	**V**isual **S**ignal-to-**N**oise **R**atio
WiMAX	**W**orldwide **I**nteroperability for **M**icrowave **A**ccess

Chapter 1

Introduction

This work presents the development of an adaptive Forward Error Correction mechanism for resilient video transmission over wireless networks. Through the analysis of the state-of-the-art of video transmission and error correction techniques, the drawbacks and advantages of the current mechanisms are identified.

The first section presents the motivation for the work developed, while the second and third section describe the objectives and contributions, respectively. In the last section the organization of the thesis is explained.

1.1 Motivation

The use of real-time services over multi-hop wireless and mobile networks, more specifically video streaming, has seen a dramatical increase in the last few years (comScore, 2011). In Europe alone, the growth of mobile video viewing was of 162% in just one year, topping of at 48.1 million unique viewers (comScore, 2013). The video streaming service can be used in varied ways and in several types of devices, such as laptops connected to a wireless networks or mobile phones connected to 3G or 4G type cellular networks. It is a means for professionals from all kinds of companies to arrange remote meetings and to improve collaboration amongst workers, therefore helping to reduce all travel related costs (i.e., time and money). Home users also make use of this service to create, share and view multimedia content, no matter if from a cabled or from a wireless connection. The key point here, is that real-time video streaming is now an integrant part of our everyday lifestyle.

It is known that network resources are not unlimited, and therefore, many factors can hinder the transmission by causing packets to be corrupted or lost, thus degrading video quality and negatively impacting the end user experience. Some of these factors are internal such as the video codec type, the content of the video and the bitrate. Then, there are the external factors like propagation and path loss, node interference, channel noise, shadowing and congestion. To mitigate the effect of the above on the quality of the transmitted video over a wireless network, we need to employ mechanisms that combat this loss/corruption of information. There are already three types of techniques that are commonly used. These are retransmission by means of Automatic Repeat reQuest (ARQ), Forward Error Correction (FEC) and Hybrid solutions.

One of the greatest problems faced with these mechanisms is that most of them are rigid in the way they perform. In the case of retransmission, packets are usually discarded at the receiver and the retransmission process involves the consequent retransmission of the packet, wasting network resources and interfering with the temporality of the service in the case of real-time services. Concerning FEC mechanisms, the main issue is that some of them offer fixed allocation of the protection/redundancy budget or have some type of adaptation in the form of pre-processing. Also, these schemes tend not to take into account network conditions for redundancy allocation. This negatively impacts the quality on the receiver's end. Furthermore, this strict fixed use of redundancy is not optimal in terms of network overhead. Such approaches can lead to further network congestion, worsening network performance and the packet drop rate, and thus negatively impacting the received video quality.

1.2 Objectives

The goals of the work presented in this thesis aim at overcoming the current limitations of the existent mechanisms, such as the staticity of the existing FEC mechanisms, and the inability to take into consideration the video's motion parameters, which are crucial to the quality of experience for the end user. This innovative proposed scheme, solves these problems by adaptively selecting the amount of redundancy given to individual frames, through the analysis of their type and their motion characteristics. The adaptive protection of the most important parts of the video causes reduction in network overhead, which in turn, increases the received video quality when compared to a fixed redundancy scheme. This way video files which consist of scenes with a great amount of movement will have their respective critical frames better protected against loss by the means of added redundancy.

1.3 Contributions

This thesis provides a general overview of several types of Error Correcting Codes, such as Linear Block Codes, Convolutional Codes and Fountain Codes. Afterwards, a series of mechanisms which use FEC are discussed through a detailed analysis that points out their strengths and weaknesses.

Concerning the topic of video classification according to motion intensity, the author performed an analysis of several schemes, such as Cluster Analysis, Principal Component Analysis and Neural Networks, exploring the concept of Neural Networks with more depth in the case of Random Neural Networks.

In the subject related to optimization, the analysed mechanisms were those of Game Theory, Metaheuristics and a specific case of Metaheuristics that is Ant Colony Optimization.

The mechanism presented in this thesis performs video transmission shielding in wireless environments. It does so through the adaptive use of redundancy in order to mitigate the effects of packet losses. It is based on the modular concept which is presented in a previous study (Immich et al., 2013). The mechanism differs from the aforementioned one in the sense that it uses a Random Neural Network (RNN) to perform motion intensity classification of the video frames, and uses a Metaheuristic approach to adaptively perform the allocation of redundancy. This mechanism improves on the base concept by taking into account network loss information. This information is used to characterize the error bursts present in the network, in order to protect the data being transmitted with even more accuracy than a mechanism which uses only motion intensity to compute the allocation of redundancy. The mechanism went through various incremental development stages which led up to the final iteration which is comprised by the Loss Prediction mechanism.

- The author also developed the neuralFEC mechanism presented in Chapter 5 which is a motion intensity categorization module. The neuralFEC mechanism is based on a Random Neural Network. From several characteristics of a video sequence, the RNN performs a numeric categorization of the current frame in terms of motion intensity. To do so, a RNN was developed from scratch with the help of RNNSimv2 MATLAB module. The RNN implementation present in RNNSIMv2 was thoroughly experimented with and resulted in a stable system suitable for motion intensity classification. The resulting RNN was trained and validated with the characteristics of several video sequences in the MATLAB environment. The results of the RNN were injected onto the Evalvid trace files corresponding to different videos which are used for evaluation of the mechanism in NS-3. From there, a simple selection mechanism uses an Error Correcting Code to perform redundancy allocation according to the degree of motion intensity present in a given frame.

- The AntMind mechanism presented in Chapter 6 was developed by the author and is an enhancement over neuralFEC. The AntMind mechanism performs dynamic redundancy allocation based on ACO and an Error Correcting Code. It takes into account several characteristics of the video, information retrieved from neuralFec about motion intensity and the instantaneous network loss rate. The ACO module was developed based on the libaco C++ library which is comprised of several ant-based problem solving systems. The several ant systems were explored and the simple ant system was selected for use and integration with Network Simulator 3 (NS-3). The libaco implementation developed to solve the travelling salesman problem was modified to work with the requirements and particularities presented by our problem. Finally, a path graph which describes the several characteristics of the video and network loss state was designed, in order to enable the use of ACO to solve our redundancy allocation problem.

- The author developed the Loss Prediction mechanism presented in Chapter 6 which is an enhancement over AntMind combining the use of the ACO dynamic redundancy allocation module with a module which performs error prediction. The error prediction module was developed from scratch based on the concept of good and bad gaps presented in (Karner et al., 2007). A feedback mechanism was implemented to enable the retrieval of loss statistics. These statistics provide information on the occurrence of errors based on the size of the gaps of the already received frames. From this, the error prediction mechanism estimates if losses will occur in the next block of packets to be transmitted. Therefore, the error prediction mechanism has an influence on ACO leading to an adjustment in redundancy based on the prediction of the occurrence/non-occurrence of an error.

- A detailed study of several Error Correcting Codes was performed by the author to assess which was the Error Correcting code suitable for use on the devised transmission scenario. The Reed-Solomon code was selected due to its simplicity and flexibility. In order to use Error Correcting Codes in video transmission in a simulation environment such as NS-3, an external C++ library was used. The IT++ library and its Reed-Solomon code were modified and integrated with the simulation scenario developed in NS-3.

Throughout the development of the mechanisms, several publications and presentations took place:

- The development and experimental evaluation performed by the author of the neuralFEC module resulted in a paper in collaboration with the co-supervisor. The paper is titled *Adaptive Motion-aware FEC-based Mechanism to Ensure Video Transmission* included in Appendix A which was presented at The Nineteenth IEEE Symposium on Computers and Communications (ISCC 2014) (Immich et al., 2014a).

- The development and experimental evaluation of the AntMind mechanism was solely performed by the author. It resulted in a paper named *AntMind: Enhancing Error Protection for Video Streaming in Wireless Networks* included in Appendix A which was presented at The Fifth International Conference on Smart Communications in Network Technologies (SaCoNet 2014) (Immich et al., 2014b).

- Based on the work presented in this thesis, a presentation by the author describing the mechanism in general and preliminary results with a RNN was conducted at the Rede Temática de Comunicações Móveis (RTCM) Seminar.

1.4 Structure

This document is structured as follows:

The second chapter presents techniques that improve video transmission which are used in wireless networks, namely retransmission, FEC and error correcting codes, and hybrid. Chapter three introduces Quality of Experience metrics used for quality assessment of the video transmitted over the network. In chapter four, mechanisms for classification and grouping of data are studied, along with several optimization schemes.

The basic mechanism for resilient video transmission and the methodology followed are presented in detail in chapter five, as well as the tools used for its modelling and evaluation. In the sixth chapter, the same procedure is followed to describe two further enhancements to the base mechanism for resilient video transmission.

The seventh chapter exposes the planning and progress of the project throughout the first and second semesters.

Chapter 2

Techniques for Improved Video Transmission

In order for someone to receive a video stream over a wireless network while experiencing satisfying quality, the network and applications at both sides must obey a set of parameter requirements. Quality of Service (QoS) characterizes these and by taking them into account, the network protocols and applications can ensure that quality is safeguarded at the network level. In turn, this does not mean that the quality perceived by the user is ideal as there are other factors that can affect the video transmission.

As the signal is transmitted omnidirectionally over the air, it tends to suffer form physical effects that degrade its quality. These effects can be caused by shadowing and multipath fading for example, which usually originate errors in packets. These are not the only variables that cause packet corruption. Antenna range, hidden terminal, incapacity of undergoing several tasks at the same time by the antenna and node overpopulation of an area are also causes of packet corruption.

To answer these problems, there are several video transmission techniques that prevent and mitigate the effects of packet corruption or loss in the data transmitted. To be truly effective, these techniques must ensure the in-time reception of the data, as users of real-time video applications will suffer from packet delays caused by routing, errors, error correction and retransmission.

This chapter describes different types of known video transmission techniques, like Retransmission, Forward Error Correction, and Hybrid schemes.

The concept of a retransmission scheme is that of waiting for acknowledgements from the receiver after a transmission to confirm the successful reception of a packet. If this confirmation is not received by the sender, a retransmission process is triggered.

Forward Error Correction mechanisms prepare the data during the transmission process for expected loss. To do this, they add redundancy to it, as a means to prevent degradation of quality of the received data due to packet loss or corruption.

Hybrid schemes aim to combine the two previous mechanisms as they combine retransmission and Forward Error Correction schemes to achieve the best possible outcome of a given network scenario in terms of packet loss.

2.1 Retransmission Based Techniques

Automatic Repeat ReQuest (ARQ), also known as Automatic Repeat Query, is an error control protocol that makes the receiver initiate the process of retransmission after a frame is received in a flawed way or incorrectly. When the transmitter fails to receive an acknowledgement that confirms the reception of the data, it usually retransmits the data after a predefined timeout and repeats this process until a successful acknowledgement is received or the maximum number of retries is reached. Three of the main types of the ARQ schemes (Shu Lin, 2004), are Stop-and-Wait ARQ, Go-Back-N ARQ and Selective Repeat ARQ.

Stop-and-Wait ARQ is the simplest of the three, having one frame sent at a time with no additional frames being sent until reception of the previous one is confirmed. With this behavior, the mechanism aims to achieve the best possible rate of packet delivery, but its drawback is that it is not suitable for real-time applications where the in-time reception of the data is crucial for the user experience.

Go-Back-N ARQ is more complex, allowing frames to be sent even if the previous ones were received without confirmation. This scheme keeps track of the sequence of frames. When the last frame is received, a request for retransmission of the frames without confirmation is issued. This process is repeated until all the frames are received with an acknowledgement signal. This method allows for many frames to be sent multiple times.

Selective Repeat ARQ can be used for the delivery and acknowledgement of sent frames or the delivery of subdivided messages in sub-units. In the first case, the scheme continues to accept and acknowledge frames after an initial error. The process continues until a specified number of frames has been received, thus depicting the window size. There is one of these for both transmitting and receiving and they must be equal. The unreceived frames have their sequence numbers sent to the sender for retransmission. The sender continues sending frames of unconfirmed data until its windows size is reached. When this happens, the sender re-sends the frame numbers given by the acknowledgement signals and continues where it left off. To ensure the reception of all of the frames, the size of the sending/receiving windows must be half of the maximum sequence number.

In the case of subdivided messages, each message contains a variable number of sub-blocks. Each non-acknowledged response carries an additional bit flag indicating each sub-block successfully received. Each re-transmission decreases in length, as it only contains the non-acknowledged sub blocks.

Automatic Repeat ReQuest schemes with variable length messages have increased difficulty with longer messages, because each repeated message is full length. However, selective retransmission of variable length messages, when using Selective Repeat ARQ, eliminates the difficulty with delivering longer messages. This is evidenced by retention of successfully delivered sub-blocks after each transmission and the decrease in the number of outstanding sub-blocks after each transmission.

2.2 Forward Error Correction

A Forward Error Correction scheme can be used to improve the quality of video being transmitted over a noisy and lossy network. To do this, the forward error correcting scheme adds redundancy to the original data, so that it is possible to correct, without retransmission, eventual errors or losses due to the characteristics of the network, therefore increasing robustness of the transmitted video.

2.2.1 Categories of Forward Error Correction

There are several ways at which FEC can be used, being two of the most used ones recent implementations of Byte-Level FEC and Packet-Level FEC are found in (Basalamah and Sato, 2007) (Nafaa et al., 2008), respectively.

Byte-Level FEC recovers byte errors and provides small-scale error protection on the link layer of wireless networks. Packet-Level FEC uses blocks of packets on the MAC layer. Since these have sequence numbers, when one is lost, the decoder has information about its position. Because coding theory defines an error as a corrupted symbol in an unknown position and an erasure as a corrupted symbol in a known position, the decoder can use this erasure information to improve the FEC error correcting capability (Basalamah and Sato, 2007).

Concerning packet-level FEC there are several performance metrics that can be used (Nafaa et al., 2008) like redundancy ratio, decoding inefficiency ratio and, encoding/decoding times, as well as bandwidth.

Firstly, the redundancy ratio n/k where k are the source packets and n are the redundancy packets, is usually referred to as the stretch factor of an erasure code, and it quantifies the amount of redundancy respecting the size of the source data. Secondly, the decoding inefficiency ratio represents the minimum number of packets necessary to recover a FEC block divided by the number of source packets. Lastly, through the encoding/decoding times and resulting bandwidth we can calculate the amount of time needed to encode/decode a FEC block of a determined size for a certain erasure code class. With this value, the achievable bandwidth for a real-time streaming system is computed, along with the suitability of such codes for resource-constrained wireless devices.

Static FEC mechanisms take a source data stream and split it into separate blocks, with each one of these holding k source packets. For each one of these blocks the FEC encoder generates a fixed number of redundancy packets identified by n. This is a static approach to FEC, meaning that the encoder always generates the same number of FEC redundancy packets. This means that more network resources are consumed to transmit this increased number of packets, which leads to a trade-off between resilience and the degradation of video quality and network resources usage.

Another known approach is that of interleaved FEC protection (Nafaa et al., 2008), which is commonly used in video transmission to reduce the effects of loss. The transmitter reorders the packets in order to separate adjacent packets by a distance that can vary over time. The use of this mechanism disperses loss and mitigates the effects of bursty loss in multimedia decoding. The main advantage is that this offers better resilience to errors while not increasing bandwidth requirements. Too much interleaving can lead to uncertain improvement of FEC efficiency, depending on the loss pattern exhibited by the wireless connection. It is important to analyze the channel dynamics through QoS metrics to determine the suited interleaving level which maximizes the recovery efficiency of the FEC mechanism. The drawbacks of this scheme are that it increases latency due to additional buffering at the transmitter, and that the efficiency of the provided protection depends on the loss pattern exhibited by the network and is severely hindered by burst error patterns. Because of this, it should not be used in time sensitive applications without a proper parameter configuration that takes into account the network's dynamics and the error pattern.

Adaptive FEC mechanisms have been gaining an increased interest because they allow for the redundant packets to be used according to the video and network characteristics. For example, Unequal Error Protection (UEP) techniques allow for the unequal allocation of redundancy to parts of the video. Making it possible to better protect the most sensitive parts of the video, taking into account its characteristics, therefore reducing the impact of packet loss on the video quality. Through this it is possible to distribute the FEC redundancy in a way that best meets predefined objectives.

2.2.2 Error Correcting Codes

A FEC code can be described as an algorithm that defines a sequence of symbols such that any errors that are introduced can be detected and corrected based on the remaining symbols. This section presents three categories of error correcting codes: Linear Block Codes, Convolutional Codes and Fountain Codes, which are also known as rateless erasure codes. The choice fell on these three types of codes for the following reasons: Firstly, Linear Block Codes are simple and widely used in research (Lecuire, 2012) (Immich et al., 2013) Hassan2010. Secondly, Convolutional Codes are used in mobile communication networks (Valenti and Sun, 2001) (Raychaudhuri and Mandayam, 2012) and WiMAX (Andrews et al., 2007) (Berrou et al., 2005) (Reddy and Lakshmi, 2013). Lastly, Fountain Codes, due to their performance and efficiency (Digital Fountain, 2005) (Ahmad et al., 2011).

2.2.2.1 Linear Block Codes

A block code is a rule for converting of source bits s, of length k, into a transmitted sequence t of length n bits. To add redundancy, n is made greater than k. In a linear block code, the extra $n - k$ bits are linear functions of the original k bits, which are called parity-check bits (MacKay, 2003).

- **Reed-Solomon:** The RS codes were developed in 1960 by Irving S. Reed and Gustave Solomon (Reed and Solomon, 1960). They are mostly known by their use in digital storage devices like CD, DVD and even Blu-Ray and digital communication, namely WiMAX, space transmission and DVB-S. According to (Jr. and Cain, 1981) and (Wicker and Bhargava, 1994) RS codes can recover the maximum possible number of erasures for a given transmission overhead. These codes are not very complex and therefore provide the capability to achieve better performance in real-time services (Neckebroek et al., 2010).

  RS codes take k data symbols of s bits each, and then parity symbols are added to construct a n symbol codeword. With the addition of these parity symbols, a RS decoder can recover up to t symbols with errors in a given codeword, where $2t$=n-k.

  In (Chen et al., 2005) the authors compare RS codes against Low Density Parity Check (LDPC) over noise burst channels, which the results suggest that non-binary LDPC codes offer an advantage in magnetic recording systems with a performance increase of about 2.7dB over Reed-Solomon codes in burst noise error recovery for a frame error rate of 10^{-4}. Reed-Solomon can handle variable packet sizes with low additional overhead but with computational complexity and performance implications (3GPP, 2005).

- **Low Density Parity Check:** Despite having been introduced by Gallager in the 1960s (Gallager, 1963), LDPC codes were mostly overlooked for more than 30 years, and were rediscovered by MacKay And Neal in 1995 (MacKay et al., 1995). Further research has been developed by Luby, Shokrollahi et al., which lead to the development of other codes such as Tornado (Byers et al., 1998), LT (Luby, 2002) and Raptor (Shokrollahi, 2006). These codes are known for approaching the Shannon Limit (MacKay and Neal, 1996), this is, a noise threshold can be set very closely to the theoretical maximum for a symmetric memory-less channel (MacKay, 2003). LDPC codes are defined by a sparse parity-check matrix which is often randomly generated, subject to the sparsity constraints. These codes are used in the DVB-S2 standard (Eroz et al., 2004) for satellite transmission of digital television. Also they were selected to be used in the ITU-T G.hn standard over Turbo codes for their lower decoding complexity at operating data rates close to 1Gbps (Oksman and Galli, 2009). They are also used as part of the IEEE 802.11 standard, specifically, the optional part of IEEE 802.11n and IEEE 802.11ac, in the High Throughput (HT) PHY specification (IEEE, 2009).

LDPC codes usually use a Belief Propagation system to correct errors. Considering two types of nodes, check nodes and message nodes, all message nodes send a message to all their connected check nodes with the bit they believe to be correct. Every check node calculates a response with parity-check equations using the messages they received from before. The use of parity-check equations forces all message nodes that connect to a particular check node to sum to 0. Finally, the message nodes then use the messages from check nodes to decide the value of the bit by majority rule and send this bit value to the check nodes. This must be repeated until all the equations at all check nodes are satisfied.

2.2.2.2 Convolutional Codes

Convolutional codes, contrary to block codes, do not divide the source stream of data into blocks, but instead read and transmit bits continuously. Convolutional codes use streams of bits and symbols of arbitrary length which are transformed according to a code rate of m/n, with $n \geq m$, where m is the size of the original symbol and n is the resulting symbol size. Convolutional codes take into account a constraint length k, that represents the input bits from the original symbol used to perform the transformation (MacKay, 2003).

- **Turbo:** Turbo codes were developed by Berrou, Glavieux and Thitimajshima in 1993 and were the first practical codes to come close to the channel capacity (Berrou et al., 1993). They are widely used in 3G and 4G communication networks specifically in the HSPA and LTE standards (Valenti and Sun, 2001). Other applications include the satellite communication systems' interaction channel (ETSI, 2009), deep space communication and IEEE 802.16 WiMAX in the form of block turbo codes and convolutional turbo codes (Andrews et al., 2007) (Berrou et al., 2005). There are many forms and instances of turbo codes, using different component encoders, interleavers, input and output ratios, and even puncturing patterns.

  A classic turbo encoder sends three sub-blocks of bits. The first sub-block is composed by m-bits of payload data. The second sub-block contains $n/2$ parity bits for the payload data, computed using a recursive systematic convolutional code (RSC). The last block contains $n/2$ parity bits for a known permutation of the payload data, computed by RSC as well. This gives us a total size of the block of $m+n$ bits of data, thus the code has a rate of $m/(m+n)$.

2.2.2.3 Fountain Codes

Digital fountain codes, like LT codes, were invented by Luby in 1998, and they work as follows. We can look at the encoder like a fountain that produces and endless supply of encoded packets (water drops). If we take an original file of size Kl bits, and each drop contains l encoded bits. To receive the encoded file, we need to receive "drops" until the number of "drops" is a little larger than K. With these, the original file can be recovered. Digital fountain codes are rateless in the sense the number of encoded packets that can be generated form the source message is potentially limitless, and the number of packets generated can be determined in real-time. Regardless of the error/loss rates of the given channel, as many encoded packets as are needed can be sent in order to successfully recover the source data (MacKay, 2003).

- **Raptor:** Raptor error correcting codes are very popular due to their efficiency (Digital Fountain, 2005). Raptor codes presently support all identified Multimedia Broadcast Multicast Services (MBMS) requirements, like bearer rates, file sizes, loss rates, packet sizes, packet size variability and protection period for streaming without negative tradeoffs. A Raptor code is a type of fountain code, capable of producing an unlimited sequence of encoded symbols from a block of k fixed-length source symbols. These codes allow the actual number of encoded symbols and thus the code rate to be determined as needed to combat the current level of network packet loss (Digital Fountain, 2005). The performance of the code in all cases is close to the ideal code. Concerning resource consumption, Raptor codes have a low resource consumption, requiring at the most 0.2% of the transmission resources of an ideal code at high loss rates (10% Block Error Rate (BLER)). Compared to Reed Solomon erasure codes they require much less processing power for encoding and decoding, with this factor increasing linearly with the level of provided protection. Raptor codes have a major drawback in that they are heavily covered with patents and that makes them a poor candidate for use in the proposed mechanism. Other drawback is the fact that they are not systematic, which means that the original information is not contained in the coded message. This causes problems to some applications that require the original information for better performance.

2.2.3 Adaptive FEC approaches

Some adaptive mechanisms use packet loss information from an access point or the receiver as a means to regulate the amount of redundancy sent (Hassan and Landolsi, 2010) (Tsai et al., 2011a).

The adaptive packet and block length forward error correction (APB-FEC) general architecture is presented in Figure 2.1. This scheme aims to solve the problem of conventional packet level FEC by the use of smaller packet length, while increasing the FEC block length (Tsai et al., 2011a). In turn this reduces the packet error rate and therefore enhances the recovery performance. Due to the smaller packet size in each FEC block, the total size of the FEC mechanism will not increase when the FEC block length is increased. This mechanism has low recovery impact in the sense that a small lost packet is replaced by another smaller redundant packet. The large block length can overcome a higher error rate, as when the transmission has a high bit error rate, where the mechanism has more opportunities to recover the error than a conventional FEC mechanism.

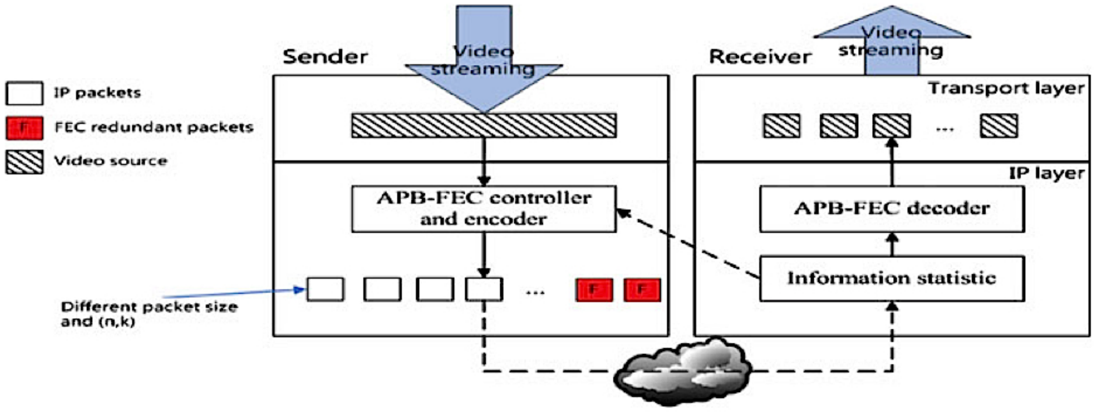

Figure 2.1: APBFEC Architecture
Source: (Tsai et al., 2011a)

When the sender has a video stream to send over the wireless network, it buffers the stream at the video streaming buffer in the transport layer. The sender then defines and adds the appropriate packet size and block size (n,k) depending on the network environment. After the packet is received, the receiver checks all the packets. If the number of packets with errors is less than or equal to (n-k), the decoder recovers the error source packet by using FEC redundancy. The mechanism calculates the average packet loss rate and burst packet loss length which will be fed back to the sender. These informations are what the sender relies on to decide the packet length and FEC parameters.

The results show that through the use of APB-FEC, which determines the appropriate FEC mechanism parameters in order to avoid too many IP headers and FEC redundancy overheads, does give better performance and PSNR (Peak Signal-to-Noise Ratio) value than conventional FEC mechanisms.

Similar to the first mechanism presented in this section, the need for buffers and the transmission of data from the receiver to the sender is not optimal, since the buffers used introduce delay, which degrades video transmission. Also, there is no guarantee that the data transmitted from the receiver to the sender arrives correctly, thus limiting the usability of this mechanism.

As an alternative to using feedback information from the receivers, other mechanisms just use the video information present at the sender side to assign the UEP accordingly (Talari et al., 2013) (Lecuire, 2012) (Immich et al., 2013).

An optimized cross-layer forward error correction coding scheme for H.264 AVC video transmission over wireless networks is proposed where FEC codes are used at both the application layer and the physical layer (Talari et al., 2013). At the application layer, Unequal Error Protection Luby Transform codes are used (Rahnavard et al., 2007) (Rahnavard and Fekri, 2006). As for the physical layer, the choice fell on rate-compatible punctured convolutional (RCPC) codes (Hagenauer, 1988). Figure 2.2 presents the architecture of the OCLFEC mechanism.

In the Moving Picture Experts Group (MPEG) standard as well as in H.264 AVC the video frames are grouped into Groups of Pictures (GoP) and each one is encoded as a unit, in this case each GoP is comprised by 30 frames. A fixed video part size is used, which leads to the aggregation of the macro blocks of a frame to form a fixed video part size. This mechanism assigns different priorities to the GoPs in order to compute the adequate FEC protection. The priority is assigned through the commutative mean squared error (CMSE) of a video par which takes into consideration the error propagation within the entire GoP.

The video parts can either be directly passed to the physical layer or encoded before being passed. When no encoding is performed at the application layer, each video part forms and application layer frame and is given to the lower network layers. When encoding is performed, LT-coded application layer frames are generated, keeping the same size, thus increasing the number of frames to be processed.

At the physical layer, cyclic redundancy check (CRC) bits are added to each application layer frame to detect RCPC decoding errors. Next, the application layer's frames are individually encoded using the RCPC code.

The last step involves the optimization of the FEC codes with different parameters for different situation in order to maximize the PSNR for a given available bandwidth.

The results show that this scheme outperforms other FEC schemes that use only UEP at the physical layer, and it also outperforms schemes that use UEP both at the application layer and at the physical layer.

The major drawback of this scheme is its optimization feature, in the sense that it needs to be done prior to transmission for every specific video taking into account the video's characteristics. Since this scheme does not account for the storage of the optimization information of different videos, its usability is limited in the manner that the optimization process for each different video takes time prior to transmission, thus not being suitable for real-time video transmission.

Regarding the Transport Audiovisuel avec Protection Inégale des Objets et Contrôle d'Admission (TAPIOCA) system, it encodes the video data units of each

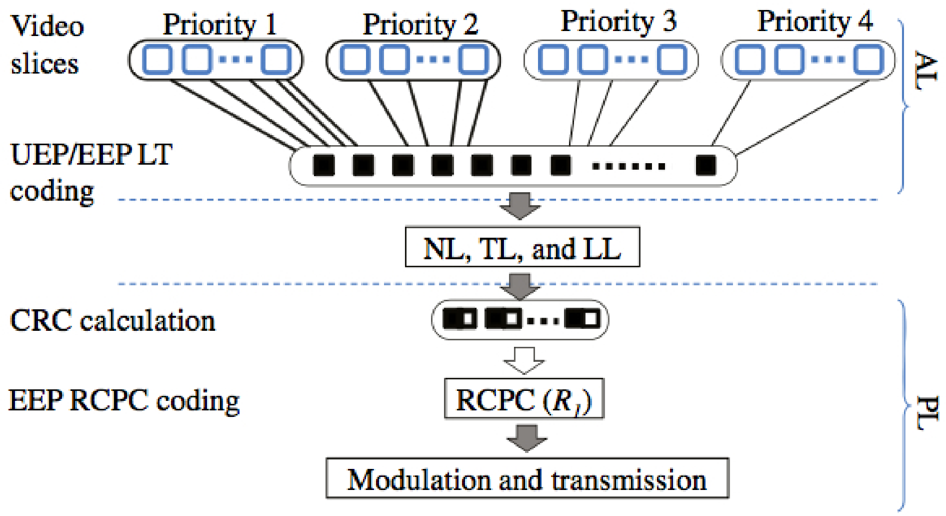

FIGURE 2.2: Optimized cross layer FEC scheme Architecture
Source: (Talari et al., 2013)

layer separately as opposed to Priority Encoding Transmission (PET) systems, which usually fill a block of packets with parts of data from all layers (Lecuire, 2012). The PET systems need to collect several video frames, or Group of Pictures, before starting the encoding process. This in turn, means that a longer the GoP size incurs in a larger delay. The longer GoP size also makes for a larger k dimension (source data) and length n (redundancy data) of the erasure code, thus needing more computation time to process the GoP.

Considering that the video stream is divided into several layers and that the topmost one has the most significant data in the stream, the TAPIOCA system computes the utility of the layers to adequately distribute the global ratio of FEC overhead. Then, the video data unit is packetized into several packets of the same size, with the last packet being padded if its size is not a multiple of the block size. Then, the packets are encoded with Reed-Solomon codes. For each video data unit to be transmitted, the optimal code dimension and length, and block size are computed prior to encoding so that two conditions are satisfied. The first is the lowest number of packets necessary to transmit the data unit with no protection which represents the k original data dimension of the code; The second is the value which minimizes the error margin or the n redundancy value of the code.

The results show that the TAPIOCA mechanism is rather simple and the FEC assignment scheme requires only a few simple calculations. It is shown that in the most cases for MPEG-4 video streams, TAPIOCA achieves a decodable frame rate significantly higher than that of a Priority Encoding Transmission with the same ratio of FEC overhead.

This mechanism requires that the original video must be split by layers. This means that if a video file is not already in this form, it needs to be transformed which requires pre-processing and ultimately means that the mechanism can not be used in real-time.

The **VI**d**E**o-a**W**are FEC-based mechanism (ViewFEC) (Immich et al., 2013) for packet loss resilient video transmission with unequal error protection has two modules that compute several parameters at the network layer: **C**luster **A**nalysis **K**nowledge bas**E** (CAKE) and **C**ross-**LA**yer infor**M**ation (CLAM). Figure 2.3 presents a general overview of the stages which constitute the ViewFEC mechanism. The CLAM module supports the functions that will compute the parameters for the required amount of redundancy needed so that video quality is good. It accesses information like video characteristics at the application layer, from the network layer, with its objective being the identification of the GoP length. The larger the GoP, the larger the influence packet loss has over the video quality. The CAKE module's objective is to optimize video transmission, through the use of a database with video and motion complexity that is built using hierarchical clustering before the video distribution.

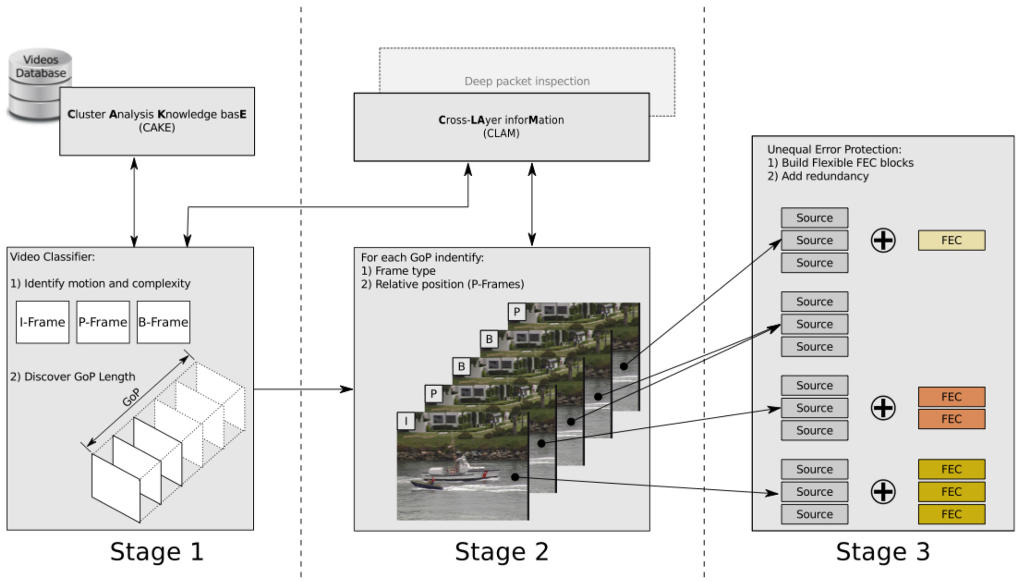

FIGURE 2.3: ViewFEC stages
Source: (Immich et al., 2013)

This mechanism can be implemented at access points, routers and the video server due to its deployment flexibility derived from the decision-process being at the network layer. Its flexibility comes from the possibility of changing the modules to obtain the desired behavior. When a cross-layer approach is not a viable way to obtain information from the application layer, for example in a router or access point, the CLAM module can be exchanged for another that can offer the means to obtain the desired information.

The ViewFEC mechanism is comprised of three stages. The first stage is where information from the two modules is fetched, and the mechanism identifies several important video characteristics, namely motion, complexity levels and GoP length. In the second stage, more information from the video data is retrieved, such as type and relative position of the frames inside the GoP. In the third stage, the needed amount of redundancy is calculated and adjusted in real-time according to the details obtained from the previous stages.

The results show that the mechanism outperforms non-adaptive FEC schemes in terms of video quality, and especially network overhead. The overhead measured was on average 40% less than while using a standard FEC mechanism. Therefore, this adaptive mechanism is able to enhance video transmission without adding unnecessary network overhead, which in turn leads to a better usage of the wireless network resources. However, this mechanism only takes into account the module's informations when adding redundancy to the data. This means that it discards the network information and cannot adequately adjust the amount of redundancy for bursty networks or other problematic network states.

2.3 Hybrid Techniques

There are some hybrid schemes which combine the Automatic Repeat ReQuest scheme with Forward Error Correction, thus trying to improve on the use of a single mechanism. Hybrid schemes can work in different ways: they can simply use a retransmission based scheme when loss is detected as in HARQ, or they can use the receiver network information to adjust previously to transmission the amount of redundancy and configuration of retransmissions as in AHECM. There are also hybrid adaptive FEC schemes which make use of both retransmission and adaptive FEC protection such as ACFEC (L. Han and In, 2010).

Hybrid Automatic Repeat ReQuest (HARQ) has two types of mechanisms. Type one HARQ (Tsai et al., 2010) (Tsai et al., 2009) contains additional symbols for error detection and correction in transmission blocks. This scheme repairs and accepts a block in the receiver as the correct block, if the corrupted symbol count is lower or equal to the maximum number of repairable errors. If the number of errors is superior to the recoverable value, the scheme rejects the block and it is retransmitted. The type two of the HARQ scheme (Moid and Fapojuwo, 2008) (Le et al., 2007) does not discard corrupted blocks but keeps them in a buffer for further processing. Only the stronger protection part of the previous corrupted block is transmitted instead of the whole block. The decoder in the receiver is used to repair the stored corrupted block according to the protection symbols. Due to this, the type two scheme can preserve system resources such as network bandwidth and CPU time. On the downside, present HARQ mechanisms cannot retransmit packets to the receiver in time, this means that the video applications on the receiver cannot use the data carried by packets that do not arrive in time. These HARQ mechanisms also cannot dynamically adjust the FEC redundancy according to network conditions, raising the probability for network congestion due to inappropriate FEC redundancy.

The Adaptive Hybrid Error Correction Model (AHECM) mechanism collects meta-information about the average packet loss rate, the average Round-Trip Time (RTT), and the available bandwidth at the receiver (Tsai et al., 2011b). When meta-information changes due to channel or network conditions, AHECM makes the receiver send the meta-information to the sender in order to adjust its parameters. With these updates, AHECM decides the appropriate FEC redundancy for different video frame types. With information about the average RTT and tolerable end-to-end delay at the receiver side, AHECM can find the maximum retransmission slot time and retransmits the lost packet to the receiver in time to reduce FEC redundancy. With information about the average packet loss and the available bandwidth, AHECM can calculate the appropriate FEC parameter to avoid network congestion and unnecessary FEC redundancy. When the end-to-end requirement is met, AHECM will only transmit the necessary number of redundant packets to the receiver.

Another alternative is a FEC and retransmission scheme to control the channel rate (Hassan and Landolsi, 2010). It takes into account that quality is seen as the continuity of played back video, and measures it by monitoring the occupancy of the playback buffer. Also, this scheme has a preloading phase that is used to stabilize the playback in the ongoing transmission. This is done by always maintaining a number of frames in the playback buffer to combat buffer underflow and therefore starvation. In turn, this adds delay every time a transmission is started.

The video server encodes into frames the video information that will be packetized at the link layer, which then passes them to the error control module which adds to each of the packets a certain number of FEC bits according to the channel state. The number of FEC bits is applied in a varied fashion to the packets of the same video frame, according to information about the channel state and the content of the playback buffer. At the receiver side, the received video frames are decoded for playback, while information is sent to the video server about the channel state and the occupancy of the decoder buffer. The control packets used to relay information to the server are small, and this means they can adequately be protected with FEC alone and/or higher SNR. This means that in theory the feedback channel has less probability for the occurrence of errors. Also, this information can be fed back through the out-of-band Real Time Streaming Protocol (RTSP) control protocol. In turn, the video sender uses this information to optimize the transmission parameters to send the next frame. If a packet is received in error, it is discarded and a retransmission is requested.

As stated above, the existence of a playback buffer means that a startup delay is present. Adding to this is the delay inherent to all retransmission mechanisms, which can guarantee a certain level of packet delivery, but the downside is that it may not be in-time and/or sequence. Therefore this means that this scheme is not well suited for real-time video streaming applications.

The Adaptive Cross-layer FEC mechanism (ACFEC) operates at the wireless access point and follows the integrated approach paradigm, focusing on the use of functionalities from different layers (L. Han and In, 2010). The adaptive FEC controller cooperates with the UDP protocol and with the MAC protocol. It has two main components, a FEC packet generator and a packet loss monitor which are represented in Figure 2.4.

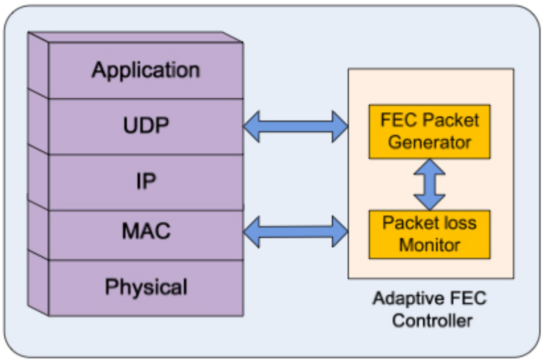

FIGURE 2.4: ACFEC Architecture
Source: (L. Han and In, 2010)

The video server encapsulates the video data in Real-time Transport Protocol (RTP) (Schulzrinne, 1996) packets and delivers them to the receiver through the wireless access point. When a packet arrives at the wireless access point, the adaptive FEC controller analyses the header and identifies the type of the packet. When a frame is sent, the access point does not send any more frames until it receives an acknowledgement form the receiver. If the acknowledgement is not received before the time-limit, the frame is sent again. The process stops if the retransmission counter exceeds the limit of retries, and the frame is given as lost. Traditional transport protocols do not use this information, contrary to the adaptive FEC controller, which uses it to control the redundancy ratios adaptively.

When the packets are routed to a receiver through the wireless access point, the adaptive FEC controller classifies the video data packets and groups them into blocks, which are then stored in a buffer for FEC packet generation. The packet loss monitor analyses the results of the video data packets transmission by obtaining the failure information from the MAC layer. If a packet is lost, the loss counter of the monitor is increased. After sending a block of video data packets, the loss monitor calculates the number of redundant FEC packets to be generated. Due to the fact that the packet loss monitor detects the number of losses in each block, the ACFEC mechanism is able to accurately produce relevant amounts of FEC packets for several different amounts of packet losses.

This mechanism has a major drawback that is the need to use a FEC controller in a wireless access point, which has problems in terms of scalability. Therefore its use in wireless mesh networks or Mobile Ad Hoc Networks (MANET) is somewhat limited in the sense that each node is a router and this will add complexity and strain its resources.

These mechanisms can surely improve the packet loss rate, which does not properly mean that we can expect an increase in video quality. This is due to the fact that in video transmission, sometimes it is not only important to receive the highest amount of packets possible, but also to receive them in time and in sequence, or the user will experience buffering issues and stuttering.

2.4 Summary

Several techniques for error prevention and recovery that can be used in video transmission were studied in this section.

Retransmission techniques, as the name implies, involve the retry of the transmission of the data if it fails to arrive at the destination due to errors occurred in route. There are many retransmission techniques ranging from simpler schemes like Stop-and-Wait ARQ to more complex ones like Go-Back-N ARQ.

FEC techniques try to recover data before it is lost. They do so by adding redundancy to the data to be sent with the help of Linear Block Codes, Convolutional Codes and Fountain Codes. This can be done in a static way, with redundancy being the same for every frame of the video, or it can be done in an adaptive way selecting the adequate amount of redundancy according to some specific rule. Retransmission vs FEC.

Hybrid techniques merge both of the previous techniques to offer a more complete protection of the data thus aiming to achieve better resilience against data loss in video transmission.

Chapter 3

QoE Assessment Methods

Quality of Experience (QoE) represents the measurement of a user's experience with a service. This chapter addresses video QoE assessment approaches which use several metrics to assess video quality as perceived by the user. It differs from Quality of Service (QoS) in the sense that it assesses the performance of the network in terms of the video quality after transmission, as opposed to the raw performance of the network in terms of error rates, bandwidth, throughput, delay and jitter. There are at least three kinds of evaluations that can be used to assess the quality of experience, namely subjective, objective and hybrid. Subjective approaches require the use of humans in order to assess the perceived quality of the video, while objective approaches tend to use metrics that emulate in some way the perception of a human viewer and its consequent assessment of quality. Hybrid evaluation mechanisms usually employ some kind of subjective human evaluation apart from the objective evaluation. Typically, this is done to calibrate some kind of classification mechanism that will then be used to assess the video quality without any human interaction.

3.1 Subjective Quality of Experience Evaluation and Metrics

The subjective type of evaluation usually produces results to which users identify themselves more closely. This occurs because there is no better indicator of quality than the subjective evaluation given by a panel of human beings. Nevertheless, this assessment depends on his/her own experience. It can also be impacted by the user's mood and by the responsiveness of the system as well as many other variables that affect video quality in the system where the evaluation is occurring. This type of evaluation consists in a group of human observers that evaluate video sequences depending on their point of view and perception of the quality of the video.

The introduction of digital television compression produces scene-dependent time-varying impairments to the picture quality. Due to this, the quality can rapidly fluctuate depending on the content of the scene, which results in short bursts of quality reduction. The conventional ITU-R methods alone, some of them presented below, like Double-Stimulus Impairment Scale (DSIS), Double-Stimulus Continuous Quality-Scale (DSCQS) and Single-Stimulus (SS), are not enough to assess this type of material. As the double stimulus laboratorial testing does not replicate the SS home viewing conditions, new subjective metrics which measure the video quality continuously have been presented in (ITU-R, 2012), namely Single Stimulus Continuous Quality Evaluation (SSCQE) and Simultaneous double stimulus for continuous evaluation (SDSCE).

3.1.1 Mean Opinion Score

In multimedia, the Mean Opinion Score (MOS) provides a numerical indication of the perceived quality from the users' perspective of received media after compression and/or transmission. This translates as a single number ranging from 1 to 5, where 1 is the lowest perceived quality, and 5 is the highest perceived quality measurement. It was originally designed to evaluate the quality of voice telephone calls (ITU-T, 1994) (ITU-T, 2003), but is now widely used in video quality evaluation as well (ITU-R, 2012) (ITU-R, 2008). Although being accurate, this type of assessment is costly in terms of human resources and time and is not viable for on-line video evaluation.

3.1.2 Double-Stimulus Impairment Scale

In the double-stimulus impairment scale (DSIS) method (EBU method) (ITU-R, 2012) there must be a previous selection of test material to allow for a meaningful evaluation, as well as the establishment of which test conditions should be used. The double-stimulus (EBU) method is cyclic. The assessor is first presented with an unimpaired reference, then with the same reference but impaired. In evaluation sessions that last up to half an hour, a series of pictures or sequences in random order and with random impairments are presented to the assessor. At the end of the group of sessions, the mean score for each test condition and test picture is calculated. This method uses the impairment scale, which usually finds that the results are more stable for small impairments rather than large impairments.

3.1.3 Double-Stimulus Continuous Quality-Scale

The double-stimulus continuous quality-scale (DSCQS) method (ITU-R, 2012) is thought to be especially useful when it is not possible to provide test stimulus conditions that exhibit the full range of quality. This method is also cyclic as DSIS, as the assessor is subjected to the viewing of a pair of pictures. One directly from the source and the other via the process under evaluation. Afterwards he is asked to assess the quality of both. The sessions last up to half an hour, the assessor is presented with a series of random ordered picture pairs, with random impairments. At the end of the sessions, the mean scores for each test condition and test picture are calculated.

3.1.4 Single-Stimulus Methods

In single-stimulus (SS) methods (ITU-R, 2012), a single image or sequence of images is presented and the assessor provides an index of the entire presentation. The test material can include only test sequences, or it might include both the test sequences and their corresponding reference sequence. There are three types of SS methods that have been used for television assessments: Adjectival categorical judgement methods in which the observers assign an image or image sequence to one of a set of categories that are defined in semantic terms; Numerical categorical judgement methods use an 11-grade categorical scale and offer good sensitivity and stability for the single stimulus numerical categorical scale when no reference is available; Non-categorical judgement methods where observers assign a value to each image sequence shown. The value assigned can be in form of a continuous scale with semantic labels or a numeric scale with a defined range.

3.1.5 Single Stimulus Continuous Quality Evaluation

In the (SSCQE) method (ITU-R, 2012) an electronic recording handset is used to obtain the continuous quality assessment from the subjects. The sessions are thirty to sixty minutes long and consist of programme segments of at least five minutes of duration, processed according to the quality parameters under evaluation. The assessors continuously rate the video with the help of the mechanical slider in the recording handset device.

3.1.6 Simultaneous Double Stimulus for Continuous Evaluation

The (SDSCE) method(ITU-R, 2012), involves the viewing of two sequences at the same time, one is the reference sequence and the other is the test condition sequence. The test session consists of a series of video segments and test condition pairs in a pseudo-random order. Each session contains every video session and test condition but not all the combinations of the two. Subjects are requested to observe the differences in the two sequences and to continuously judge the fidelity by moving the slider of the handset-voting device. The device has a scale of 0 to 100, being 0 the worst quality and 100 the best.

3.2 Objective Quality of Experience Evaluation and metrics

Since the subjective approach is not suitable for a practical real-time implementation, there are other approaches that comprise the use of algorithms and formulas with the current network parameters and information to automatically give a result. These are called objective QoE metrics. Even though there have been recent advances in the fields related with objective assessment of video image quality following human perception, there are still some setbacks related to physiology, psychology, sight research and computer science.

Regarding the way the metrics are developed we can still differentiate between two sets, namely signal based techniques which are designed to analyse the signal, and techniques that are developed with basis on modelling the human visual system and its reactions.

The main issue with Signal-to-Noise Ratio (SNR) based on other signal fidelity based techniques is that, although they are simple and widely used, they can predict poorly the perceived video quality, namely when noise is not additive (Karunasekera and Kingsbury, 1995) (Limb, 1979). Even though the physical alterations relative to Mean Squared Error (MSE), SNR and Peak Signal-to-Noise Ratio (PSNR) can be observed in the video quality change, the above mentioned signal fidelity based metrics fail to predict a human visual system (HVS) perception due to some problems. Firstly, not all changes in a video image are noticeable. Secondly, through the user's viewing, not every region of the image receives the same level of attention. Thirdly, not every change applied to the video, be it intentional or not, leads to distortion. And lastly, not every alteration to the video has the same effect in perception with the same amount of change (Lin and Kuo, 2011).

Perceptual Video Quality Metrics (PVQM) are defined by the objective models that predict subjective visual scores. There are two categories of PVQMs: vision-based modelling and a signal-driven approach. In the first category PVQMs are developed with information from the systematical modelling of relevant psychophysical properties and physiological knowledge like temporal, spacial and color decomposition, contrast sensitivity function, luminance adaptation and masking effects. The second category bases itself on signal extraction and analysis like statistical features, structural similarity, luminance or color distortion, and common visual artifacts. In this last category, metrics take into account how pronounced the features of the video image are, in order to estimate overall quality. In turn, this does not means that the human vision system knowledge is disregarded, it just means that image content and distortion analysis is the basis of the design.

Very important to the design of PVQMs is image/video decomposition, visual feature and artifact detection, just-noticeable distortion (JND) modelling and visual attention (VA) map generation.

Image/video decomposition follows the HVS in the sense that it has separate processing for chromatic and achromatic signals, different visual pathways for signals with fast and slow motion, and special cells in the brain's visual cortex for distinctive orientations. This leads us to the fact that by decomposing the video image into its different constituents facilitates the evaluation of changes in the signal, and unequal treatment of each signal component, just like in the HVS.

In features and artifact detection it is known that the detection of several features of the signal and artifacts is a common component of visual quality evaluation. The HVS considers a bigger part of signal contrast than that of the absolute signal strength, due to its specialized cells that process signal contrast. In this sense, contrast is a central part of the Contrast Sensitivity Function (CSF), luminance adaptation, contrast masking and VA. Adding to this, certain structural artifacts occur in the signal compression and network transmission process, which in turn results in an annoying experience for the viewer. Some common coding artifacts are blockiness, blurring, edge damage and ringing (Shen and Kuo, 1998) (Yuen and Wu, 1998). This is where traditional measures, such as MSE or PSNR fail to reflect the perceptual effect the structural artifacts have.

The just-noticeable distortion modelling refers to a visibility threshold below which a change cannot be detected by the majority if viewers (Jayant et al., 1993) (Keelan, 2002) (Lin, 2005). This modelling addresses the problem of visual similarity, and reflects the local characteristics of the HVS.

The VA refers to the selective awareness/responsiveness to visual stimuli (M.M. Chun, 2001) (Bryan Kolb, 2008), as a consequence of human evolution. There are two types of cues that shift attention to a particular point in the video image: Bottom-up, which refer to the external stimuli; Top-down which are caused by a voluntary shift in attention like when one is ordered to direct attention to a specific location. The VA process can be regarded as having two stages: in the pre-attentive stage, all information is processed across the entire visual field; in the attention stage, the features may be bound together or the dominant feature is selected. Most of the existing computational VA models are bottom-up, this is, based upon the evaluation of contrast of several low-level features in video images, with the objective to determine which locations exhibit a different behavior than their surroundings.

Many of the existing schemes follow the methodolgy presented in Figure 3.1 where the original and distorted videos are decomposed and then pass through a contrast gain control function. Finally the features of the information are compared and a measure is generated. Some of model-based PVQM metrics are JND-Metrix and Perceptual distortion metric for digital color video.

The signal-based PVQMs are where most of the recent research effort has been directed to, dervived from the model-based PVQMs being more costly in terms of computation, and also because of the disparity between the vision research knowledge and the need for engineering the model previous to usage.

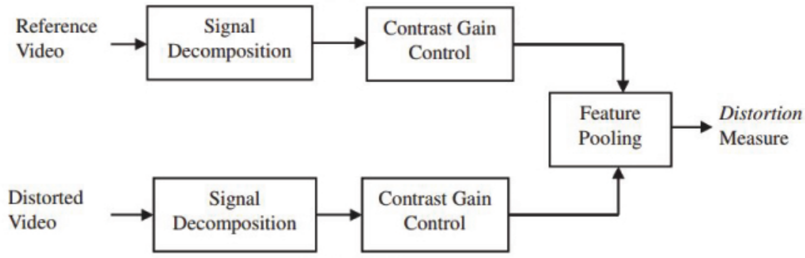

FIGURE 3.1: Model-based PVQMs methodology
Source: (Lin and Kuo, 2011)

There are some metrics that make use of both PVQM categories like the multi-metric objective picture-quality measurement model for MPEG video (Tan and Ghanbari, 2000a) which switches between a model-based scheme and a signal-based one, when the amount of blockiness in the video demands it. There is also the Vision-model-based impairment metric (Yu et al., 2002) which is a model-based metric that was used in blockiness dominated areas, through a signal-based measure.

3.2.1 Categories of Objective Quality of Experience Metrics

Quality metrics can be generally classified into full-reference, reduced-reference and no-reference with basis on the amount of information required from the reference video (Winkler, 2005).

3.2.1.1 Full-Reference

FR metrics perform a frame by frame comparison between the original video and the test video. The entire referenced video is required in an unimpaired and uncompressed form. This imposes some restrictions on the practical usability of this type of metrics. Also, a precise spatial and temporal alignment of the two videos is required in order to match every pixel in every frame between the two videos. The registration of the temporal information is another strong restriction in the sense that it can be very difficult to achieve due to frame drops, repeats or variable delay. Furthermore, these metrics also do not respond well to global shifts in brightness, contrast or color, and therefore they require a calibration of the videos. PSNR, Structural Similarity Index (SSIM), Video Information Fidelity (VIF) are some examples of full-reference metrics.

3.2.1.2 No-Reference

NR metrics only analize the test video, without needing an explicit reference clip. This makes these metrics more flexible than FR metrics as it can be difficult to obtain the reference in some cases. It is possible to have some NR metrics that combine the features of the video, like contrast, blurring, blockiness and motion. And it is also possible to have single-factor NR metrics like blockiness (Tan and Ghanbari, 2000b) (Wu and Yuen, 1997), blurring (Marziliano et al., 2002) (Wu et al., 2007) and jerkiness (Chen and Thropp, 2007) (Yang et al., 2007). Since the video can be classified according to four visual content classes (Oelbaum et al., 2009), like low rate, blur, blocking and general. Different NR metrics must be chosen to be applied to every different class of video, since a single NR metric cannot be well tuned for all situations. Another challenge that NR metrics face is that of the possibility of mistaking the actual content as distortion, a distinction humans are able to make from experience. NR metrics always have to make assumptions about the video content and distortions of interest.

3.2.1.3 Reduced-reference

RR metrics are a compromise between full-reference and no-reference metrics. They retrieve a number of features from the reference video, and the evaluative comparison of the videos is then based upon those features. This approach allows the circumvention of some of the assumptions and drawbacks of no-reference metrics while keeping a low amount of reference information. These metrics have alignment requirements that are less restrictive than those of full-reference metrics, as only the retrieved features need to be aligned.

3.2.1.4 Common Metrics

With basis on this classification we can attribute uses to each type of metrics. FR metrics are more useful for offline video quality measurement, like codec tuning or lab testing, where the conditions are well controlled and the precise evaluation of the video is more important than the immediacy of the results. NR and RR metrics have more use in monitoring or in-service video systems, where real-time measurement is essential.

No-Reference metrics cannot be used singularly to assess several types of videos. Since we have the whole original information set available Reduced-Reference can be dismissed of use in this work. Therefore, below are presented some of the most used and commonly known Full-Reference metrics (Sheikh et al., 2006) (Pedersen and Hardeberg, 2009).

Peak Signal-to-Ratio (PSNR) is a widely used metric by researchers and is very simple. It is the relationship between the maximum power of a signal and the noise power that corrupts and affects the representation's accuracy (Arun

N. Netravali, 1995). It is defined form Mean Squared Error (MSE) between an original frame o and the distorted frame d. The more similarity between the two pictures the higher the PSNR value is, represented in dB. Some authors consider that the PSNR value can be associated to a MOS for better understanding of the results as shown in Table 3.1 The great drawbacks of PSNR are that it is not suited to be used in real-time mechanisms because it can only be used on the receiver side after the image is rebuilt, and also, that it does not consider human perception, which can result in some video disruptions not being correctly detected (Piamrat et al., 2009).

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

Table 3.1: PSNR to MOS convertion
Source: (Khan et al., 2010)

JNDmetrix is a model presented in (Lubin, 1995), where the Gaussian pyramid is used for decomposition with luminance and chrominance components in the video. There are several stages to the process. Taking two paired images, these go through processes of simulating the response of the human eye due to its physical characteristics. The resulting values are then transformed into a Laplacian pyramid with seven frequency bandpass levels. These go through filtering and result in energy responses which mimic the response of the visual cortex. Finally the pyramid levels are processed to take into account contrast and human foveal sensitivity resulting in two vectors which represent each image. The result is a spatial map of JND values, representing the degree of discriminability between the two, and can be reduced to a single aggregate value, which shows the magnitude and position of noticeable differences.

Structural similarity (SSIM) is presented in (Wang and Bovik, 2002) (Wang et al., 2004) is a full reference metric which combines luminance, contrast and structural similarity of the images to compare the correlation between the original image and the one received. SSIM is designed to improve on other methods like PSNR and MSE, which are inconsistent with the perception of the human eye. The main difference between SSIM and, PSNR and MSE is that the latter estimate perceived error and the former considers image degradation as perceived change in structural information.

Video information fidelity (VIF), introduced in (Sheikh and Bovik, 2006) is based on the random field (RF) from a subband from the test image D can be expressed as $D = GU + V$, where U denotes the RF from the corresponding subband from the reference image , G is a deterministic scale gain field, and V is a stationary additive zero-mean Gaussian noise RF. This model takes into account additive noise and blur distortion.

The visual signal-to-noise ratio (VSNR) is a metric presented in (Chandler and Hemami, 2007) which uses two stages. In the first stage, it computes the contrast threshold for distortion detection in the image via wavelet-based models of visual masking and visual summation, in order to determine if the distortion in the test image is visible. If the distortion is below the threshold for detection, the test image has perfect visual fidelity. If not, a second stage is used, which uses the property of perceived contrast, and the mid-level visual property of global precedence. The two properties are then modelled in terms of Euclidean distance in the distortion-contrast space of multiscale wavelet decomposition and VSNR is computed based on a simple linear sum of the distances.

The Video Quality Metric (VQM) improves on the PSNR and SSIM by detecting artifacts perceived by humans on the images (Xiao et al., 2000), because it is based on a simplified human spatial-temporal contrast sensitivity model. The metric has five steps for the determination of quality. The first three are identical to Watson's Digital Video Quality (DVQ) model (Watson, 1998). The fourth step is the conversion of local contrast (LC) to just-noticeable differences by applying one spatial contrast sensitivity function (SCSF) matrix for static frames and one matrix for dynamic frames in one step. The last step involves the weighted pooling of mean and maximum distortion, where the two sequences (the original and the transmitted one) are subtracted, and then contrast masking is incorporated into a simple maximum operation and then it is weighted with the pooling means distortion.

The perceptual distortion metric for digital color video metric presented in (Winkler, 1999) is based on a contrast gain control model of the human visual system that incorporates spatial and temporal aspects of vision as well as color perception. This model achieves a close fit to contrast and sensitivity and contrast masking data from several different psychophysical experiments of both luminance and color stimuli.

3.3 Hybrid Quality of Experience Evaluation

The hybrid QoE metrics have a component of subjective evaluation as well as a component of objective evaluation. The subjective evaluation is usually only done once and it can then be used as many times as necessary in conjunction with the objective parameters. To be noted that the subjective can be used several times as a means to further refine the process.

3.3.1 PSQA

The pseudo subjective quality assessment (PSQA) (Mohamed and Rubino, 2002) is based on statistic learning using random neural network (RNN). Before PSQA is used in real-time, it needs to perform three-steps beforehand. This process needs to be done for each given application.

The first step is called Quality-affecting factors and Distorted Video Database Generation. In this stage, a set of quality affecting factors with impact on quality such as the codec, bandwidth, loss, delay and jitter are selected. This set of parameters and their specific parameters is called configuration. After this a distorted video database is generated by varying the configurations.

The second step is the Subjective Quality Assessment, where the chosen configurations are evaluated by means of a subjective evaluation campaign. A panel of human observers evaluates the distorted videos and the MOS is computed through the average score of all the observers. The configurations and the MOS values are stored in two separated databases called training and validation.

The last step is Learning of the quality behavior with RNN. The RNN is trained to learn the mapping of configurations and scores defined in the training database. Once the RNN is trained, a function $f()$ comes into existence that can map any possible value of parameters into MOS. The RNN is validated by comparing the value given by $f()$ at the point corresponding to each configuration in the validation database.

After this, PSQA can be run anywhere in real-time. It is only necessary to measure the parameters that affect the quality at a given time and then evaluate these with the RNN to obtain the instant perceived quality.

3.3.2 HyQoE

The Hybrid Quality of Experience (HyQoE) mechanism presented in (Aguiar et al., 2012) is a quality estimator specially designed to assess real-time multimedia applications. The HyQoE mechanism is designed based on the framework of the PSQA tool. It is comprised of four main stages: (1) Quality-affecting factors in which a set of factors relevant to service quality are defined as evaluation parameters. They are then used in conjunction with a RNN to construct a correlation between network level impairments and user level perceptual experience; (2) Generation of a distorted video database. Ten videos were selected and were transmitted through a Wireless Mesh Network which results in a realistic loss of quality; (3) Subjective Quality Assessments where a panel of human observers evaluates the generated database of distorted videos following ITU-T recommendations. The results of subjective evaluation are then separated into two sets, one for training and one for validation; (4) Learning about the quality behaviour through RNN where a RNN constructs the structure and parameters of a HyQoE model aiming to maximize the fit of the model output to the subjective ratings.

Once the RNN is trained and accordingly validated, the mechanism can be used for real-time QoE prediction without any intervention by human viewers.

The scores given by HyQoE are expressed in terms of MOS and are as close as possible to those voted by human users.

The results show that HyQoE performs well compared to other common hybrid and subjective measurement schemes. Also, an improvement over PSQA of 17% for a GoP size of 10 is shown.

3.4 Summary

QoE assessment methods involve subjective, objective and hybrid approaches.

A subjective approach mandatorily involves the presence of a human participant to perform the assessment and rate the quality of the video. These can be very precise but the requirement of a human presence can be a hindrance in terms of resources and real-time execution.

The objective approaches can be derived from the human physiology, signal-based or model based. The human physiology based ones apply algorithms that mimic the eye response to a certain stimulus for example. On the other hand signal-based approaches take the values of the video signal and assess the quality from the obtained values. They can be classified in three categories according to the amount of information they use from the original source. These categories are Full-Reference, No-Reference and Reduced-Reference.

Hybrid approaches combine both of the approaches to try to achieve a more complete evaluation of the video quality.

Chapter 4

Mechanisms for Resilient Video Transmission in Wireless Networks

In order to ensure that a given video being transmitted over a wireless network arrives with good quality for the user there are some methods that can be employed. We suggest video motion intensity classification through the training and use of a Neural Network mechanism and adaptive FEC techniques based on evolutionary computing to achieve this. The produced video motion intensity classification scheme will allow for the configuration of the adequate initial parameters of the FEC mechanism, so that the video already has beforehand, a better chance of maintaining quality than if no protection was used. The developed adaptive FEC mechanism, selects the way that a specific part of the video is given more protection following an evolutionary ant colony algorithm. It dynamically changes the amount of redundancy to be attributed, through the complex relation between the of monitoring several parameters of the video/network and path finding. This, ensures that during transmission, the parameters are regulated in order to quickly respond to changes in network state that can negatively impact the video quality.

4.1 Video Content Classification according to motion intensity

Video streams can have different and varied complexity in terms of movement (Khan et al., 2010). This means that as videos have different motion intensities, some of the information has greater importance such as the one that describes movement in contrast to the information that describes a still part of the image and therefore has less importance. Therefore, these streams or video files do not behave the same way under loss conditions depending on from which part of the video was the information lost.

In video compression different algorithms are used and produce several frame types according to the Moving Pictures Expert Group (MPEG) standard (ITU-T, 2013), I, P and B. An I frame is a complete picture in the sense that it holds all the information needed to construct a video image. The P frame or predicted picture contains a small quantity of spatial information and temporal information about the changes in the picture. Finally, the B frame only contains temporal information and is specified as the differences between the current frame and both the previous and the following frames. According to this classification, the loss of a I frame on a still video does not have the same impact as the same loss in a video with greater movement complexity. This is because the P and B frames need the information from a I frame to produce the correct video images. Therefore in a video with a larger quantity of motion, the loss of a I frame will produce a more severe impairment.

Therefore, in order for the proposed mechanism to improve the quality of experience on the receiving end, the video content being sent must be categorized (Khan et al., 2010) in a way that the amount of redundancy can be optimized taking into account the movement characteristics of the video and motion intensity. It is possible to do this with several methods like Cluster Analysis, Principal Component Analysis and Neural Networks.

4.1.1 Cluster Analysis

In Cluster Analysis or Clustering, objects are grouped such that the objects in the same group or cluster have greater similarity to each other as opposed to those in other groups or clusters. Cluster analysis is a technique for statistical data analysis that is used for image analysis (Immich et al., 2013) (Khan et al., 2010) as well as pattern recognition (Jain et al., 2000), information retrieval (Manning et al., 2008) and bioinformatics (Sturn et al., 2002).

The clustering algorithm and parameter settings to use depend on the given individual data set and intended use of the results. The process is iterative and comprises knowledge discovery and interactive multi-objective optimization which involves trial and error, often being necessary to modify the data preprocessing and model parameters to achieve the desired results. Since the notion of a cluster cannot be precisely defined, this produces the wide variety of clustering algorithms in existence (Estivill-Castro, 2002). Despite this, there are several typical cluster models.

Hierarchical clustering (Johnson, 1967) is based on distance connectivity thus a connectivity model. Objects are more related to nearby objects than to those that are farther away. These are connected to form clusters based on their distance. Objects at different distances form different clusters which are represented by a dendrogram which is a graph that represents clusters.

Centroid-based Clustering (Radev et al., 1999) is the model in which cluster are defined by a central-vector, which does not need to be a member of the data set. K-means clustering (Marroquin and Girosi, 1993) is used when the number of clusters is fixed to k. This turns into an optimization problem where the goal is to find the k cluster centers and assign the objects to the nearest cluster center, where the squared distances from the cluster are minimized.

In Distribution-based Clustering (Xu et al., 1998) clusters are defined as objects which belong to the same distribution. This type of clustering can suffer from overfitting, which occurs when the model is excessively complex and describes random error or noise instead of the underlying relationship. One typical method of this type of model is the Gaussian mixture (Banfield and Raftery, 1993) where the expectation-maximization algorithm is used and overfitting is avoided by using a fixed number of Gaussian distributions which are randomly initialized and with parameters that are optimized to fit better with the data set.

Density-based Clustering (Kriegel et al., 2011) is the method in which a cluster is defined as an area with higher density of objects than the rest of the data. The objects which stand on the sparse areas are considered to be noise and border points. This method can feature a well-defined cluster model which is called density-reachability (Ester et al., 1996). It is based on the connection of points which satisfy a fixed distance limit. The objects are only connected if they satisfy a density criterion which is specified as the minimum number of objects within the radius.

Cluster analysis is good for grouping data where there are many groups. Its limitations reside on the approaches to identify and define the clusters, as different methods can achieve different results. Due to this, this mechanism is of an exploratory nature, which means it is used to find relationships and group data in order to be used for something else. These facts mean it is clearly not suitable for video motion intensity classification in real-time.

4.1.2 Principal Component Analysis

Principal Component Analysis is a method that analyses a data set representing observations described by several dependent variables (Abdi and Williams, 2010), which are generally inter-correlated. It aims to extract the important information from the data table and to express this information as a set of new orthogonal variables called principal components. PCA also represents the pattern of similarity of the observations and the variables by displaying them as points in maps (Jolliffe, 2005) (Saporta and Niang, 2009).

The goals of PCA are divided in four steps. The first is the extraction of the most important information from the data set. The second involves the compression of the size of the data set by keeping only important information. The third allows for the simplification of the description of the data set. Finally, the fourth step is the analysis of the structure of the observations and the variables.

To fulfil these goals, PCA computes new variables called principal components which are obtained as linear combinations of the original variables (Abdi and

Williams, 2010). The first component is required to have the largest possible variance (inertia) which extracts the largest part of the inertia from the data table. The second component is computed under the constraint of being orthogonal to the first one, and to have the largest possible inertia. The remainder of the components are calculated likewise. The values that these new variables give are called factor scores and can be interpreted geometrically as the projections of the observations onto the principal components.

If we take into account a set of points in Euclidean space, the first component corresponds to a line that passes through the multidimensional mean and minimizes the sum of squares of the distances of the points from the line. The second component corresponds to the same process after all the correlation with the principal component has been subtracted from the points. PCA rotates the set of points around their mean in order to align with principal components, which moves as much of the variance as possible into the first two dimensions. The remaining values from the other dimensions can be discarded as they are small and represent a very small amount of information. PCA is therefore the optimal orthogonal transformation that keeps the subspace with the largest variance. However this comes at the price of great computational requirements when compared to discrete cosine transform.

The applicability of PCA is limited by certain constraints (Shlens, 2005) and its results depend on the scale of the variables. In turn this can cause problems to arise when the variables have different units. This means that they are not in an Euclidean space (Pearson, 1901), which makes PCA an arbitrary method of analysis.

PCA is good for finding patterns in data and to represent the data in a manner it highlights its similarities and differences. The main disadvantages of PCA are that it tends to reduce the data to its simpler, reduced form. Also it does not suit the necessity for a mechanism that can classify video in real-time according to its characteristics.

4.1.3 Neural Networks

Neural Networks or Artificial Neural Networks (ANN) (Abraham, 2005) in terms of computer science and artificial intelligence are computational models inspired by biological central nervous systems, which are able to go through the process of machine learning and pattern recognition. These networks are represented by neurons and their interconnections, where the inputs are computed into values by passing information through the network.

The artificial neurons or nodes, are the basic constitutive units of an artificial neural network and their function is to receive one or more inputs and to sum them to produce an output as shown in Figure 4.1.

Typically the sums of each node are weighted, and the resulting sum is passed through a non-linear function which is known as an activation function. A simple

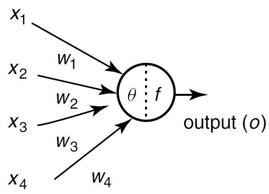

Figure 4.1: Artificial Neuron
Source: (Abraham, 2005)

representation of this type of model is one with three layers as represented in Figure 4.2. One for input with input neurons which send data via synapses to the second layer, usually called the hidden layer, and from there, the data is sent again via synapses to the third layer of output neurons.

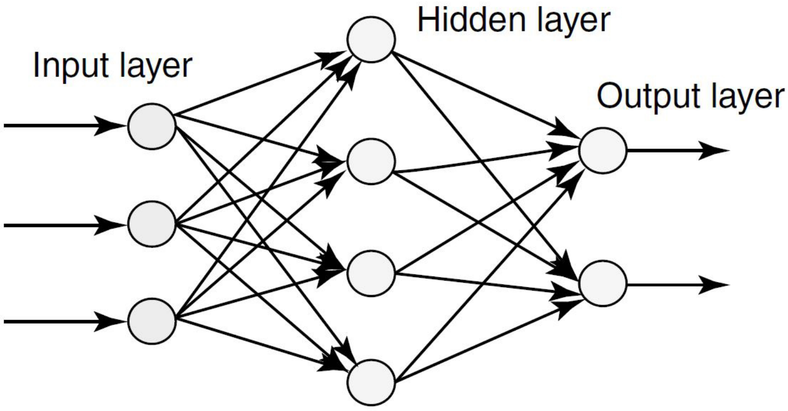

Figure 4.2: Artificial Neural Network
Source: (Abraham, 2005)

The parameters called weights are stored in the synapses, that manipulate the data in the calculations and serve the purpose of the neural network's learning process. Neural networks have the possibility of learning, going from a given task and a class of functions, they use a set of observations and find which optimally solves the task. For this a cost function must be defined in equation 4.1, such that for the optimal solution, no other solution has a lower cost.

$$C : F \to \mathbb{R} \tag{4.1}$$

$$f : X \to Y \tag{4.2}$$

The cost function is a measure of how far away a determined solution is from the optimal one. The choice of the cost function depends on the desired task. A neural network has to be configured in such a way that the use of a set of inputs produces the desired outputs. There are several methods to set the strengths of the connection between the nodes. One of these methods is to explicitly set the weights, using previous knowledge. Another way is to train the neural network by feeding it teaching patterns and letting it change the weights according to some learning rule. The three main sorts of learning tasks are Supervised Learning, Unsupervised Learning and Reinforcement Learning.

4.1.3.1 Neural Network Learning

In Supervised Learning a set of example pairs is given and the goal is to find a function as shown in Equation 4.2 that matches the examples. A commonly used cost function is the mean-squared error, which tries to minimize the average squared error between the network's output and the target value over all the example pairs. When the cost is minimized by gradient descent for the class of neural networks called multilayer perceptrons, the backpropagation algorithm for training neural networks is obtained. Supervised learning uses a function to provide continuous feedback on the quality of the solutions and is useful for pattern recognition, also known as classification and regression or function approximation.

In Unsupervised Learning only some uncategorised data is given as well as the cost function to be minimized, which can be any function of the data, and the output of the network. The system is supposed to discover statistically salient features of the input population. The cost function is dependent on the task and on the previously defined assumptions about the model or variables/parameters. The tasks that can be performed by unsupervised learning are estimation problems like clustering, statistical distributions, compression and filtering.

Reinforced learning is the task where no data is given, but is generated by an agent's interactions with the environment which is modelled as a Markov decision process. The agent performs a determined action in a given time, and the environment generates an observation as well as an instantaneous cost. The goal is to find a policy defined as a conditional distribution over actions given the observations for the choice of actions that minimize some measure of a determined long-term cost. If we take the environment and the policy together, their combination defines a system which we a Markov Chain. The tasks performed by reinforcement learning are control problems, games and sequential decision tasks.

These techniques result in an adjustment of the weighted connections between the neurons with the help of a modification rule. One of the most influential contributions was that of Hebb (Hebb, 2002), by presenting a theory of behaviour close to the physiology of a biological nervous system. The concept of learning emerged in the sense of how it occurred. It is based on the modification of the synaptic connections between neurons. Basically when an axon of a determined cell is near enough to another in order to excite it, and this second cell takes part continuously in firing it, a metabolic change or growth process takes place in one or both the cells in a way that the efficiency of the first cell in firing the second is increased. This process is then called Hebbian Learning, which is the base for most of neural networks' learning techniques. The point is that if two neurons are active at the same time, their connection must be strengthened. The downside of Hebbian Learning is that it continually strengthens the weights without bound if the input data is not normalized.

The perceptron is a single layer neural network with weights and biases can be trained to produce a correct target output when fed with the corresponding input (Raudys, 1998). The training technique for this is called the perceptron learning rule and this mechanism is particularly suited for simple problems in the classification of patterns. This procedure is very similar to Hebbian Learning with the only difference being that when the network responds correctly, no weights are modified. To build on the simplicity of these mechanisms comes backpropagation learning.

To build upon the simplicity of these mechanisms comes backpropagation learning. While the simple perceptron is not able to handle anything more than linearly separable or linearly independent problems. Backpropagation takes the negative of partial derivative of the error of the network and adds it to the weight. This decreases the error until it reaches a local minima. In other words, if the error is increasing with the increase of weights, a negative value must be added to the weight and vice versa. Because this process takes place from the output layer to the hidden layer and then to the input layer, this algorithm is called backpropagation learning.

Also related to training are two different modes in which it can be performed, namely online and batch modes. In the online mode the weight updates are computed for each input data sample, and the weights are modified after each of the samples. In batch mode, the weight update is computed for each sample but is then stored through the execution of the training set which is called an epoch. At the end of the epoch, all the contributions are added and the corresponding composite weight update is performed. This method follows the gradient more closely due to the cumulative weight update.

4.1.3.2 Types of Artificial Neural Networks

There are several types of Artificial Neural Networks inspired in the real life behaviour of neurons and the electrical messages they produce.

The Feed-forward Neural Network (Bebis and Georgiopoulos, 1994) is probably the most simple type or artificial neural network besides the perceptron. In this type of ANN, the data only moves in one direction which is forward, as there are no directed cycles or loops in the network. They can be built from the previously mentioned perceptron in a single-layer form and in a multi-layer form. In the multi-layer perceptron form, each neuron of a layer has connections to the neurons of the next layer. Usually the activation function in these types of units is a sigmoid function. The most popular learning technique in multi-layer networks is the backpropagation algorithm which was previously described. By using it, the network converges to a state where the error is minimized. The use of backpropagation is hindered if the number of training samples is small because the network can overfit the data and fail to capture the true statistical process generating the data. Another problem of backpropagation is the speed of convergence and the possibility of stalling in a local minimum of the error function.

In the Radial Basis Function Network (Lippmann, 1989) radial basis functions (RBF) are used as activation functions (Broomhead and Lowe, 1988). A RBF is a powerful technique for interpolation in multidimensional space in which it is a function that has built into a distance criterion with respect to a center. They can be used as a replacement to the sigmoidal hidden layer activation function of multi-layer perceptrons in feed-forward networks. The RBF network has two layers of processing. In the first, the input is mapped onto each RBF in the hidden layer, usually through a Gaussian distribution. When they are used in regression problems, the output layer is a linear combination of hidden layer values representing mean predicted output, where the output value is interpreted in the same way as a statistics' regression model. In both cases the performance is improved by a shrinkage technique known as ridge regression which corresponds to using small parameter values to produce smoother output functions in a Bayesian framework.

These networks do not suffer from the local minima problem in the same way as the multi-layer perceptrons, because the only parameters that are adjusted in the learning process are the linear mappings from the hidden layer to the output layer. The disadvantage of RBFs is that they require good coverage of the input space by the radial basis functions. This is because RBF centres are determined with reference to the distribution of the input data but without reference to the prediction task, which leads to a waste of representational resources on irrelevant areas of the input space for the learning task.

Recurrent Neural Networks are models with bi-directional data flow where data is also propagated from later processing stages to previous stages. They can follow several architectures, such as the fully recurrent network, the Hopfield network (Hopfield, 1982), the Elman network (Elman, 1990) and Jordan network (Jacobs and Jordan, 1993) and more. The fully recurrent network is the basic architecture where each neuron has a directed connection to every other neuron as shown in Figure 4.3. Each of the neurons has a time-varying real-valued activation and each connection has a modifiable real-valued weight. Some of the nodes are for input and output and the rest are called hidden nodes. For supervised learning training sequences of real-valued inputs become sequences of activations for the input nodes, one input at a time. There can be teacher-given target activations for some of the output units at certain points in time. At any time-step the non-input units can compute their current activation as a non-linear function of the weighted sum of the activations of all units from which it receives connections.

The error for each sequence is the sum of the deviations of all target signals from the corresponding activations computed by the network, which for a training set with numerous sequences, corresponds to the total error being the sum of the

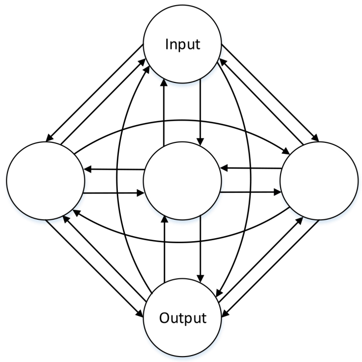

Figure 4.3: Fully Recurrent Neural Network

errors of individual sequences. For reinforcement learning there are no teacher-given targets. Instead, a fitness function or reward function is used to evaluate the performance of the RNN, which influences the input stream through output units connected to actuators affecting the environment. To minimize the total error, gradient descent can be used to change each weight in proportion to its derivative with respect to the error, granted that the activation functions are differentiable.

Random Neural Networks are simplified models of biophysical neural behaviour described in (Kandel et al., 2000), in which the neurons exchange spiking signals to modify an integer value in the neuron. At any time t, the state of a neuron i can be described by its signal potential $k_i(t)$ which is the integer associated with the added positive signals received by the neuron. A neuron is excited if $k_i(t) > 0$ and is idle or quiescent if $k_i(t) = 0$. Equation 4.3 shows the excitation probability of a neuron.

$$q_i(t) = Pr[k_i(t) > 0] \leq 1 \tag{4.3}$$

The random neural network is a recurrent model in the way that it is allowed to have complex feedback loops. Some interesting features of the random neural network are the closer representation of the manner in which signals are transmitted in a biophysical neural network where they travel as spikes rather than as fixed analogue signals. Also, they are easy to simulate and implement, as each neuron can be simply represented by a counter. Also, the neuron potential is represented as an integer which leads to more detailed information on system state, since a neuron is in a firing state if the potential is positive. The spikes can originate from outside the network or they can come from other cells in the network. Cells with a positive excitatory state can randomly send out spikes according to the exponential distribution presented in equation 4.4 with rate r_i. If a neuron fires, its total input potential is reduced by one. The signal can reach another neuron as a positive signal with probability $P^+(i,j)$ or as a negative signal with probability $P^-(i,j)$, or it departs from the network with probability $d(i)$. These probabilities must sum up to equation 4.5. When a neuron is excited it fires positive and negative signals with the rates shown in equations 4.6 and 4.7.

$$r_i = (d(i))^{-1} \sum_{j=1}^{N} [w^+(i,j) + w^-(i,j)] \tag{4.4}$$

$$\sum_{j=1}^{N} [p^+(i,j) + p^-(i,j)] + d(i) = 1, \forall i \tag{4.5}$$

$$w^+(i,j) = r_i p^+(i,j) \geq 0 \tag{4.6}$$

$$w^-(i,j) = r_i p^-(i,j) \geq 0 \tag{4.7}$$

The signals that arrive from the outside of the network are represented by Poisson processes of rates Λ_i and λ_i. The neuron's potential rises by one if it receives an excitatory spike, and drops by one if it receives an inhibitory spike. Nevertheless if the value is zero, receiving an inhibitory spike has no effect on the signal.

At a time t, the state of the network is described by the vector of signal potentials $\mathbf{k}(t) = [k_1(t), ..., k_N(t)]$. The stationary probability distribution can be described by the steady-state Chapman-Kolmogorov equations 4.8 for continuous time Markov chain systems.

$$\begin{aligned} &\pi(\mathbf{k}) \sum_{i\ =1}^{N} [\Lambda_i + (\lambda_i + r_i)\mathbf{1}_{k_i<0}] \\ &= \sum_{i=1}^{N} \Bigg\{ \pi(\mathbf{k}_i^+) r_i d(i) + \pi(\mathbf{k}_i^-)\Lambda_i \mathbf{1}_{k_i>0} + \pi(k_i^+)\lambda_i \\ &\qquad + \sum_{j=1}^{N} \left[\pi(k_{ij}^{+-} r_i p^p(i,j)\mathbf{1}_{k_j>0} + \pi(k_{ij}^{++} r_i p^-(i,j) + \pi(k_i^+) r_i p^-(i,j)\mathbf{1}_{k_j=0}\right] \Bigg\} \end{aligned} \tag{4.8}$$

The values of the stationary parameters of the network, the stationary excitation probabilities and the stationary probability distribution, are derived from the total arrival rates of the positive and negative signals $\lambda^+(i)$ and $\lambda^-(i)$, $i = 1, ...N$ that are given by the equations found in the theorem (Gelenbe, 1989) from 4.9 through 4.11.

$$\lambda^+(i) = \Lambda_i + \sum_{j=1}^{N} r_j q_j p^+(j,i) \tag{4.9}$$

$$\lambda^{-}(i) = \lambda_i + \sum_{j=1}^{N} r_j q_j p^{-}(j,i) \tag{4.10}$$

$$q_i = min\left\{1, \frac{\lambda^{+}(i)}{r_i + \lambda - (i)}\right\} \tag{4.11}$$

The theorem shows that whenever a solution to the signal flow can be found such that $q_i < 1, \forall i$, therefore the stationary joint probability distribution of the network has the simple product form of equation 4.12 associated with the marginal probabilities from each neuron, $\pi_i(k_i)$.

$$\pi(\mathbf{k}) = \prod_{i=1}^{N} \pi_i(k_i) = \prod_{i=1}^{N} (1 - q_i) q_i^{k_i} \tag{4.12}$$

Thus, the condition $q_i < 1$ is a stability condition that guarantees that the excitation level of each neuron remains finite with probability one. The product form implies that the neurons are independent even though they are coupled through the exchanged signals. The network is stable if the signal potential of each neuron does not tend to increase without bounds. In order words, stability is guaranteed if a unique solution exists for the non-linear system of equations 4.9- 4.11 and $q_i < 1, \forall i$. Furthermore, if a solution exists with $q_i < 1, \forall i$ then it is unique (Gelenbe, 1990). This is shown by taking equations 4.9- 4.10 and eliminating the q_i terms and combining them to obtain equation 4.13.

$$\boldsymbol{\lambda}^{-} - \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+}\mathbf{FP}^{-} = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{FP}^{+})^{-1}\mathbf{FP}^{-}\boldsymbol{\lambda}^{-}, \boldsymbol{\lambda}^{+}, \boldsymbol{\lambda}, \boldsymbol{\Lambda}, \in \mathbb{R}^{1\times N} and \mathbf{I}, \mathbf{F}, \mathbf{P}^{\pm} \in [0,1]^{N\times N} \tag{4.13}$$

$\boldsymbol{\lambda}^{\pm}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\lambda}$ are vectors that represent the total and external arrival rates of excitatory-inhibitory signals, $\mathbf{P}^{+}$ and $\mathbf{P}^{-}$ are square matrices whose elements are the transition probabilities $p \pm (i,j)$, $\mathbf{I}$ is the identity matrix and $\mathbf{F}$ is a diagonal matrix with elements $f_{ii} = r_i/(r_i + \lambda^{-}(i)) \leq 1$. Because $\mathbf{P}^{+}$ is sub-stochastic and all elements of $\mathbf{F}$ are smaller than 1, the series $\sum_{m=0}^{\infty}(\mathbf{FP}^{+})^m$ is geometrically convergent we have equation 4.14.

$$(\mathbf{I} - \mathbf{FP}^{+})^{-1} = \sum_{m=0}^{\infty} (\mathbf{FP}^{+})^m \tag{4.14}$$

By defining $y = \boldsymbol{\lambda}^{-} - \boldsymbol{\lambda}$ the system can be written in the fixed point form shown in equation 4.15.

$$y = g(y) = \sum_{m=0}^{\infty} (\mathbf{FP}^{+})^{m} \mathbf{FP}^{-} \tag{4.15}$$

The implication of this is that according to Brouwer's fixed point theorem, equation 4.15 has at least one fixed point solution. In this case, exactly one point must exist y^* since solution uniqueness has already been established, therefore a solution to equations 4.9- 4.11 always exists and is unique.

With respect to learning in the random neural network, Gelenbe proposed an approach based on gradient descent supervised learning, where the training examples represented by k are sequentially processed and the weights of the network are updated according to the gradient descent rule until a minimum of the error function is reached. Assuming that the error function to be minimised is a general quadratic function as shown in equation 4.16 where E_k is the error function of the kth input-output pair, $c_i \in 0, 1$ shows whether neuron i is an output neuron and $f_i(q_{ik})$ is a differentiable function of neuron i.

$$E = \sum_{k=1}^{K} E_k = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{N} c_i (f_i(q_{ik}) - y_{ik})^2 \tag{4.16}$$

The generic term $w(u, v)$ denotes either $w^{+}(u, v)$ or $w^{-}(u, v)$ and the rule for updating the weights using the $k - th$ input-output pair at step $(\tau + 1)$ is defined by equation 4.17.

$$w_{\tau+1}(u, v) = w_{\tau}(u, v) - \eta \left[\frac{\partial Ek}{\partial w(u, v)} \right]_{\tau} \tag{4.17}$$

The partial derivative of the error function with respect to $w(u, v)$ can be calculated based on equation 4.16 and results in equation 4.18, where $[]_\tau$ defines that all the calculations are performed using the weight values of step τ and the q_{ik} values are derived from solving equations 4.9- 4.11 when the current weights $(w_\tau(u, v))$ are used.

$$\left[\frac{\partial Ek}{\partial w(u, v)} \right]_{\tau} = \sum_{i=1}^{N} c_i (f_i(q_{ik} - y_{ik}) \times \left[\frac{\partial f_i(q_i)}{\partial q_i} \frac{\partial q_i}{\partial w(u, v)} \right]_{\tau} \tag{4.18}$$

Gelenbe reached equation 4.19 which is the derivation of a closed expression for the term $[\partial q_i / \partial w(u, v)]_\tau$ from equations 4.17- 4.18 that depends on the nonlinear system of equations 4.9- 4.11.

$$\frac{\partial \mathbf{q}}{\partial w(u, v)} = \gamma(u, v)(\mathbf{I} - \mathbf{W})^{-1} \tag{4.19}$$

The steps for the gradient descent learning algorithm are then the following: Firstly, initialize the weights $w^{+}(u,v)$ and $w^{-}(u,v)\forall u,v$ and appropriately choose the learning rate η. Secondly for each successive input-output pattern k, initialize Λ_{ik} and λ_{ik} according to x_{ik} which are the training values, then solve the system of equations 4.9- 4.11 using the current weight values. Based on the values obtained, calculate $\mathbf{W}$, $\gamma^{+}(u,v)$ and $\gamma^{-}(u,v), \forall u,v$. Calculate $[\partial\mathbf{q}/\partial w^{+}(u,v)]_{\tau}$ and $[\partial\mathbf{q}/\partial w^{-}(u,v)]_{\tau}$ according to equation 4.19. Update the weights from equations 4.17- 4.18. To satisfy the weight non-negativity constraint either the negative values can be set to zero or the iteration can be repeated with a smaller value of η. The last step implies the repetition of the second step until convergence is reached.

RNNs are used in (Mohamed et al., 2004) for evaluation of the quality of real-time speech in a packet network. In (Mohamed and Rubino, 2002) they are used to quantify the quality of a video flow. The results from that study show that they correlate well with the human perception.

Neural Networks in general are useful for video classification. They can be trained with a determined set of training videos from different classes so that they produce results that distinguish each video class accurately. Then, they can be used with a random video that needs to be can classified and identify the class it belongs from according to the training.

Neural networks are highly successful in pattern recognition and classification problems. This results from their capability for learning and generalization. By training the network successfully with an adequate range of video samples, we can afterwards use it to classify a given video file or part according to the complexity of the movement present in the scene. This is useful as we can then feed this information to the adaptive FEC mechanism so it can define the amount of redundancy needed for optimal QoE according to the network conditions.

4.2 Adaptive Forward Error Correction

Static FEC mechanisms use a fixed amount of redundancy for error protection. This means that every video unit has the same protection as any other. While this may seem fair, in reality it results in poor usage of the available and sometimes scarce network resources. Not every video unit needs the same quota of error protection due to the different characteristics of the video units or the state of the network. The excessive and indiscriminate use of redundancy can ultimately lead to network congestion. This can be overcome if an adaptive FEC mechanism is used based on optimizing the amount of redundancy to fit the needs of the characteristics of the video and of the state of the network. Some mechanisms that can be used to optimize the error protection allocation problem can be Game Theory, Heuristics and Ant Colony Optimization.

4.2.1 Game Theory

Game Theory (GT) or interactive decision theory, is presented in (Myerson, 2013) as the study of mathematical models of conflict and cooperation between intelligent rational decision-makers. When game theory was invented, it was used for zero-sum games where one of the player's gains are exactly equal to the net losses of the other players. The Nash Equilibrium (NE) is a strategic equilibrium discovered by Nash for non-cooperative games. The NE states that if each player has chosen a strategy and no player can benefit by changing strategies while the other players keep theirs, then the current set of strategy choices and the corresponding payoffs constitute a NE (Nash, 1951). GT was initially used in Economics (Von Neumann and Morgenstern, 1944) to understand a large set of behaviours, but has expanded to Social Sciences (Krishna, 2009), Biology (Hofbauer and Sigmund, 1998), Political Science (Morris et al., 1986) and Computer Science (Roughgarden, 2004).

The games which are studied in GT are well-defined mathematical objects. To fully define a game, we must specify the players of the game, the information and actions available to each player at a determined decision point and the gains for each outcome. These elements are used in conjunction with a solution concept to construct a set of equilibrium strategies for each player such that, when they are employed, a player can not profit by unilaterally deviating from their strategy. These equilibrium strategies determine a stable state in which either one outcome occurs or a set of outcomes occur with known probability. There are several forms in which games can be represented like the extensive and normal form which are used to define noncooperative games and the characteristic form function which is used to define most of the cooperative games.

The extensive form is used to formalize games with a time sequencing of moves and can be seen as a multi-player generalization of a decision tree (Fudenberg and Tirole, 1991). Games are played on trees where each node represents a point of choice for a player. The player is specified by a number listed on the node, and the lines out of the node represent a possible action for that player. The payoffs are specified at the bottom of the tree.

The normal form is usually represented by a matrix which shows the players, strategies and payoffs. If we generalize, this form can be represented by any function that associates a payoff for each player with every possible combination of actions. In this strategy, one player chooses the row and the other chooses the column. There are as many strategies in the game for each player as there are rows or columns respectively, whose payoffs are represented in numerical pairs inside the matrix. In normal form games it is presumed that each player acts simultaneously or without knowing the actions of the other.

In games which have removable utility separate, the rewards are not given and the characteristic function decides the payoff of each unity, where if a unity is empty, it does not receive any reward. A characteristic function is seen as (N, v), where N represents the group of people and $v : 2^N \to \mathbf{R}$ is a normal utility.

Games in game theory can be classified according to various types (Osborne, 1994):

- Cooperative, if the players are able to form binding commitments, or non-cooperative, if not. The latter can model games to the finest detail.
- Symmetric when payoffs for a particular strategy depend only on the other strategies used with no regard for the player who is playing them, and asymmetric when the strategy sets are not identical for both players.
- Zero-sum when the games do not increase or decrease the available resources and the sum of all the benefit to all players always adds to zero, this is what one player wins, another player must loose. In non-zero-sum, a gain by one player does not necessarily imply a loss by another.
- Simultaneous are those games where players move at the same time or do not have knowledge of other players movements, and sequential games are those where one plays with some knowledge about previous actions.
- Perfect information if all the players know the moves previously made by all other players and imperfect information if there is some omission of information about previous moves.
- Combinatorial games refer to those where finding an optimal strategy involves exploring the multiplicity of possible moves.

Game Theory is very useful if we have a competitive problem that needs to be solved where it brings to light the concept of bargaining. However, for dynamic optimization problems with a large number of variables, its problem solving techniques have difficulties when the matrix of pay-offs is large which makes the process very complex.

4.2.2 Heuristics

In Computer Science a heuristic is a technique designed for solving a problem when basic deterministic global optimization approaches fail (Gilli and Winker, 2009). Although aiming at high-quality solutions, due to the intrinsic complexity of the problems and stochastic elements of the algorithms, they cannot pretend to produce the exact solution in every case with certainty.Nevertheless, a stochastic high quality approximation of a global optimum is probably more valuable than a deterministic poor-quality local minimum provided by a classical method or no solution at all. Generally a heuristic can be based on the properties of an algorithm (Winker and Gilli, 2004). Firstly, a heuristic should be able to provide high-quality stochastic approximations to the global optimum at least when the amount of computational spent on a single run of the algorithm or repeated runs is increased. Secondly, a well behaved heuristic should be robust to changes in problem characteristics, this is, it should not fit only one problem but the whole class. Therefore, a heuristic should be easily implemented to many problem instances including new ones. Finally, a heuristic might be stochastic, but should not have any subjective elements.

Heuristic optimization methods can be divided into two broad classes, namely, construction methods and local search methods. Construction methods determine a tour according to some construction rules, but they do not try to improve upon this tour. A tour is successively built and parts already built remain unchanged through the algorithm. Local search uses only information about the solutions in the neighbourhood of a current solution, and for this it is very similar to hill climbing, where the choice of a neighbour solution locally maximizes a criterion. Local search methods can be divided into trajectory methods which work on a single solution and populations-based methods or evolutionary algorithms.

In trajectory methods, the concept of neighbourhood is very important and depends on the problem under consideration. It can be challenging to find efficient neighbourhood functions that lead to high-quality local optima.

Simulated annealing (Van Laarhoven and Aarts, 1987) is a method which is a refinement of local search which is based on a parallelism between combinatorial optimization and the annealing process of solids. This method is similar to local search but an improvement of the solution for a move is always accepted. The algorithm also accepts a move uphill but only with a given probability. This probability depends on a parameter called temperature that is gradually decreasing during the process.

Tabu search (Glover et al., 1997) is particularly designed for the exploration of discrete search spaces where the set of neighbour solutions is finite. This method implements the selection of a neighbourhood solution in a way to avoid cycling. To do this, a short-term memory is employed, known as the tabu list and which contains the solutions that were most recently visited.

Population-based methods work simultaneously on a whole set of solutions called the population. They are more efficient in exploring the whole search space but have a higher computational cost and more complex structures.

Genetic algorithms (Holland, 1975) (Eiben et al., 1994) are a technique that imitates the evolutionary process of species that reproduce sexually. New candidates for the solution are generated by an evolutionary mechanism called crossover, where part of the genetic composition from each parent is combined and then a random mutation is applied. If the result has good characteristics, it will have a higher probability of being used.

Differential Evolution (Storn and Price, 1997) is a population-based heuristic optimization technique for continuous objective functions. The algorithm updates a population of solution vectors by addition, subtraction and crossover, and then selects the fittest solutions among the original and updated population.

All of the previously presented strategies are a type of heuristic called metaheuristic. A metaheuristic is a set of algorithmic concepts that can be used to defined heuristic methods applicable to a wide set of different problems, this is, a metaheuristic is a general-purpose heuristic method designed to guide an underlying problem-specific heuristic toward promising regions of the search space containing high-quality solutions. In this way, a metaheuristic is a general algorithmic framework which can be applied to different optimization problems with relatively few modifications to make them adapted to a specific problem, which brings us to the Ant Colony optimization metaheuristic.

4.2.3 Ant Colony Optimization

Ant Colony optimization (ACO) is a metaheuristic that was first introduced in 1996 by Marco Dorigo's Ant System (Dorigo et al., 1996). It has been extensively developed and is the main focus of this section (Dorigo and Stutzle, 2004).

ACO works as a colony of artificial ants that cooperate in finding good solutions for discrete optimization problems. The computational resources are allocated to a set of simple agents that communicate indirectly through the environment, where good solutions emerge through the cooperative interaction. These algorithms can be used to solve both static and dynamic combinatorial optimization problems.

In ACO an artificial ant is a stochastic constructive procedure that incrementally builds a solution by adding opportunely defined solution components to a partial solution under construction. This makes it applicable to any combinatorial optimization problem for which a constructive heuristic can be defined. The artificial ants build solutions by performing randomized walks on the completely connected graph $G_C = (C, L)$ whose nodes are components C, and the set L fully connects the components C. To the G_C graph we call construction graph and to the elements of L we call connections. In most cases, the ants build acceptable solutions, but sometimes, it may be necessary or advantageous to let them explore solutions that are least plausible. To the components $c_i \in C$ and connections $l_{ij} \in L$ can be associated a pheromone trail τ (τ_i if associated with components and τ_{ij} if associated with connections), and a heuristic value η (η_i and η_{ij} respectively). The pheromone trail serves the purpose of encoding a long-term memory about the entire ant search process, and is updated by the ants themselves. The heuristic value or heuristic information represents previously known information, generally the η represents the cost or an estimate of the cost, of adding the component or connection to the solution under construction.

Each ant in the colony has the following properties: It explores the construction graph G_c to search for optimal solutions; It has a memory $\mathcal{M}^k$ that can be used to store information about the path followed so far; It has a start state x_s^k and one or more termination conditions e^k; If the ant is in a state $x_r = \langle x_{r-1}, i \rangle$, if no termination condition is satisfied, it moves to a node j in the neighbourhood $\mathcal{N}^k(x_r)$. If at least one of the termination conditions e^k is satisfied, the ant stops; It selects a move by applying a probabilistic decision rule, which is a function of, firstly the locally available pheromone trails and heuristic values, secondly the ant's private memory storing its current state and finally the problem constraints; When a component c_j is added to the current state, the ant can update the pheromone trail τ associated with it or the corresponding connection; Once a solution is created, the ant can retrace the same path backward and update the pheromone trails of the used components.

Ants in the solution finding process act simultaneously and independently, nevertheless each ant is capable to find a solution to the problem, but it is only through the collective interaction of the whole colony that good-quality solutions emerge. This is obtained through the indirect communication mediated by the information the ants read and write to in the variables storing pheromone values. Therefore, one can say that this is a distributed learning process in which the single agents, are not adaptive themselves but, adaptively modify the representation and perception of the problems for the other ants.

Regarding the ACO algorithm, it can be seen as the interactions of three distinct processes. Specifically, there is a process through which a colony of ants is managed and whose ants concurrently and asynchronously visit the adjacent nodes of the considered problem graph G_c. They travel by means of stochastic local decision rules that make use of pheromone trails and heuristic information. Through this, a solution to the optimization problem is incrementally built by the ants. When an ant has a solution, or even while the solution is being constructed, the ant can evaluate the partial solution that will be used by the process that updates the pheromones to decide how much pheromone it must deposit.

The pheromone updating process modifies the pheromone trails. The value of the trails can both increase or decrease. It can increase through the ants depositing of pheromone on the components connections they use and it can decrease through pheromone evaporation. The increase of pheromone results in an increase of the probability that the components or connections that were used by one or many ants and that have produced a very good solution will be used again by other ants. The evaporation of the pheromone serves the purpose of avoiding a too sudden convergence to a suboptimal solution and favours the exploration of new areas of the search space.

The third and last process is that of performing operations that cannot be performed by individual ants. These tasks comprise the activation of a local optimization procedure, or the collection of global information that can be used to decide whether or not it is useful to intensify the quantity of pheromone in order to somewhat bias the search process.

ACO is an optimization methodology but it is already used to solve problems from several areas that reach from network routing problems (Varela and Sinclair, 1999), to assignment (Lourenço and Serra, 2002), scheduling (Blum, 2002), machine learning (Parpinelli et al., 2002) and more.

ACO can adaptively change the solution's value according to the viability of the path, by taking into account the path and node weights. Also, its ability to run in real time, iteration by iteration, makes it suitable for use in a situation where there are several paths with changing conditions. For example, if we express the video and network characteristics as a series of paths in a graph, we can run the algorithm one step at a time and if the video or the network change, it will adapt to it. This mechanism will be presented with more emphasis on detail in the next chapter which describes the interactions between the parameters and the working of the algorithm.

4.3 Summary

In this section we presented some mechanisms like Cluster Analysis, Principal Component Analysis and Neural Networks which could serve the purpose of the classification of video sequences, more specifically the motion intensity in these video sequences. Also several techniques that could provide for the optimization of the redundancy attribution of the adaptive FEC mechanism were reviewed, namely Game Theory, Metaheuristics and Ant Colony Optimization which is a type of Metaheuristic.

Chapter 5

Advances Towards Video Transmission Optimization: neuralFEC

This document has presented several approaches to improve video transmission resilience in adverse network conditions. Even though some of them offer worthy contributions, such as video protection in a bursty error environment or FEC allocation according to video characteristics, they fail to satisfy both at the same time in a dynamic and adaptive way. The presented mechanism fulfils the goal of having dynamic FEC protection according to the characteristics of the video.

The next sections will describe the objectives of the neuralFEC mechanism, its specification and the results obtained with it.

5.1 Objectives

The neuralFEC mechanism's main objective is to provide FEC protection according to the motion intensity characteristics of a frame in the video sequence. The challenge resides in the adjustment of the quantity of redundancy according to the changing characteristics of the scenes in the video. For example, a video with high motion complexity will have more redundancy allocated compared to a video with low motion complexity. Therefore, the most QoE sensitive frames which are the ones with a greater amount of movement will be better protected.

In the end, by fulfilling this objective, the mechanism can achieve a reduction of the impact of packet loss on the video quality while using a lesser amount of redundancy. Through this, it is possible to distribute the redundancy in a way that will improve the QoE for the end user while sparing network resources.

5.2 neuralFEC Mechanism Specification

In order to achieve the aims of the proposal, there is the need for a video motion intensity classification mechanism. The neuralFEC mechanism takes into account the characteristics of the video and decides the amount of redundancy to be allocated through the use of an error correcting code (ECC), for a given video frame at a given time. To do this, we resort to the use of a Random Neural Network (RNN) (Kandel et al., 2000). The RNN mechanism was selected, because of its flexibility and ease of use, which allow it to be used for real-time video classification (Aguiar et al., 2012) (Mohamed and Rubino, 2002).

Figure 5.1 shows a general overview of the neuralFEC mechanism. Firstly, the off-line training and validation of the RNN created for motion intensity categorization purposes was performed. After completing its training, the RNN is ready to be used to categorize the videos off-line (i.e., previous to transmission).

Based on the output given by the RNN, the value which represents the amount of redundancy to be used by the Reed-Solomon (Reed and Solomon, 1960) algorithm is computed in real-time for each frame. This is done in a way which minimizes network overhead and maximizes Quality of Experience (QoE). After the video has the redundancy applied to it, the transmission occurs.

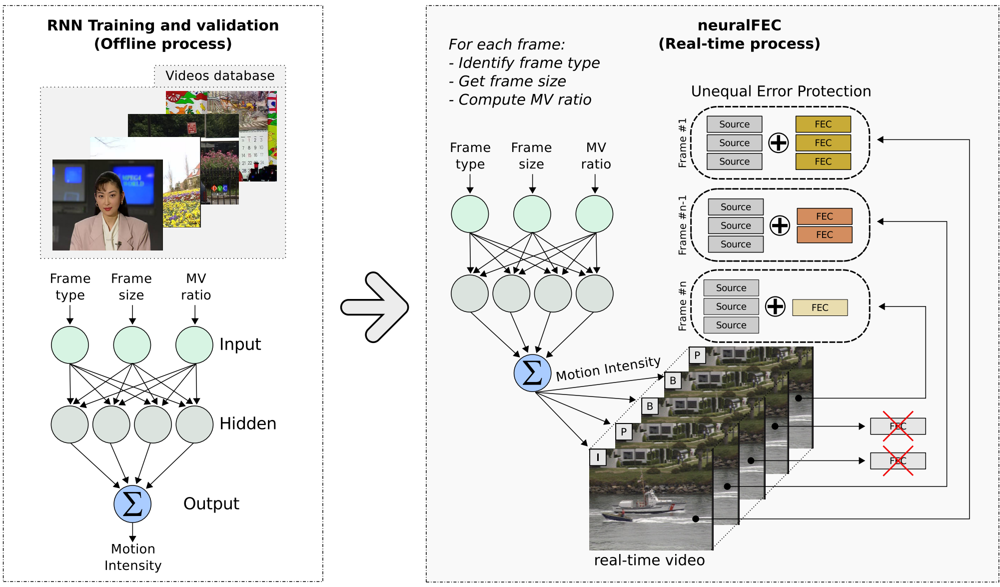

FIGURE 5.1: Stages of the neuralFEC

Regarding the hierarchical cluster analysis exploratory study, data concerning each individual video collected from the analysis of the trace files obtained for use with Evalvid was fed to the clustering process. The information collected consisted in the sum of the byte sizes of each type of frame, and a ratio r given by Equation 5.1 where the upper part of the equation represents the sum of the number of motion vectors and the lower part of the equation represents the sum of the distance those vectors travel from all the frames in the video.

$$r = \frac{\sum_{i=0}^{ttlfrms-1} MV_i}{\sum_{j=0}^{ttlfrms-1} distMV_j} \tag{5.1}$$

The R Project for Statistical Computing software (R Core Team, 2013) commonly known as just R was used to perform the cluster analysis. The clustering method used was the Ward method (Mojena, 1977) (Ward Jr, 1963), which minimizes the total within-cluster variance. At each step of the process the pair of clusters with minimum between-cluster distance are merged. Thus the lines of the dendogram represent the dissimilarity between each of the video files. The longer the path between them, or the farther away they are, the more different the videos are in terms of motion intensity and the inverse is also valid, the shorter the path, the more similar the videos are.

The dendogram represented in Figure 5.2 shows the videos grouped according to degrees of motion, which corresponds to resulting classifications present in the literature (Khan et al., 2010). Video characteristics such as frame size, frame type and motion vectors were used during the analysis. From this result, a training set representing video sequences of low motion intensity (Akiyo and Silent), medium motion intensity (City and Football) and high motion intensity (Mobile and Flower) was selected. This result implies that even in the form of a preliminary classification of the videos it is possible to build sets of videos grouped by their motion characteristics. Not only that, but it serves as a proof of concept of the RNN video motion intensity classification mechanism, showing that it is indeed possible to use the aforementioned parameters in the RNN for real-time frame-by-frame video motion intensity classification.

The training of the RNN consisted in feeding the information of this set of selected videos to the inputs of the network for about 600 iterations which was the point at which the Minimum Mean Squared Error (MSE) stabilized. During the training process the RNN's weights adjust themselves, such that the output corresponds to the expected values. After the training process is completed, the network is validated with a different set of video sequences. Afterwards, the RNN can be used for motion intensity classification in distinct scenarios. The training and validation steps of the RNN were performed in the RNNSimv2 module (Abdelbaki, 1999) within MATLAB (MATLAB, 2013). The RNN, which is used to classify all the video sequences in the evaluation of the mechanisms is the same due to the fact that it was trained with different video sequences with different characteristics. Therefore, it is able to correctly classify the motion intensity of the individual frames of the video sequences.

For each frame, the RNN receives as input the frame size, the frame type, the number of motion vectors present and the distance travelled by the motion vectors in the form of the ratio 5.1. These depict the three input neurons of the RNN. It processes these inputs and outputs a numerical value through a single output neuron as shown in Figure 5.3, which represents the classification of the frame in analysis according to a complexity scale defined during the training stage.

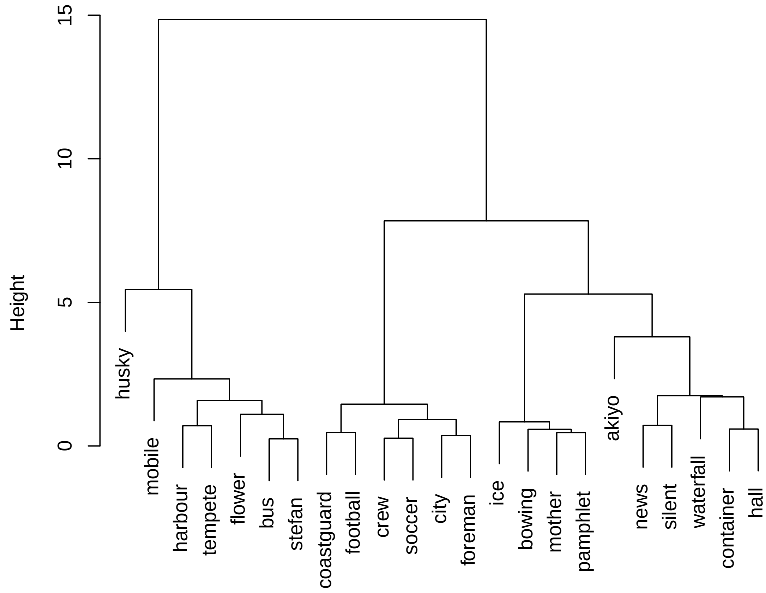

FIGURE 5.2: Cluster Analysis Results

The RNN mechanism has several features that make it suitable for use as a video motion intensity classification tool, such as: a probability distribution that is easily computed by an analytical equation; the learning algorithm has low complexity and strong generalisation capacity; the neuron potential is described by an integer instead of a binary value; it represents the signals of a biological neural network better than other Artificial Neural Networks.

In order to have a working FEC mechanism, an Error Correcting Code is needed. Reed-Solomon was the ECC selected in detriment of Low Density Parity Check (LDPC) (MacKay et al., 1995) and Turbo Codes (Berrou et al., 1993) due to its characteristics, such as the ability to use original data blocks of arbitrary sizes, which makes it very flexible, and its simplicity, which allows its modification to comply with the requirements of this study. Also, there is some previous familiarity with the Reed-Solomon code by part the advisors which provided additional support in terms of particularities regarding the encoding and decoding process. Hence, it was integrated with the Evalvid (Klaue et al., 2003) and NS-3 (Riley and Henderson, 2010) tools with the help of an external mathematical library with ECCs, namely IT++ (IT+). This required the use of data structures specific to the library in order to feed the encode and decode functions of the ECC. Therefore, extensive modifications had to be done in the Evalvid scripts in the functions that handle packet creation and processing so that the data from the encoder function could be translated to a packet and vice-versa for the decoder function.

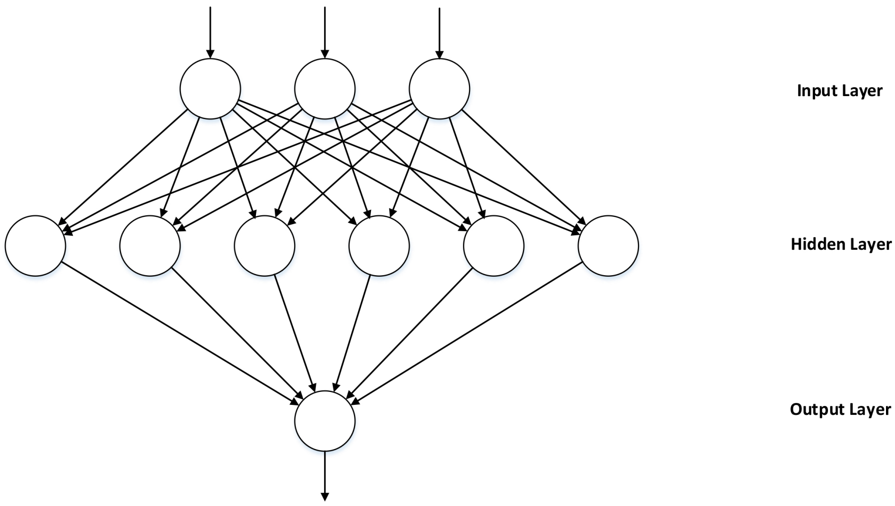

FIGURE 5.3: Random Neural Network Structure

During the exploratory analysis of the Error Correcting Codes it became evident that the information from the generated video trace files for Evalvid had to suffer modifications prior to encoding. Thus, the video trace files for Evalvid consist of lines each describing a frame. Each frame has a different size, which conflicts with the implementation present in the IT++ library. The implementation requires that the size of each data unit to be encoded is of the form present in Equation 5.5, where n is given by Equation 5.2, and with m being the exponent of the alphabet size of the code factor. The *iterations* parameter is given by Equation 5.4 where *iterations* is the number of iterations of the code, *encbitssize* is the number of bits of the frame to be encoded, and k is the parameter described by Equation 5.2, involving the m and t parameters of the Reed-Solomon code.

$$n = 2^m - 1 \tag{5.2}$$

$$k = 2^m - 1 - 2 * t \tag{5.3}$$

$$iterations = encbitssize/k * m \tag{5.4}$$

$$pktsize = iterations * n * m \tag{5.5}$$

Therefore, by taking the frame sizes of the video trace files used and by adding some bytes to them (i.e., perform padding), the frame sizes become a multiple number of the ratio described in Equation 5.4. This is required for correct and successful decoding of the received packet or else the decoded packets would be smaller than the original packets by the remainder of bytes that exceed the ratio.

5.3 neuralFEC Mechanism Evaluation

This section presents the evaluation of the neuralFEC mechanism.

5.3.1 Evaluation Objectives

The main goal of this evaluation is to assess if the neuralFEC mechanism can perform an adaptive protection of different video sequences according to their motion intensity in a lossy network scenario. The performance of the neuralFEC mechanism is assessed in terms of redundancy overhead and QoE.

Regarding the redundancy overhead used by the adaptive neuralFEC mechanism, it was measured from the data which is added to the original packets being transmitted. This provides a means of comparison between different protection mechanisms on how different video sequences with different motion intensities affect the allocation of redundancy.

The performance of neuralFEC according to a human user viewpoint was assessed by objective QoE metrics. The metric chosen for this assessment was Structural Similarity (SSIM) (Wang and Bovik, 2002). This objective evaluation metric is a method based on the analysis of luminance, contrast and structural similarity of images. SSIM is one of the most commonly used metrics for objective evaluation of QoE (Chikkerur et al., 2011). The SSIM was assessed through the Moscow State University (MSU) Video Quality Measurement (VQM) (Vatolin et al., 2011) tool. The Moscow State University (MSU) Video Quality Measurement Tool (VQMT) is a software for objective video quality assessment. It provides functionality for both full-reference and single reference comparisons.

5.3.2 Evaluation Scenario

The video files used for the experiments are encoded in the MPEG-4 part 14 standard (ISO/IEC, 2003). Each video was encoded following the same parameters: a resolution of 352x288 pixels; a GoP size of 19; two B frames after an I-frame or P-frame.

In order to use the video files in the simulation environment, the Evalvid tool is used for processing of video traces during transmission and reception.

Evalvid is a framework and tool-set for the evaluation of the quality of video transmitted over a real or simulated communication network. It can measure the QoS parameters of the underlying network as well as it can provide standard video quality metrics like PSNR and SSIM. This is done by taking video traces that represent the original video. Those traces are then used to create dummy packets and a sent packet trace file. The dummy files are received and added to a received packet trace file. When the transmission ends, Evalvid attempts to rebuild the original file with information from all three traces and can then compute the quality metrics mentioned above.

The video trace files generated consist of a description of each frame of the video with characteristics such as frame type, frame size and number of packets needed for each frame considering a 1024 bytes Maximum Transmission Unit (MTU).

The performance of the neuralFEC mechanism was assessed across several simulation experiments in Network Simulator-3 (NS-3) (Riley and Henderson, 2010). The evaluation scenario is comprised of 25 nodes in a grid disposition (5x5), located 50 meters apart. The Optimized Link State Routing Protocol (OLSR) (Clausen et al., 2003) was used as the routing protocol since all nodes have a route table with information to every node in the network. Therefore, routes are immediately available on request and as such, no increased delay due to protocol route discovery is added to the video transmission. Ten video sequences were used in this scenario, namely Bowing, Coastguard, Container, Crew, Foreman, Hall, Harbour, Mother and Daughter, News and Soccer. These sequences were selected in order to provide a wide variety of motion intensities.

To approximate the behaviour of a wireless network, a two-state discrete-time Markov chain model was implemented following a simplified Gilbert-Elliot packet-loss model (Razavi et al., 2009). This model produces error representations which are closely related to those of burst loss patterns of wireless channels (Wilhelmsson and Milstein, 1999) and real time services on the internet (Haßlinger and Hohlfeld, 2008). The simplified Gilbert-Elliot Model is shown in Figure 5.4, where the probability of packet loss in the Good state (G) was set at 0, which means no losses, and the probability of packet loss in the Bad state (B) was set at 1, where all the packets are lost. The Packet Loss Rate (PLR) can be obtained by Equation 5.6, where $P_{(BG)}$ represents the probability of transitioning from the Bad state to the Good state and vice-versa with $P_{(GB)}$.

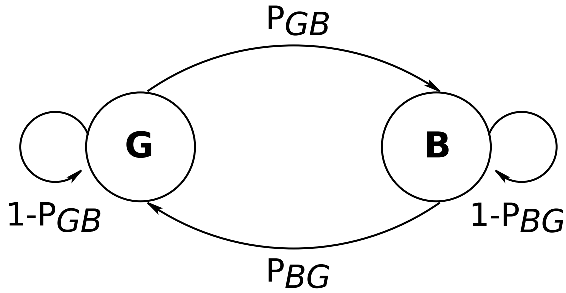

Figure 5.4: Gilbert-Elliot simplified model

$$PLR = \frac{P_{BG}}{P_{BG} + P_{GB}} \tag{5.6}$$

5.3.3 Results

In this section, the results of the simulation experiments with the neuralFEC mechanism are analysed and discussed.

Three distinct experiments with different schemes were performed. The first experiment serves the purpose of providing a baseline to the neuralFEC mechanism as there is no FEC mechanism in use. The second experiment was performed with a non-adaptive video-aware FEC mechanism (Video-aware FEC), where a fixed amount of redundancy (38%) was added to only I- and P- frames. This amount of redundancy was selected according to an extensive set of experiments representing the best trade-off between between video quality and the characteristics of the experimental scenario. The last experiment is focused on the neuralFEC mechanism.

Figure 5.5 illustrates the added overhead of the experiments during video transmission. Regarding the use of the non-adaptive Video-aware FEC mechanism, the added network overhead situated itself between 34% and 43%. If the neuralFEC mechanism is used, the added overhead is reduced to values between 13% and 24%. This means that the average redundancy added by the non-adaptive FEC mechanism has a value of 38%. On the other hand the adaptive neuralFEC mechanism achieved an average amount of redundancy of 19%. From the overhead results it is also observable that a greater amount of redundancy was attributed to video sequences such as Crew, Soccer, Harbour and Coastguard. On the contrary, video sequences with less motion intensity such as Bowing, Mother and Hall are given less redundancy. This is due to the RNN selecting higher amounts of redundancy to frames with a high degree of movement and lower amounts of redundancy to frames with a low degree of movement. This shows that the neuralFEC mechanism can correctly assess the importance of frames of frames according to their motion intensity.

Figure 5.6 depicts the SSIM scores for each video sequence while using the three aforementioned mechanisms. These show that the average SSIM value obtained for the neuralFEC mechanism is of 0,831 against a value of 0.819 for the video-aware FEC mechanism. Both FEC mechanisms obviously surpass the mechanism with no protection and thus results in an average SSIM value of 0,726. The average neuralFEC SSIM score represents a slight improvement of almost 1,5% over the video-aware mechanism. Also, from the analysis of the results for video sequences which are very distinct in terms of motion intensity such as Bowing and Harbour the following is observable: if we take into account the Without FEC mechanism, the video-aware FEC mechanism and the neuralFEC mechanism for the Bowing video sequence the SSIM scores obtained for the three mechanisms were of 0,886, 0,924 and 0,925, respectively. As for the Harbour video sequence the SSIM scores are of 0,474, 0,663 and 0,673, respectively. This is explained by the disparity in motion intensity characteristics between the two sequences. The Harbour video sequence is much heavier in terms of movement compared to the Bowing video sequence which means packet loss has a greater impact on this type of sequences. Ultimately, this results in lower SSIM scores in sequences with a

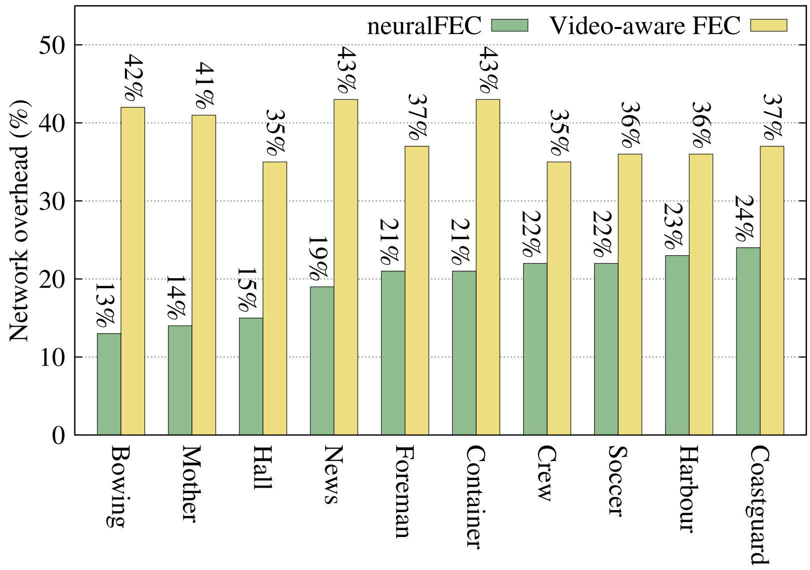

FIGURE 5.5: Network Overhead

higher degree of motion intensity. Also it highlights the greater resilience that video sequences with low motion intensity have to packet loss.

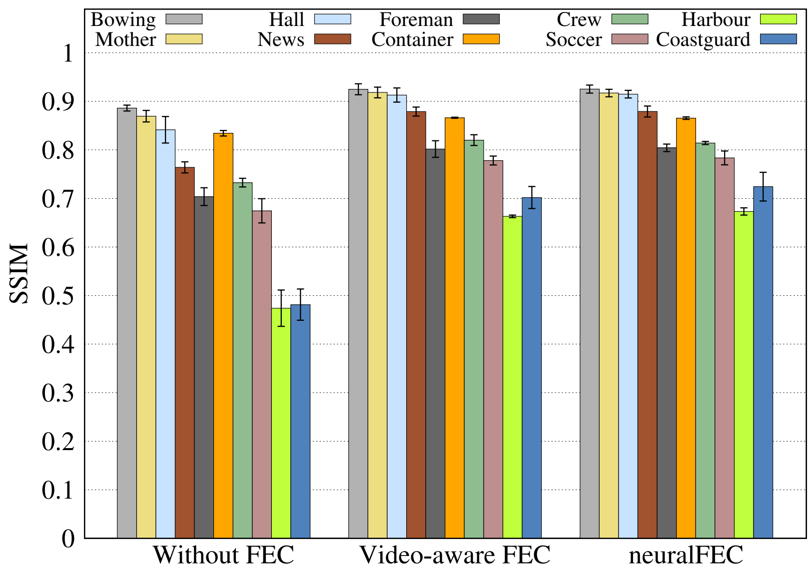

FIGURE 5.6: Network Overhead

Table 5.1 summarizes the results in terms of average overhead and SSIM of all of the video sequences for the three studied mechanisms. It shows that there was a slight improvement in the video quality by the neuralFEC mechanism, but most importantly the used amount of added redundancy was halved.

TABLE 5.1: Average SSIM and network overhead

	neuralFEC	**Video-aware FEC**	**Without FEC**
SSIM	0,831	0,819	0,726
Overhead	19,334%	38,460%	–

5.4 Summary

This chapter presented the base mechanism which functions solely based on a RNN. The neuralFEC mechanism, through an accurate motion intensity classification of video sequences with distinct characteristics, is able to add a precise amount of protection according to the degree of motion present in the video sequence. This shields video transmission, protecting the most QoE-sensitive data, maximizing the video quality while saving network resources by not sending unnecessary redundancy. Reducing overhead in wireless networks is of great importance, due to the limited nature of the wireless channel resources, which is hindered by packet loss from concurrent transmissions and network congestion. This contribution resulted in the presentation and publication of a paper titled *Adaptive Motion-aware FEC-based Mechanism to Ensure Video Transmission* included in Appendix A which was presented at The Nineteenth IEEE Symposium on Computers and Communications (ISCC 2014) (Immich et al., 2014a).

Chapter 6

Advances Towards Video Transmission Optimization: Enhanced Mechanisms

This chapter describes two enhancements performed on the base neuralFEC mechanism to further improve the adaptation of the allocated redundancy to the video and network characteristics. The following sections are organized in this fashion: Firstly, the objectives of the enhancements are described; Secondly, the tools used for the implementation and evaluation of the enhancements are presented; Finally, the improvements are specified and the obtained results are analysed.

6.1 Objectives

The main goal of the enhanced mechanisms is to improve on the neuralFEC base mechanism by further reducing the weight of redundancy needed to protect a given frame during transmission, while at the same time increasing the frame's resilience to errors due to packet loss. This can be done with an Ant Colony Optimization (ACO) scheme which takes into account several video characteristics and the network loss error rate. With this added information the enhanced mechanisms can perform a more accurate characterization of the video frame being transmitted which results in a more precise allocation of redundancy.

6.2 AntMind

The AntMind enhanced mechanism uses the Ant Colony Optimization algorithm in order to provide an even more accurate redundancy allocation than neuralFEC. This is performed by jointly using video characteristics and network loss information to assess the amount of redundancy needed for a given frame.

A directed graph was built so that the virtual ants were able to search for a solution. The graph comprises nodes representing the types of frames, movement complexity and error rate.

The graph's path weights were selected through an exploratory analysis of several different weight distributions. Each individual layer representing a video characteristic was tested with different weight relations to the nodes of the same characteristic and the nodes of other characteristics. This helped to determine which video characteristics are more important (i.e., motion intensity is more important than frame type, and higher motion intensity is more important than lower motion intensity). The distribution which produced results closer to those expected for all the motion intensity cases and video/network characteristics was used, producing the values presented in the results section.

6.3 AntMind Mechanism Specification

The main component present in this enhancement of the neuralFEC mechanism is the redundancy attribution mechanism. It is based on ACO. Other adaptive FEC mechanisms such as APB-FEC (Tsai et al., 2011a) and OCLFEC (Talari et al., 2013), and optimization methods such as Game Theory (Myerson, 2013) and several other Heuristics (Gilli and Winker, 2009) were considered. They were dismissed due to the fact that ACO allows for the dynamic change in the conditions of the problem and works around this to adaptively build a solution every time the said conditions change.

Figure 6.1 shows a general overview of the AntMind adaptive FEC mechanism. It is composed of three main modules, where the two first are based on the neuralFEC mechanism. The RNN module is trained and validated with several video sequences and is then used to classify each video sequence according to its motion intensity. The relationship between the neuralFEC and ACO is described by the path which represents the passing of the motion intensity classification information to the ACO module.

The novel ACO module present in this enhancement takes the previous module's (neuralFEC) numeric classification of the video frame and starts with a predefined path graph as shown in Figure 6.2. The numeric motion intensity classification value and size of the frames are translated into three intervals each. These intervals can accurately represent the situation at hand and facilitated the construction of the graph. The search graph itself consists of five layers with fourteen nodes depicting video and network characteristics. The distribution of the nodes in five layers has to do with the several characteristics of the video/network being analysed. Therefore, the first layer only has one node which serves merely as a starting point. Then the next layer has three nodes which represent the classification given by the RNN in terms of motion intensity, namely low, medium and high obtained by the hierarchical cluster analysis presented in Figure 5.2. The third layer has two nodes which represent the frame type, I- or P-. The fourth layer has

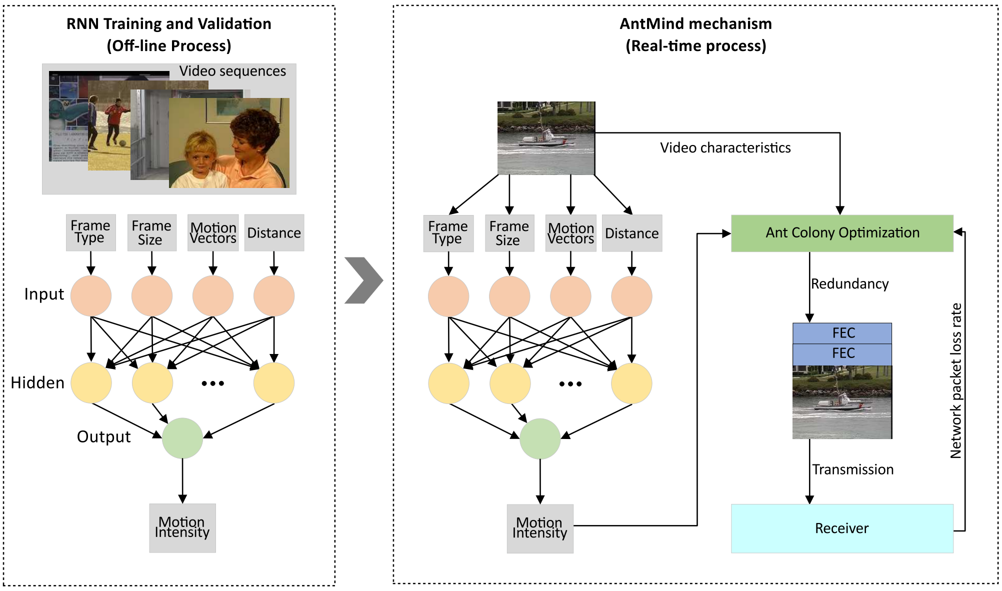

FIGURE 6.1: AntMind mechanism

three nodes depicting the frame size as small, medium and large. The last layer has five nodes representing the instantaneous network loss rate. The placement of the layers relative to each other is irrelevant, except for the first node, since the solution is not affected by the position of each layer. By attributing different orders of significance to each variation in a characteristic, we specify what will have more weight and consequent impact in the final redundancy value.

The ACO algorithm is executed, populating the graph with 10 virtual ants,

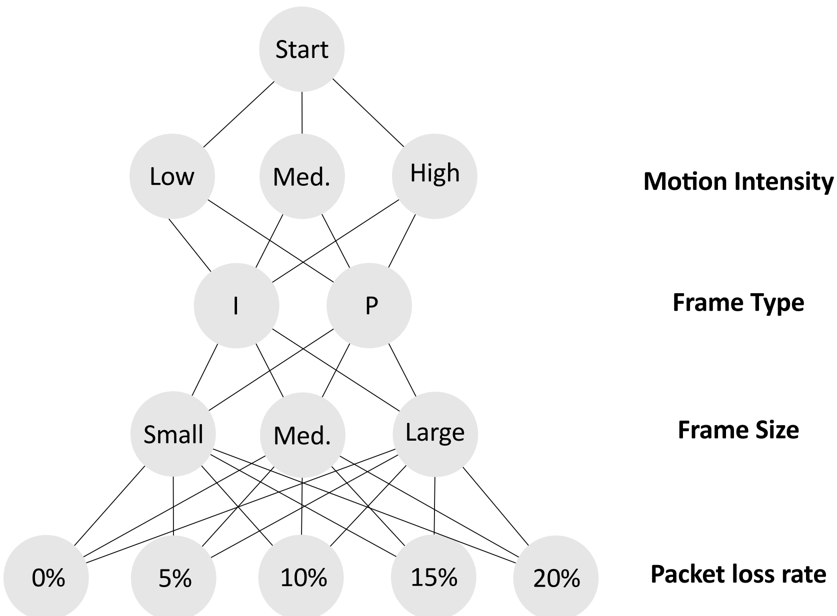

FIGURE 6.2: ACO graph used in AntMind

which according to the algorithm's path-finding routines will find the most adequate path according to the graph's path weights. This process is repeated 10 times for each of the frames being transmitted. The number of ants and repetitions was calibrated so that the execution of the ACO algorithm would not have a significant effect on the delay of the video transmission. The resulting path cost determines the redundancy value to be used by the Reed-Solomon Error Correcting Code. Should the network's error characteristics or the video's characteristics change, the ACO algorithm selects other corresponding paths, thus resulting in a different value of redundancy which is more adequate to the current set of parameters.

6.4 AntMind Mechanism Evaluation

This section presents the evaluation of the AntMind mechanism.

6.4.1 Evaluation Objectives

The performed evaluation aims at showing that the AntMind mechanism can effectively reduce network overhead while preserving video quality in comparison to other adaptive mechanisms. Similarly to neuralFEC, the redundancy overhead and objective QoE metrics are used to examine the performance of the mechanism. An additional objective QoE metric, Video Quality Metric (VQM) (Xiao et al., 2000) is used in order to more thoroughly assess the quality of the transmitted video. Contrary to SSIM, where higher values are an indicator of higher quality, in VQM, values closer to 0 are the ones which represent superior image quality.

6.4.2 Evaluation Scenario

The same network scenario developed for the neuralFEC mechanism's evaluation is used. Additionally, the Gilbert-Elliot packet-loss model was configured for four distinct packet loss rate values such as 5%, 10%, 15% and 20% to provide a greater insight into the effects of loss in the transmitted video.

6.4.3 Results

Four distinct data protection schemes were used, namely one with no protection, one with non-adaptive protection and two with adaptive protection. As with neuralFEC, the first scheme serves as a baseline for assessment of improvements in the video quality and measurement of the generated overhead. The second scheme is the Video-aware Equal Error Protection (VaEEP) scheme and the third scheme is the Video-aware UEP (VaUEP). VaEEP corresponds to the video-aware

FEC scheme used in the neuralFEC mechanism. In VaEEP a fixed amount of redundancy is added to I- and P- frames. In VaUEP the same frames are protected with fixed but distinct amounts of redundancy for each type. These amounts were selected from a thorough set of simulation experiments which resulted in values that represent the best possible trade-off between redundancy overhead and video quality in terms of QoE. Thus, the resulting values are of 38% and 30% for VaEEP and VaUEP, respectively. The last studied mechanism is AntMind which combines the use of an RNN and ACO for Unequal Error Protection.

Table 6.1 shows the results in terms of the average added overhead applying all of the PLRs using the three FEC schemes. The use of VaEEP resulted in an amount of overhead ranging from 35% to 43% averaging at 38%. Using VaUEP, the resulting overhead ranged from 25% to 36% averaging at 30%. The AntMind mechanism achieved overhead values which ranged from 9% to 19% averaging at 15%. This means that in some cases the reduction of added overhead was in the order of 77% over VaEEP and 72% over VaUEP. Therefore, far less redundancy data is used by the AntMind mechanism, opposed to VaEEP and VaUEP translating into reductions of 61% and 50%, respectively.

TABLE 6.1: Average Overhead

	VaEEP	**VaUEP**	**AntMind**	**AntMind STDEV**
Bowing	42%	36%	10%	0,239
Coastguard	37%	27%	19%	0,504
Mother	41%	33%	9%	0,323
Soccer	36%	27%	16%	0,293
Hall	35%	27%	13%	0,304
Container	43%	35%	18%	0,493
Crew	35%	25%	16%	0,376
Foreman	37%	28%	15%	0,361
News	43%	35%	17%	0,400
Harbour	36%	26%	18%	0,378
Overall	38%	30%	15%	0,367

From these results it is already clear that the adaptive AntMind mechanism assesses the importance of frames according to their MI characteristics. Therefore, the mechanism attributes a lesser amount of redundancy to videos with a low degree of motion intensity and a greater amount of redundancy to frames with a higher degree of motion intensity. This is clearly observable for Bowing and Mother which are categorized as low motion intensity video sequences, opposed to Coastguard, Harbour and Container which are categorized as high intensity video sequences. Additionally it is possible to notice that in the Mother and Soccer video sequences the VaEEP mechanism allocates a higher amount of redundancy to Mother (41%) than to Soccer (36%). In this video sequence the proportional size of the I- and P- frames is higher than that of the B- frames due to its low motion intensity. In the Soccer video sequence this is not valid because it has a higher motion intensity. Therefore, the AntMind mechanism allocated a higher amount of redundancy to Soccer (16%) in contrast to Mother (9%).

Two sets of assessments were preformed using the SSIM and VQM objective QoE metrics. The different values which can be observed are due to the particular characteristics of each video sequence. This highlights the importance of performing the experiments with videos with different types of motion spanning a broad array of motion intensity levels/characteristics.

Figure 6.3 shows the obtained SSIM results for each of the transmission schemes and for all of the video sequences. The average SSIM value was computed from all PLRs for each individual video sequence. The mechanism with no FEC averaged a value of 0,805. The VaEEP and VaUEP mechanisms obtained values of 0,881 and 0,882, respectively, which are closely matched by a value of 0,876 for the AntMind mechanism.

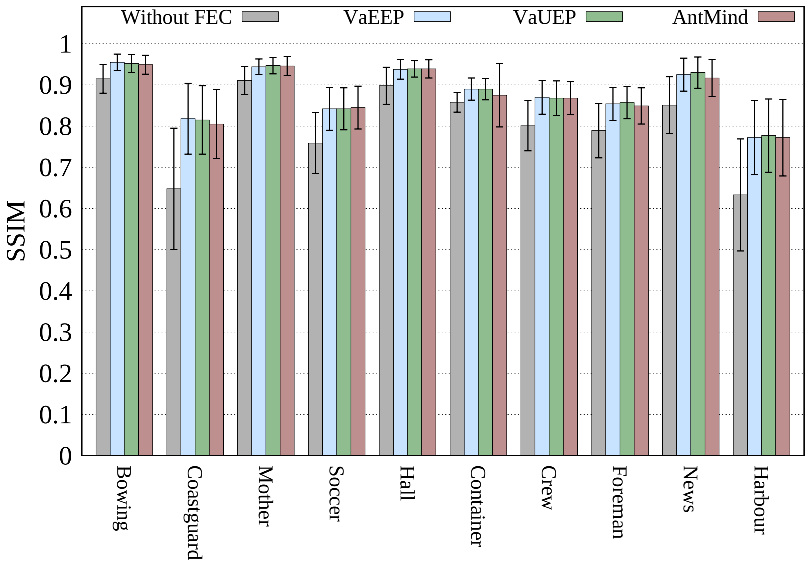

Figure 6.3: Objective QoE assessment (SSIM)

In Table 6.2 the Coastguard and Mother sequences are analysed in further detail. For Coastguard, a higher PLR represents SSIM scores which are considerably worse than for a lower PLR. If we take the AntMind results for 20% PLR and 5% PLR, the SSIM scores are 0,688 and 0,906, respectively. This represents a reduction of quality of about 25%. The Mother video sequence does not exhibit this type of behaviour. If we take the same PLRs of 20% and 5% the SSIM scores are 0,971 and 0,915,respectively. In turn, this represents a reduction of quality of about 5%. Such reduction of quality of the video sequence is expected as the PLR increases. Nevertheless, a variation of such magnitude as observed for Coastguard and Harbour means that packet loss has a greater impact on video sequences with a higher intensity of motion, contrary to Mother and Bowing. This highlights the resilience towards packet loss that video sequences with lower motion intensity have.

Table 6.2: SSIM variation through the different PLRs

	Without FEC		**AntMind**	
	Coastguard	**Mother**	**Coastguard**	**Mother**
PLR 20%	0,482	0,868	0,688	0,915
PLR 15%	0,574	0,899	0,784	0,938
PLR 10%	0,678	0,923	0,841	0,959
PLR 5%	0,859	0,955	0,906	0,971

Figure 6.4 shows the VQM results of the four schemes. The VaEEP and VaUEP mechanisms achieved an average VQM of 3,895 and 3,860, respectively, while the AntMind mechanism obtained an average of 3,940. This means that AntMind was able to maintain video quality. These results follow a similar trend to that of SSIM which helps corroborate the improvements of the proposed mechanism in overhead reduction with video quality shielding.

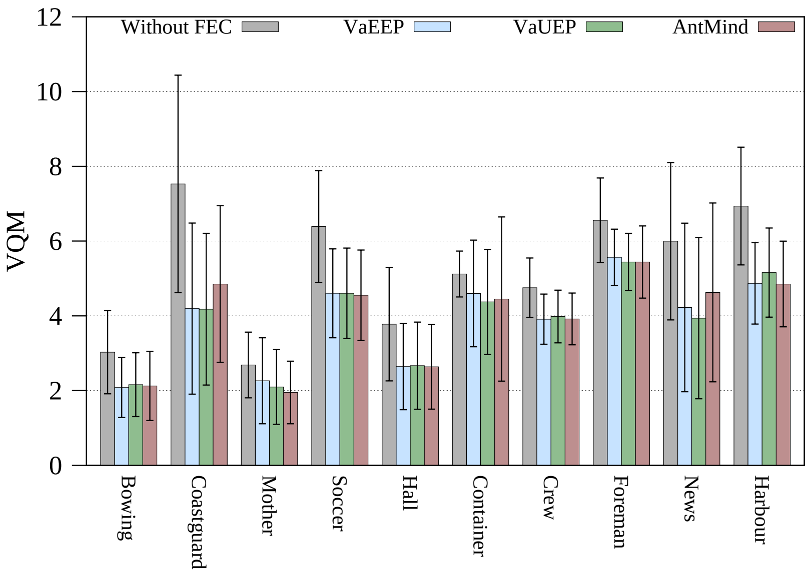

Figure 6.4: Objective QoE assessment (SSIM)

Table 6.3 summarizes the SSIM, VQM, and network overhead results. It shows that the AntMind mechanism managed to achieve a considerable reduction in network overhead by avoiding the addition of unnecessary redundancy. Furthermore, it did so while preserving video quality. This is of great importance in wireless network environments due to their limited network resources.

Overall the results showed that the AntMind mechanism can considerably reduce network overhead. AntMind performs a dynamic protection of the data which impacts QoE the most. Video quality is preserved while greatly reducing the use of network resources. The simulation results evidenced interesting behaviours from some video sequences, showing that those with a higher amount of motion

Table 6.3: Average SSIM, VQM and network overhead

	AntMind	VaEEP	VaUEP	Without FEC
SSIM	0,876	0,881	0,882	0,806
VQM	3,940	3,895	3,860	5,278
Overhead	14,898%	38,460%	29.827%	–

intensity require a distinct amount of protection, contrary to video sequences with lower amounts of motion intensity. AntMind takes into account motion intensity and network loss to adequately shield a video stream with varying video characteristics in a lossy wireless network scenario. AntMind was able to reduce network overhead by 61% on average, while keeping video quality similar to a best-case scenario, thus showing an improvement over both non-adaptive and adaptive FEC mechanisms.

6.5 Loss Prediciton

The Loss Prediction mechanism further improves on AntMind by performing error prediction from feedback data, instead of just using the instantaneous network loss rate to attribute redundancy to the video sequence being transmitted. The use of a loss prediction scheme enables a further reduction of added overhead by balancing the allocation of redundancy data between the network's good and bad states.

6.6 Loss Prediction Enhanced Mechanism Specification

The main improvement the Loss Prediction mechanism offers over AntMind is the prediction the occurrence of errors according to feedback information from the receiver. This feedback information comprises the distribution of good and bad gaps during transmission. As shown in Figure 6.5, a good gap is considered as the interval of packets that were successfully received between two bad gaps (white squares). A bad gap is the interval of packets during which a burst of errors is occurring (Dark red squares). The feedback information is collected by the receiver and sent to the transmitter in the form of a vector containing the size of every gap of each type. From this, the average gap size is computed on the server side to be used as a predicting value of a higher probability for the occurrence of an error burst.

Since the distribution of the size of both types of gaps tends to be located around a specific value for each packet loss rate, the average size of the good and

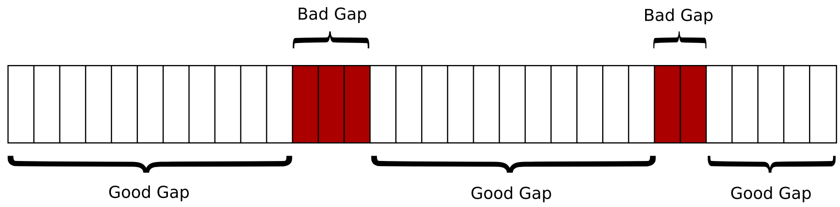

FIGURE 6.5: Packet gaps during transmission

bad gaps can be used to estimate if an error will occur during transmission of block of packets.

In order to be able to use error prediction, instead of performing redundancy allocation on a frame-by-frame basis, the mechanism was modified to support the transmission of each frame on blocks of 10 packets of original data. The size of the blocks was selected through extensive experimentation to provide more flexibility of the mechanism, while dealing with the error correcting code, and better control over the data. Furthermore, this provides enough granularity to comply with the sizes of the gaps. That is, it was observed that the sizes of gaps ranged from 9 to 22 packets for the good gaps and from 1 to 3 packets for the bad gaps. Therefore, the selected packet block size was that which allowed for the block to be isolated (i.e., whole block inside a good gap) from an error gap through error prediction in all the packet loss rate conditions.

Figure 6.6 represents the redesigned ACO graph for Loss Prediction. It is the same as for the AntMind mechanism but the last layer of nodes represents the possibility of occurrence of an error instead of the instantaneous packet loss rate. The five nodes represent five different scenarios which can occur: NE (No Error) is a scenario where no error is accounted for the current block; SSE (Shared Single Error) is a scenario where a single error which will be shared by this block and the next is predicted; SE (Single Error) is a scenario where a single error is predicted only for the current block; SLE (Shared Large Error) is a scenario where the occurrence of two error gaps is predicted in the current block continuing to the next one; LE (Large Error) is a scenario where two blocks of errors occur only on the current block.

6.7 Loss Prediction Mechanism Evaluation

This section presents the evaluation of the Loss Prediction mechanism.

6.7.1 Evaluation Objectives

The performed evaluation aims at showing that the Loss Prediction mechanism can effectively reduce network overhead while preserving video quality in comparison to other adaptive mechanisms. Similarly to AntMind and neuralFEC, the

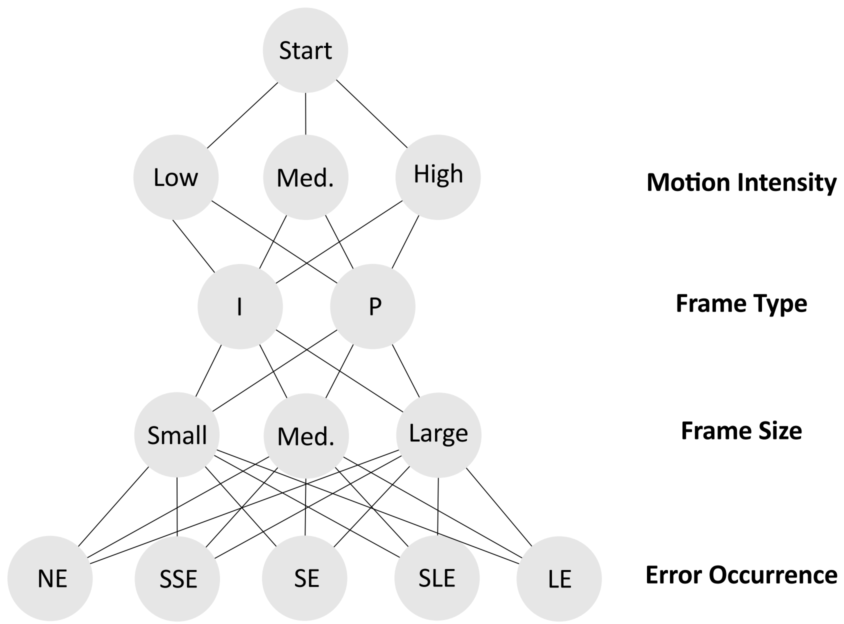

FIGURE 6.6: ACO graph used in Loss Prediction

redundancy overhead and objective QoE metrics are used to examine the performance of the mechanism. The objective QoE metrics used are again SSIM and VQM to provide a means of comparison with AntMind.

6.7.2 Evaluation Scenario

The same network scenario developed for the AntMind and neuralFEC mechanisms' evaluation is used as well as the Gilbert-Elliot error loss model with the same parameters.

6.7.3 Results

Following the same methodology used for the AntMind mechanism, four distinct data protection schemes were used, namely one with no protection, one with non-adaptive protection and two with adaptive protection. The results are presented in terms of the average of the values for all the four studied PLRs.

Figure 6.7 represents the average added redundancy overhead for all the studied video sequences. It is observable that the Loss Prediction mechanism managed to achieve an even better result than the AntMind mechanism. On average the gains obtained with Loss Prediction are in the order of 10% less added redundancy, ranging from 2,728% to 16,860%. The biggest reductions were achieved on the video sequences which have greater amounts of motion intensity, such as Harbour (16,860%), Hall (12,933%) and Foreman(11,745%). The lowest reductions were found on the videos that are opposite to these three in terms of motion intensity, specifically Mother (2,728%), Bowing (6,001%) and Container (7,152%). Since the Loss Prediction mechanism is an enhancement of AntMind, the greater gain in the overhead reduction has already been achieved. Therefore, this explains why the video sequences with lower intensities of motion had small amounts of reduction compared to AntMind. Regarding the standard deviation, it is observable that the overhead has a greater variation in the Loss Prediction mechanism contrary to AntMind. The reduced variation of the overhead with different packet loss rates for AntMind is due to the division of the frame in blocks which absorbs the fluctuation of the added redundancy. Also, the packet loss rate parameter was given a lower importance than the video characteristics in the ACO mechanism present in AntMind.

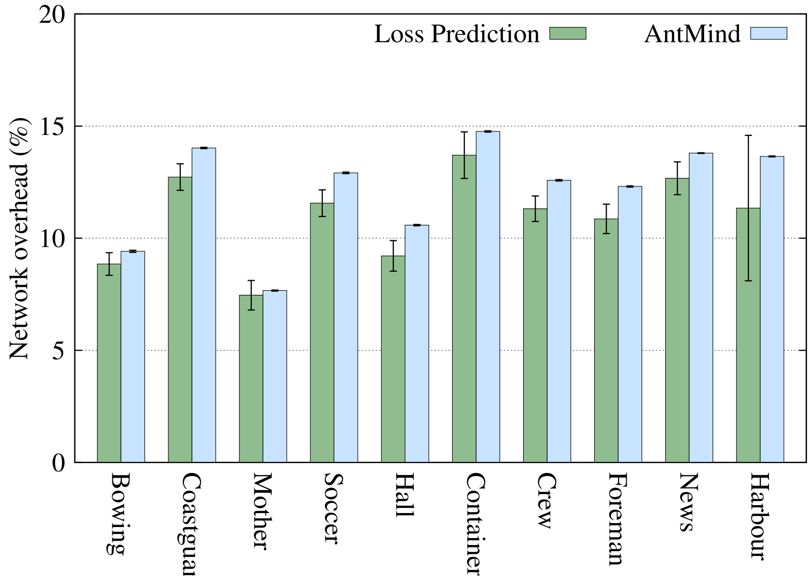

Figure 6.7: Average Overhead

Figure 6.8 shows the results in terms of SSIM of the Loss Prediction mechanism over the AntMind mechanism. The Loss Prediction mechanism achieved an average SSIM score of 0,884 versus 0,886 of the AntMind mechanism. This means that in terms of SSIM, the Loss Prediction mechanism behaved as intended presenting results similar to those of AntMind.

Figure 6.9 presents the VQM results of the Loss Prediction mechanism against the AntMind mechanism. With Loss Prediction the average VQM situated around 3,664 against 3,611 for the AntMind mechanism. This shows that VQM follows a similar trend to that of SSIM.

These results show therefore, that the Loss Prediction mechanism offers in terms of video quality an identical behaviour to that of AntMind. Taking this into account and the reduction of the added overhead we can say that it provides an even more precisely tailored protection scheme than the AntMind mechanism. This is due to enhancement performed over the AntMind-based mechanism which already combines motion intensity classification with ACO. In conjunction with

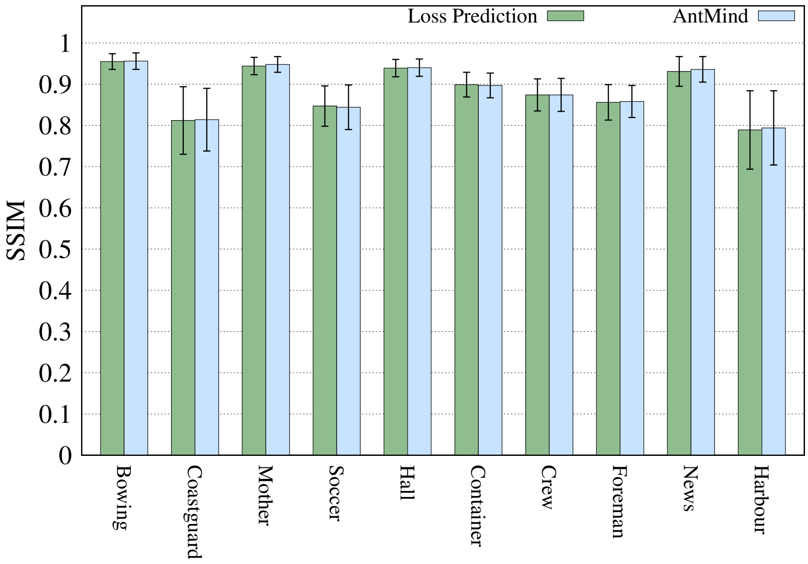

Figure 6.8: Objective QoE assessment (SSIM)

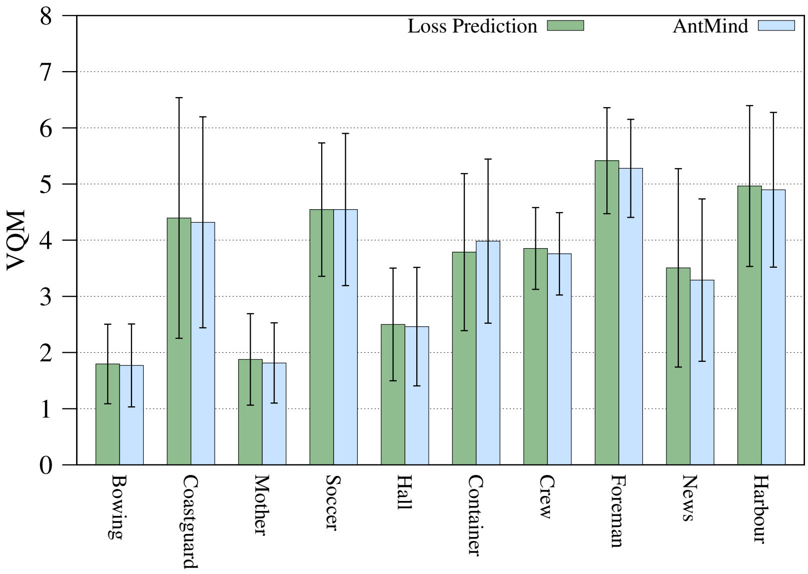

Figure 6.9: Objective QoE assessment (VQM)

the prediction of when good or bad gaps will occur, the Loss Prediction mechanism enables further reduction of the added overhead due to redundancy. At the same time, it shields the video against loss due to network errors with results similar to those obtained for AntMind meaning that the video quality is maintained.

6.8 Computational Complexity of the Enhanced Mechanisms

The AntMind and Loss Prediction mechanisms resort to implementations of two somewhat complex components such as the Reed-Solomon code Ant Colony Optimization system. The motion intensity classification is performed previously by the RNN off-line and has no runtime impact on computational complexity. Also, the feedback mechanisms, namely the instantaneous network loss feedback mechanism and the loss prediction mechanism, only perform a few simple operations and therefore their impact on computational complexity is negligible. Nevertheless, it is important to understand that these processes have a computational impact while being executed during runtime the developed mechanisms.

Regarding ACO, the complexity of finding a solution with an expected number of iterations is represented by Equation 6.1 (Attiratanasunthron and Fakcharoenphol, 2008) in a graph with n nodes, m edges and where ρ is the evaporation rate of the pheromone used by the ants.

$$O(\frac{1}{\rho} n^2 \, m \, log \, n) \tag{6.1}$$

Concerning Reed-Solomon, the encoding computational complexity is comprised of two steps, namely the pre-computing of the Generator Matrix (GM) of the code, followed by the multiplication of the source vector by the GM. Equation 6.2 represents the total computational complexity of the encoding per element (Lacan, 2009), where k represents the rows and n represents the columns of the GM matrix.

$$O(\frac{k}{(n-k)} * (log \, k)^2 + log \, k) \tag{6.2}$$

The decoding steps of the Reed-Solomon code involve he computation of the $k * k$ sub-matrix of the GM. Afterwards, this matrix is inverted and multiplied by the received vector in order to recover the original vector. The computational complexity (Lacan, 2009) per element of these steps is represented by Equation 6.3 where k is the number of received elements.

$$O((log \, k)^2) \tag{6.3}$$

6.9 Summary

This chapter presented the enhancements developed upon the base neuralFEC mechanism, namely the AntMind mechanism and the Loss Prediction mechanism.

The AntMind Mechanism utilizes Ant Colony Optimization metaheuristic which consists of representing the behaviour of a colony of ants in real life to a problem solving evolutionary algorithm. The ACO module receives the motion intensity classification values of the video sequences provided by the Random Neural Network already present in the neuralFEC mechanism. Along with this, it takes into account the characteristics of the video and the instantaneous network error loss rate to perform an adaptive allocation of redundancy and therefore shield the video against errors. The Loss Prediction mechanism improves on AntMind by replacing the use of the instantaneous network error loss rate with a module which retrieves transmission statistics. It performs the prediction of when good or bad periods of transmission will occur and allocates more redundancy to the periods of when loss is expected and less redundancy to the periods of when loss is not expected.

From the examination of the results it is observable that the Loss Prediction mechanism managed to provide an improvement against the AntMind mechanism. This is due to the fact that AntMind already provided the greatest amount of enhancement over the VaUEP and VaEEP mechanisms. Nevertheless, the results provided by the Loss Prediction must not be neglected since they provide an even further enhancement to an already good mechanism such as AntMind. Such improvement is advantageous since it managed to provide good video quality with reduced amounts of overhead leading to a better utilization of network resources and further improved QoE in general.

The contribution presented with the AntMind mechanism resulted in the presentation and publication of a paper titled *AntMind: Enhancing Error Protection for Video Streaming in Wireless Networks* included in Appendix A which was presented by the author at The Fifth International Conference on Smart Communications in Network Technologies (SaCoNet 2014) (Immich et al., 2014b).

Chapter 7

Project Management

This chapter presents the work plans for the first and second semesters of the Master Thesis proposal presented in this document.

The work plan for the first semester is presented in the first section and the next section describes the second semester's work plan.

7.1 First semester work plan

In the first semester, due to the nature of exploring several libraries and dealing with some bugs and particularities of the code, some delays occurred which are shown in Figure 7.2 .

- **Study of state-of-the-art in video transmission techniques**

  This task involved the study of techniques like Forward Error Correction and Retransmission schemes.

- **Analysis of error correction techniques and QoE assessment methods**

  The analysis of error correction techniques involved the study of several types of error correcting codes, namely linear block codes, convolutional codes and fountain codes. The second step involved the study of QoE assessment techniques, which can be subjective, objective and hybrid.

- **Identification of issues in existing mechanisms related to video optimization and the perceived video quality (QoE)**

  In this step, several research contributions were studied and the respective advantages and disadvantages were identified, thus providing a starting ground on what the main problems are when trying to design a mechanism for resilient video transmission.

- **Specification for mechanisms for resilient video transmission in wireless networks**

  With the issues for resilient video transmission identified, this task involved the study of mechanisms for video classification and also adaptive redundancy allocation that could then be used in the design of the proposal.

- **Design, implementation and evaluation of a proof-of-concept mechanism in a simulation environment**

  After the suitable methods for the achievement of the objects were identified, the design of the mechanism was defined as having a video classification mechanism which uses Random Neural Networks and an adaptive redundancy allocation mechanism based on Ant Colony Optimization. This task is delayed due to the non-existence of dedicated C libraries for RNNs and also to the inexperience of working with such a mechanism. Also there have been problems with the error correcting codes, namely Reed-Solomon that was producing unexpected results but are now soved. Therefore this step of the project only involves the implementation of the classification mechanism and preliminary results obtained by testing the RNN for classification.

- **Writing a paper with the state-of-the-art and preliminary results of the proposed mechanism**

  This task has not yet been started.

- **Writing of intermediate report**

  This task involved the writing of the intermediary report and has been underway since the beginning of the project.

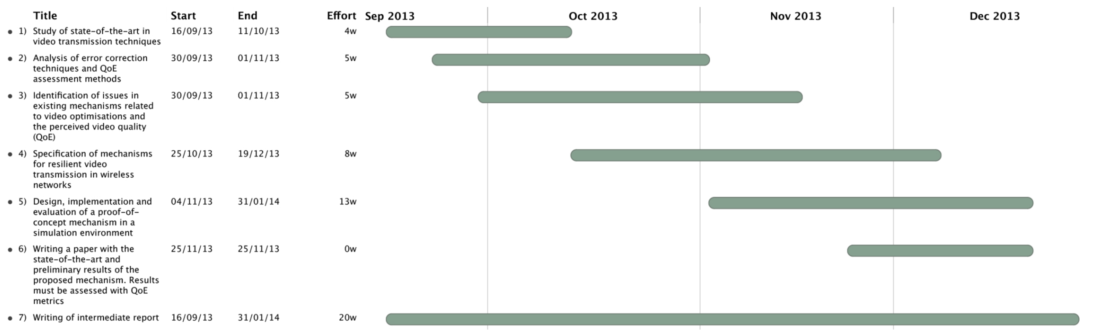

Figure 7.1: Projected first semester work plan

By the analysis of the diagrams of the first semester work plan, it is easily seen that some tasks have suffered delays.

The second task was delayed due to the study of state-of-the-art in video transmission techniques, and an exploratory analysis of three types of error correcting codes in the NS-3 simulation environment to assess the performance and particularities of each of them.

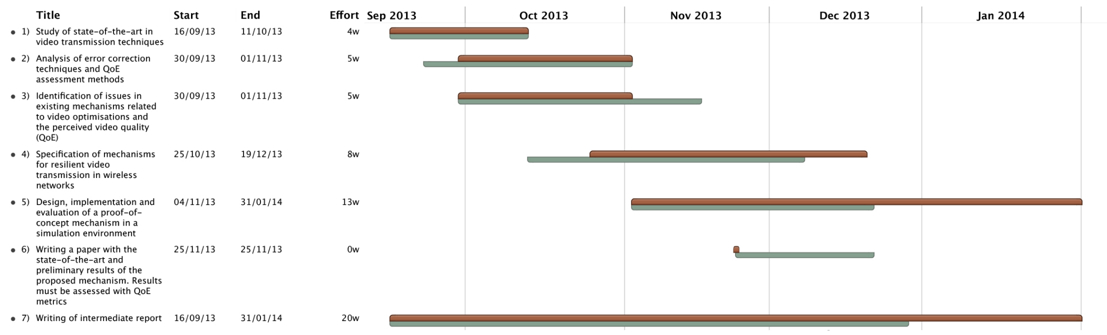

FIGURE 7.2: Final first semester work plan

The fourth task also suffered a minor delay from some indecisions regarding which optimization method to use, and how it could be used with the available parameters.

The fifth task is a task that has been continuously in progress and will continue on as the first task of the second semester.

The sixth task has not yet been started derived from the setback of not yet having fully functioning classification and adaptive FEC mechanisms.

The seventh task prolonged itself onto the date of the submission of the intermediate report for improvements from the suggestions given by the supervisors.

7.2 Second semester work plan

The second semester work plan is presented in this section. Due to delays on some tasks in the first semester, these will follow on through the second semester and will be integrated in the work plan.

- **Implementation of the proposed mechanism for resilient video transmission in wireless networks**

  This task involves the complete integration of both the classification and adaptive FEC mechanisms and consequent implementation in the NS-3 simulator in wireless network environments.

- **Implementation and evaluation of the state-of-the-art mechanisms**

  In this task the performance of the existing state-of-the-art mechanisms such as the ViewFEC mechanism (Immich et al., 2013) is assessed to provide for a baseline to compare and evaluate the performance of the proposed mechanism.

- **Proposal validation, assessment and comparison with state-of-the-art mechanisms**

  This task involves the validation of the proposal with extensive testing with several network conditions and different video files. Firstly, each scenario will be evaluated with all of the selected videos. Secondly, the network loss conditions will be increased gradually and the performance of the mechanism will assessed. Finally, the results will be compared against the results of the state-of-the-art mechanisms assessed in the previous task.

- **Writing of two papers with the proposed mechanism compared to state-of-the-art mechanisms**

  Throughout the semester, two articles will be written exploring the performance of the proposed mechanism against other ones.

- **Writing of the final report**

  This task involves the writing of the final report comprising the completed proposed mechanism and the obtained results throughout the semester.

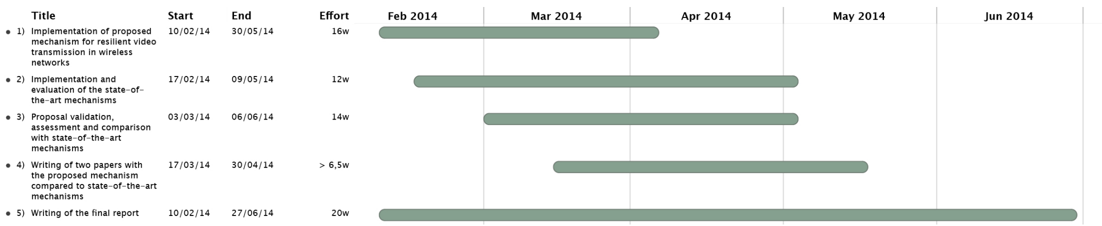

FIGURE 7.3: Projected second semester work plan

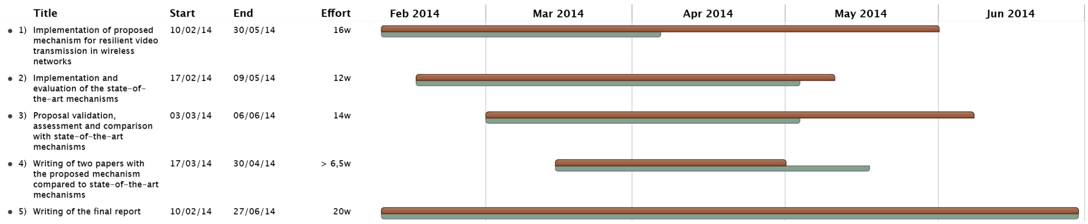

Figure 7.4: Final second semester work plan

Taking into account the diagram corresponding to the final second semester work plan, it is observable that some of the tasks work effort differed from what was projected. Nevertheless, there were no major delays, meaning that the variations are due to adjustments to the work plan.

The first task lasted until the end of May because after performing the implementation of the proposed mechanism, it was decided to enhance it further in the form of loss prediction instead of using the network error loss rate from the feedback information received.

The second task which is the implementation of the state-of-the-art mechanisms was extended. This was due to the necessity of modifying the mechanisms to be used as comparison to asses the performance of the Loss Prediction mechanism.

Regarding the third task, since it was decided to enhance the AntMind mechanism, the performance assessment and validation of the mechanism required more time, thus explaining the larger work effort.

The fourth task represents the writing of the two published papers. This task was performed in a shorter time than expected in order to meet deadline constraints. The first submission deadline was at the end of March and the second submission deadline was at the end of April. Considering this, the work effort for the papers was of about one month to each.

The fifth comprised the writing of the final report and suffered no adjustments.

Chapter 8

Final Considerations

This document has addressed the subject of techniques for resilient video transmission in wireless networks. There are already several mechanisms in the literature which make video transmission less subject to degradation due to losses and therefore improve its Quality of Experience. However, these techniques alone are not enough to achieve an optimal protection of the video data during transmission which calls for a new approach. In order to attain an acceptable video Quality of Experience, the mechanisms used to ensure a resilient video transmission have to be prepared to deal with error losses(i.e., Forward Error Correction), specifically burst error losses which are very common in wireless networks. Nevertheless, in order to avoid the unnecessary transmission of excessive data, the use of an adaptive FEC mechanism instead of a static one is important.

From the analysis of Forward Error Correction Techniques it became clear that static FEC schemes do not offer optimal allocation of the redundancy, therefore compromising the protection of the transmitted video. The examination of the types of codes revealed that the Reed-Solomon code is simple and flexible enough to allow for its use in an adaptive FEC solution. Furthermore, the use of an adaptive FEC scheme has an obvious advantage over a static FEC scheme: it can protect the data more efficiently by allocation of different amounts of redundancy to each individual frame, while also helping to reduce network overhead.

The study of the QoE metrics revealed that the subjective evaluation metrics are very precise in terms of assessing the quality of the video. However, they have a relevant downside, which is the requirement for a human intervenient. This makes them not suitable for use in an environment requiring real-time assessment of the video, such as the developed adaptive FEC mechanism. Therefore, in this study were only used objective QoE metrics such as SSIM and VQM to assess the quality of the video.

As the adaptive FEC approach needed a mechanism that classifies the video motion intensity, the analysis of several classification schemes was performed. The mechanism that was shown to be the most adequate to the requirements was the Random Neural Network because it provides a low complexity and adjustable environment for training, as well as strong generalisation characteristics. In addition, mechanisms for dynamic allocation of redundancy were explored, with Ant Colony Optimization revealing itself as being an interesting and suitable candidate for the task. It allows the alteration during execution of the weights of the paths of the graph enabling us to feed external parameters to it during transmission. This results in the modification of the solution which gives the FEC mechanism its dynamic nature in redundancy allocation.

The goals of achieving a better allocation of redundancy while improving/-maintaining QoE led to the development of three mechanisms. Each mechanism provides an enhancement to the previous one in succession (i.e. AntMind improves on neuralFEC and Loss Prediction improves on AntMind).

The first developed mechanism was the neuralFEC and comprises a Random Neural Network for video motion intensity classification. This RNN works together with a simple selection mechanism which attributes amounts of redundancy according to the motion intensity classification value. The video frame is then coded by the Reed-Solomon error correcting code and is sent. The results showed that by classifying the video frames according to motion intensity the mechanism can provide an adaptive and adequate amount of protection while at the same time saving network resources in the form of reduced overhead.

The AntMind mechanism improves on neuralFEC by using an Ant Colony Optimization algorithm to perform an even more accurate characterization of the video combined with information about the instantaneous network loss rate. The ACO algorithm takes into account the values of the RNN and combines it with video/network parameters, adaptively selecting an amount of redundancy which better suits the current conditions. The AntMind mechanism improved on neuralFEC providing further reduction of added overhead while maintaining the video quality assessed.

The Loss Prediction mechanism is a modification of AntMind where the instantaneous network loss rate scheme is replaced with a statistical analysis of intervals of transmission comprised of successful transmission and errors. The analysis of this information is translated into gaps of good and bad packets which are used to construct a model of the current transmission in terms of the size of the successful and unsuccessful transmissions. These gaps can be used to perform the prediction of when an error burst will occur. Therefore, combined with the calculated redundancy for protection according to the characteristics of the video, further redundancy is added to account for the occurrence of the error burst. The Loss Prediction mechanism managed to attain even further reduction in added overhead over AntMind which already managed to do so in a large amount. Furthermore, the mechanism achieved this while maintaining the video quality of the sequences according to the two studied metrics (SSIM and VQM). Taking this into account, the Loss Prediction mechanism proved to be the best of the three providing a suitable amount of protection according to video characteristics and the network loss patterns.

Taking into account the limitations of the performed study, future work should be focused on the use of other network scenarios involving mobility such as cellular mobile networks. Also, it would be of interest to explore a better implementation of Reed-Solomon or another more efficient error correcting code. Furthermore, in order to further assess the performance of the mechanism, it would be pertinent to experiment with videos which are larger and involve varying motion intensity.

Nevertheless, the work presented in this thesis shows that through the analysis of all the related mechanisms and the implementation of these in the adaptive FEC scheme, the goal of providing a robust and resilient mechanism for video transmission error protection in a wireless network environment was achieved.

Bibliography

Ming-Fong Tsai, Naveen Chilamkurti, and Ce-Kuen Shieh. An adaptive packet and block length forward error correction for video streaming over wireless networks. *Wireless Personal Communications*, 56(3):435–446, 2011a. ISSN 0929-6212.

Ali Talari, Sunil Kumar, Nazanin Rahnavard, Seethal Paluri, and JohnD Matyjas. Optimized cross-layer forward error correction coding for h.264 avc video transmission over wireless channels. *EURASIP Journal on Wireless Communications and Networking*, 2013(1):1–13, 2013.

Roger Immich, Eduardo Cerqueira, and Marilia Curado. Cross-layer fec-based mechanism for packet loss resilient video transmission. In Ernst Biersack, Christian Callegari, and Maja Matijasevic, editors, *Data Traffic Monitoring and Analysis*, volume 7754 of *Lecture Notes in Computer Science*, pages 320–336. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-36783-0.

S.-S. Kang L. Han, S. Park and H. P. In. An adaptive fec mechanism using cross-layer approach to enhance quality of video transmission over 802.11 wlans. *TIIS*, pages 341–357, 2010.

Weisi Lin and C.-C. Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297 – 312, 2011. ISSN 1047-3203. URL `http://www.sciencedirect.com/science/article/pii/S1047320311000204`.

Ajith Abraham. Artificial neural networks. *handbook of measuring system design*, 2005.

A. Khan, L. Sun, E. Jammeh, and E. Ifeachor. Quality of experience-driven adaptation scheme for video applications over wireless networks. *Communications, IET*, 4(11):1337–1347, 2010. ISSN 1751-8628. doi: 10.1049/iet-com.2009.0422.

comScore. More than 200 billion online videos viewed globally in october,. Tech. rep., comScore inc., www.comscore.com/Press Events/Press Releases/2011/12/More than 200 Billion Online Videos Viewed Globally in October, 2011.

comScore. 2013 europe digital future in focus. Whitepaper, comScore, March 2013.

Wolfgang Karner, Olivia Nemethova, and Markus Rupp. Link error prediction in wireless communication systems with quality based power control. In *Communications, 2007. ICC'07. IEEE International Conference on*, pages 5076–5081. IEEE, 2007.

Roger Immich, Pedro Borges, Eduardo Cerqueira, and Marilia Curado. Adaptive motion-aware FEC-based mechanism to ensure video transmission. In *19th IEEE Symposium on Computers and Communications (IEEE ISCC 2014)*, Madeira, Portugal, June 2014a.

Roger Immich, Pedro M Borges, Eduardo Cerqueira, and Marilia Curado. AntMind: enhancing error protection for video streaming in wireless networks. In *5th International Conference on Smart Communications in Network Technologies 2014 (SaCoNeT 2014)*, Vilanova i la Geltru, Spain, June 2014b.

Daniel J. Costello Shu Lin. *Error Control Coding*. Prentice Hall, 2nd edition, June 2004.

A. Basalamah and T. Sato. A comparison of packet-level and byte-level reliable fec multicast protocols for wlans. In *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, pages 4702–4707, 2007. doi: 10.1109/GLOCOM.2007.892.

A. Nafaa, T. Taleb, and L. Murphy. Forward error correction strategies for media streaming over wireless networks. *Communications Magazine, IEEE*, 46(1):72–79, 2008. ISSN 0163-6804. doi: 10.1109/MCOM.2008.4427233.

Vincent Lecuire. Unequal error protection under bitrate constraint for video streaming over internet. *Computer Communications*, 35(3):287 – 297, 2012. ISSN 0140-3664. URL `http://www.sciencedirect.com/science/article/pii/S0140366411003161`.

M. C. Valenti and J. Sun. The umts turbo code and an efficient decoder implementation suitable for software defined radios. *International Journal of Wireless Information Networks*, 8:203–216, 2001.

Dipankar Raychaudhuri and Narayan B Mandayam. Frontiers of wireless and mobile communications. *Proceedings of the IEEE*, 100(4):824–840, 2012.

K.S. Andrews, D. Divsalar, S. Dolinar, J. Hamkins, C.R. Jones, and F. Pollara. The development of turbo and ldpc codes for deep-space applications. *Proceedings of the IEEE*, 95(11):2142–2156, 2007. ISSN 0018-9219. doi: 10.1109/JPROC.2007.905132.

C. Berrou, R. Pyndiah, P. Adde, C. Douillard, and R. Le Bidan. An overview of turbo codes and their applications. In *Wireless Technology, 2005. The European Conference on*, pages 1–9, 2005. doi: 10.1109/ECWT.2005.1617639.

B Siva Kumar Reddy and B Lakshmi. Adaptive modulation and coding in cofdm for wimax using lms channel estimator. 2013.

Inc. Digital Fountain. *Why Digital Fountain's Raptor Code Is Better Than Reed Solomon Erasure Codes For Streaming Applications.* 2005.

Shakeel Ahmad, Raouf Hamzaoui, and Marwan M Al-Akaidi. Unequal error protection using fountain codes with applications to video communication. *Multimedia, IEEE Transactions on*, 13(1):92–101, 2011.

David J.C. MacKay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, 2003.

Irving S. Reed and Gustave Solomon. Polynomial codes over certain finite fields. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):300–304, June 1960.

G. C. Clark Jr. and J. B. Cain. *Error-Correction Coding for Digital Communications.* Springer, 1981.

S. B. Wicker and V. K. Bhargava. *Reed-Solomon codes and their applications.* IEEE Press, 1994.

J. Neckebroek, M. Moeneclaey, and E. Magli. Comparison of reed-solomon and raptor codes for the protection of video on-demand on the erasure channel. In *Information Theory and its Applications (ISITA), 2010 International Symposium on*, pages 856–860, 2010. doi: 10.1109/ISITA.2010.5649587.

Junbin Chen, Lin Wang, and Yong Li. Performance comparison between non-binary ldpc codes and reed-solomon codes over noise bursts channels. In *Communications, Circuits and Systems, 2005. Proceedings. 2005 International Conference on*, volume 1, pages 1–4 Vol. 1, 2005. doi: 10.1109/ICCCAS.2005.1493349.

3GPP. Report for fec selection for mbms. Technical report, 3GPP, Lisbon,Portugal, February 2005.

R. G. Gallager. Low density parity check codes. *Cambridge, MA: MIT Press*, 1963.

David J.C. MacKay, David J. C. Mackay, Radford M. Neal, and Radford M. Neal. Good codes based on very sparse matrices. In *Cryptography and Coding. 5th IMA Conference, number 1025 in Lecture Notes in Computer Science*, pages 100–111. Springer, 1995.

John W. Byers, Michael Luby, Michael Mitzenmacher, and Ashutosh Rege. A digital fountain approach to reliable distribution of bulk data. In *Proceedings of the ACM SIGCOMM '98 conference on Applications, technologies, architectures, and protocols for computer communication*, SIGCOMM '98, pages 56–67, New York, NY, USA, 1998. ACM. ISBN 1-58113-003-1. doi: 10.1145/285237.285258. URL `http://doi.acm.org/10.1145/285237.285258`.

M. Luby. Lt codes. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 271–280, 2002. doi: 10.1109/SFCS.2002.1181950.

A. Shokrollahi. Raptor codes. *Information Theory, IEEE Transactions on*, 52(6): 2551–2567, 2006. ISSN 0018-9448. doi: 10.1109/TIT.2006.874390.

D. J C MacKay and R.M. Neal. Near shannon limit performance of low density parity check codes. *Electronics Letters*, 32(18):1645–, 1996. ISSN 0013-5194. doi: 10.1049/el:19961141.

Mustafa Eroz, Feng-Wen Sun, and Lin-Nan Lee. Dvb-s2 low density parity check codes with near shannon limit performance. *International Journal of Satellite Communications and Networking*, 22(3):269–279, 2004.

V. Oksman and S. Galli. G.hn: The new itu-t home networking standard. *Communications Magazine, IEEE*, 47(10):138–145, 2009. ISSN 0163-6804. doi: 10.1109/MCOM.2009.5273821.

IEEE. 802.11n, October 2009.

C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error-correcting coding and decoding: Turbo-codes. 1. In *Communications, 1993. ICC '93 Geneva. Technical Program, Conference Record, IEEE International Conference on*, volume 2, pages 1064–1070 vol.2, 1993. doi: 10.1109/ICC.1993.397441.

ETSI. Digital video broadcasting (dvb); interaction channel for satellite distribution systems, May 2009.

Mohamed Hassan and Taha Landolsi. A retransmission-based scheme for video streaming over wireless channels. *Wireless Communications and Mobile Computing*, 10(4):511–521, 2010. ISSN 1530-8677.

N. Rahnavard, B.N. Vellambi, and F. Fekri. Rateless codes with unequal error protection property. *Information Theory, IEEE Transactions on*, 53(4):1521–1532, 2007. ISSN 0018-9448. doi: 10.1109/TIT.2007.892814.

N. Rahnavard and F. Fekri. Generalization of rateless codes for unequal error protection and recovery time: Asymptotic analysis. In *Information Theory, 2006 IEEE International Symposium on*, pages 523–527, 2006. doi: 10.1109/ISIT.2006.261771.

J. Hagenauer. Rate-compatible punctured convolutional codes (rcpc codes) and their applications. *Communications, IEEE Transactions on*, 36(4):389–400, 1988. ISSN 0090-6778. doi: 10.1109/26.2763.

Ming-Fong Tsai, Ce-Kuen Shieh, Chih-Heng Ke, and Der-Jiunn Deng. Sub-packet forward error correction mechanism for video streaming over wireless networks. *Multimedia Tools Appl.*, 47(1):49–69, March 2010. ISSN 1380-7501. doi: 10.1007/s11042-009-0406-5.

Ming-Fong Tsai, Ce-Kuen Shieh, Wen-Shyang Hwang, and Der-Jiunn Deng. An adaptive multi-hop forward error correction protection scheme for video streaming over wireless mesh networks. *Int. J. Commun. Syst.*, 22(10):1297–1318, October 2009. ISSN 1074-5351. doi: 10.1002/dac.v22:10.

A. Moid and A. Fapojuwo. Three-dimensional absorbing markov chain model for video streaming over ieee 802.11 wireless networks. *Consumer Electronics, IEEE Transactions on*, 54(4):1672–1680, 2008. ISSN 0098-3063. doi: 10.1109/TCE.2008.4711219.

L.B. Le, E. Hossain, and M. Zorzi. Queueing analysis for gbn and sr arq protocols under dynamic radio link adaptation with non-zero feedback delay. *Wireless Communications, IEEE Transactions on*, 6(9):3418–3428, 2007. ISSN 1536-1276. doi: 10.1109/TWC.2007.06020038.

Ming-Fong Tsai, Tzu-Chi Huang, Chih-Heng Ke, Ce-Kuen Shieh, and Wen-Shyang Hwang. Adaptive hybrid error correction model for video streaming over wireless networks. *Multimedia Systems*, 17(4):327–340, 2011b. ISSN 0942-4962. doi: 10.1007/s00530-010-0213-x.

Henning Schulzrinne. Rtp: A transport protocol for real-time applications. 1996.

ITU-R. Recommendation itu-r bt.500-13 methodology for the subjective assessment of the quality of television pictures. Technical report, ITU-T, January 2012.

ITU-T. A method for subjective performance assessment of the quality of speech voice output devices. Technical Report P.85, ITU-t, June 1994.

ITU-T. Mean opinion score (mos) terminology. Technical Report P800.1, ITU-T, March 2003.

ITU-R. Subjective video quality assessment methods for multimedia applications. Technical Report P.910, ITU-T, April 2008.

S.A. Karunasekera and N.G. Kingsbury. A distortion measure for blocking artifacts in images based on human visual sensitivity. *Image Processing, IEEE Transactions on*, 4(6):713–724, 1995. ISSN 1057-7149. doi: 10.1109/83.388074.

John O. Limb. Distortion criteria of the human viewer. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(12):778–793, 1979. ISSN 0018-9472. doi: 10.1109/TSMC.1979.4310129.

Mei-Yin Shen and C.-C.Jay Kuo. Review of postprocessing techniques for compression artifact removal. *Journal of Visual Communication and Image Representation*, 9(1):2 – 14, 1998. ISSN 1047-3203. URL `http://www.sciencedirect.com/science/article/pii/S1047320397903783`.

Michael Yuen and H.R. Wu. A survey of hybrid mc/dpcm/dct video coding distortions. *Signal Processing*, 70(3):247 – 278, 1998. ISSN 0165-1684. URL `http://www.sciencedirect.com/science/article/pii/S0165168498001285`.

N. Jayant, J. Johnston, and R. Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10):1385–1422, 1993. ISSN 0018-9219. doi: 10.1109/5.241504.

Brian W. Keelan. *Handbook of Image Quality*. CRC Press, 2002.

Weisi Lin. *Digital Video Image Quality and Perceptual Coding*, chapter Computational Models for Just-noticeable Difference, page 281–303. CRC Press, 2005.

J.M. Wolfe M.M. Chun. *Blackwell Handbook of Perception*, chapter Visual attention, page 272–310. Wiley-Blackwell, January 2001.

Ian Q. Whishaw Bryan Kolb. *Fundamentals of Human Neuropsychology*. Worth Publishers, 2008.

K. T. Tan and M. Ghanbari. A multi-metric objective picture-quality measurement model for mpeg video. *Circuits and Systems for Video Technology, IEEE Transactions on*, 10(7):1208–1213, 2000a. ISSN 1051-8215. doi: 10.1109/76.875525.

Zhenghua Yu, Hong Ren Wu, S. Winkler, and Tao Chen. Vision-model-based impairment metric to evaluate blocking artifacts in digital video. *Proceedings of the IEEE*, 90(1):154–169, 2002. ISSN 0018-9219. doi: 10.1109/5.982412.

Stefan Winkler. *Digital Video Image Quality and Perceptual Coding*, chapter Perceptual Video Quality Metrics - A Review, pages 155–179. CRC Press, 2005.

K. T. Tan and M. Ghanbari. Blockiness detection for mpeg2-coded video. *Signal Processing Letters, IEEE*, 7(8):213–215, 2000b. ISSN 1070-9908. doi: 10.1109/97.855443.

H.R. Wu and M. Yuen. A generalized block-edge impairment metric for video coding. *Signal Processing Letters, IEEE*, 4(11):317–320, 1997. ISSN 1070-9908. doi: 10.1109/97.641398.

P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. A no-reference perceptual blur metric. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 3, pages III–57–III–60 vol.3, 2002. doi: 10.1109/ICIP.2002.1038902.

Shiqian Wu, Weisi Lin, Zhongkang Lu, Ee Ping Ong, and Susu Yao. Blind blur assessment for vision-based applications. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 1639–1642, 2007. doi: 10.1109/ICME.2007.4284981.

J.Y.C. Chen and J.E. Thropp. Review of low frame rate effects on human performance. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 37(6):1063–1076, 2007. ISSN 1083-4427. doi: 10.1109/TSMCA.2007.904779.

Kai-Chieh Yang, C.C. Guest, K. El-Maleh, and P.K. Das. Perceptual temporal quality metric for compressed video. *Multimedia, IEEE Transactions on*, 9(7):1528–1535, 2007. ISSN 1520-9210. doi: 10.1109/TMM.2007.906576.

T. Oelbaum, C. Keimel, and K. Diepold. Rule-based no-reference video quality evaluation using additionally coded videos. *Selected Topics in Signal Processing, IEEE Journal of*, 3(2):294–303, 2009. ISSN 1932-4553. doi: 10.1109/JSTSP.2009.2015473.

H.R. Sheikh, M.F. Sabir, and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *Image Processing, IEEE Transactions on*, 15(11):3440–3451, 2006. ISSN 1057-7149. doi: 10.1109/TIP.2006.881959.

Marius Pedersen and Jon Yngve Hardeberg. Survey of full-reference image quality metrics. 2009.

Barry G. Haskell Arun N. Netravali. *Digital Pictures: Representation, Compression and Standards*. Plenum Press, 2nd edition, January 1995.

K. Piamrat, C. Viho, J. Bonnin, and A. Ksentini. Quality of experience measurements for video streaming over wireless networks. In *Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on*, pages 1184–1189, 2009. doi: 10.1109/ITNG.2009.121.

Jeffrey Lubin. *Vision Models for Target Detection and Recognition*, chapter A visual discrimination model for imaging system design and evaluation. World Scientific, 1995.

Zhou Wang and A.C. Bovik. A universal image quality index. *Signal Processing Letters, IEEE*, 9(3):81–84, 2002. ISSN 1070-9908. doi: 10.1109/97.995823.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004. ISSN 1057-7149. doi: 10.1109/TIP.2003.819861.

H.R. Sheikh and A.C. Bovik. Image information and visual quality. *Image Processing, IEEE Transactions on*, 15(2):430–444, 2006. ISSN 1057-7149. doi: 10.1109/TIP.2005.859378.

D.M. Chandler and S.S. Hemami. Vsnr: A wavelet-based visual signal-to-noise ratio for natural images. *Image Processing, IEEE Transactions on*, 16(9):2284–2298, 2007. ISSN 1057-7149. doi: 10.1109/TIP.2007.901820.

Feng Xiao et al. Dct-based video quality evaluation. *Final Project for EE392J*, page 769, 2000.

Andrew B. Watson. Toward a perceptual video-quality metric, 1998.

Stefan Winkler. A perceptual distortion metric for digital color video. In *in Proc. SPIE*, pages 175–184, 1999.

Samir Mohamed and Gerardo Rubino. A study of real-time packet video quality using random neural networks. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(12):1071–1083, 2002.

E. Aguiar, A. Riker, M. Mu, S. Zeadally, E. Cerqueira, and A. Abelem. Real-time qoe prediction for multimedia applications in wireless mesh networks. In *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*, pages 592–596, 2012. doi: 10.1109/CCNC.2012.6181017.

ITU-T. Recommendation h.262, 3 2013.

A.K. Jain, R. P W Duin, and Jianchang Mao. Statistical pattern recognition: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22 (1):4–37, 2000. ISSN 0162-8828. doi: 10.1109/34.824819.

Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*, volume 1. Cambridge University Press Cambridge, 2008.

Alexander Sturn, John Quackenbush, and Zlatko Trajanoski. Genesis: cluster analysis of microarray data. *Bioinformatics*, 18(1):207–208, 2002.

Vladimir Estivill-Castro. Why so many clustering algorithms: A position paper. *SIGKDD Explor. Newsl.*, 4(1):65–75, June 2002. ISSN 1931-0145. doi: 10.1145/568574.568575. URL `http://doi.acm.org/10.1145/568574.568575`.

StephenC. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967. ISSN 0033-3123.

Dragomir R Radev, Vasileios Hatzivassiloglou, and Kathleen R McKeown. A description of the cidr system as used for tdt-2. In *Broadcast News Workshop'99 Proceedings*, page 205. Morgan Kaufmann Pub, 1999.

Jose L Marroquin and Federico Girosi. Some extensions of the k-means algorithm for image segmentation and pattern classification. Technical report, DTIC Document, 1993.

Xiaowei Xu, Martin Ester, H-P Kriegel, and Jörg Sander. A distribution-based clustering algorithm for mining in large spatial databases. In *Data Engineering, 1998. Proceedings., 14th International Conference on*, pages 324–331. IEEE, 1998.

Jeffrey D Banfield and Adrian E Raftery. Model-based gaussian and non-gaussian clustering. *Biometrics*, pages 803–821, 1993.

Hans-Peter Kriegel, Peer Kröger, Jörg Sander, and Arthur Zimek. Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(3):231–240, 2011.

Martin Ester, Hans peter Kriegel, Jörg S, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996.

Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, 2010.

Ian Jolliffe. *Principal component analysis.* Wiley Online Library, 2005.

Gilbert Saporta and Ndèye Niang. Principal component analysis: application to statistical process control. *Data analysis*, pages 1–23, 2009.

Jonathon Shlens. A tutorial on principal component analysis. *Systems Neurobiology Laboratory, University of California at San Diego*, 2005.

Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

Donald Olding Hebb. *The organization of behavior: A neuropsychological theory.* Psychology Press, 2002.

Šarūnas Raudys. Evolution and generalization of a single neurone: I. single-layer perceptron as seven statistical classifiers. *Neural Networks*, 11(2):283–296, 1998.

George Bebis and Michael Georgiopoulos. Feed-forward neural networks. *Potentials, IEEE*, 13(4):27–31, 1994.

Richard P Lippmann. Pattern classification using neural networks. *Communications Magazine, IEEE*, 27(11):47–50, 1989.

DS Broomhead and David Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. Technical report, DTIC Document, 1988.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8): 2554–2558, 1982.

Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

Robert A Jacobs and Michael I Jordan. Learning piecewise control strategies in a modular neural network architecture. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(2):337–345, 1993.

Eric R Kandel, James H Schwartz, Thomas M Jessell, et al. *Principles of neural science*, volume 4. McGraw-Hill New York, 2000.

Erol Gelenbe. Random neural networks with negative and positive signals and product form solution. *Neural computation*, 1(4):502–510, 1989.

Erol Gelenbe. Stability of the random neural network model. *Neural computation*, 2(2):239–247, 1990.

Samir Mohamed, Gerardo Rubino, and Martın Varela. Performance evaluation of real-time speech through a packet network: a random neural networks-based approach. *Performance Evaluation*, 57(2):141–161, 2004.

Roger B Myerson. *Game theory: analysis of conflict.* Harvard university press, 2013.

John Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295, 1951.

John Von Neumann and Oskar Morgenstern. *Theory of games and economic behavior (commemorative edition)*. Princeton university press, 1944.

Vijay Krishna. *Auction theory*. Academic press, 2009.

Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics*. Cambridge University Press, 1998.

James H Morris, Mahadev Satyanarayanan, Michael H Conner, John H Howard, David S Rosenthal, and F Donelson Smith. Andrew: A distributed personal computing environment. *Communications of the ACM*, 29(3):184–201, 1986.

Tim Roughgarden. Stackelberg scheduling strategies. *SIAM Journal on Computing*, 33(2):332–350, 2004.

Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.

Martin J Osborne. *A course in game theory*. Cambridge, Mass.: MIT Press, 1994.

Manfred Gilli and Peter Winker. Heuristic optimization methods in econometrics. *Handbook of computational econometrics*, pages 81–119, 2009.

Peter Winker and Manfred Gilli. Applications of optimization heuristics to estimation and modelling problems. *Computational Statistics & Data Analysis*, 47 (2):211–223, 2004.

Peter JM Van Laarhoven and Emile HL Aarts. *Simulated annealing*. Springer, 1987.

Fred Glover, Manuel Laguna, et al. *Tabu search*, volume 22. Springer, 1997.

John H Holland. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press, 1975.

Agoston E Eiben, P-E Raue, and Zs Ruttkay. Genetic algorithms with multi-parent recombination. In *Parallel Problem Solving from Nature—PPSN III*, pages 78–87. Springer, 1994.

Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

Marco Dorigo, Vittorio Maniezzo, and Alberto Colorni. Ant system: optimization by a colony of cooperating agents. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 26(1):29–41, 1996.

Marco Dorigo and Thomas Stutzle. *Ant Colony Optimization*. MIT Press, 2004.

G Navarro Varela and Mark C Sinclair. Ant colony optimisation for virtual-wavelength-path routing and wavelength allocation. In *Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on*, volume 3. IEEE, 1999.

Helena Ramalhinho Lourenço and Daniel Serra. Adaptive search heuristics for the generalized assignment problem. *Mathware & soft computing*, 9(3):209–234, 2002.

Christian Blum. Aco applied to group shop scheduling: A case study on intensification and diversification. In *Ant Algorithms*, pages 14–27. Springer, 2002.

Rafael S Parpinelli, Heitor S Lopes, and Alex A Freitas. An ant colony algorithm for classification rule discovery. *Data mining: A heuristic approach*, 208:191–132, 2002.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

Richard Mojena. Hierarchical grouping methods and stopping rules: an evaluation. *The Computer Journal*, 20(4):359–363, 1977.

Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

Hossam Eldin Abdelbaki. Rnnsimv.2. Technical report, School of Computer Science, University of Central Florida, September 1999.

MATLAB. *version 8.2 (R2013b).* The MathWorks Inc., Natick, Massachusetts, 2013.

Jirka Klaue, Berthold Rathke, and Adam Wolisz. Evalvid - a framework for video transmission and quality evaluation. In *In Proc. of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pages 255–272, 2003.

George F. Riley and Thomas R. Henderson. The ns-3 network simulator modeling and tools for network simulation. In Klaus Wehrle, Mesut Güneş, and James Gross, editors, *Modeling and Tools for Network Simulation*, chapter 2, pages 15–34. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-12330-6.

The it++ library. URL `http://itpp.sourceforge.net`.

S. Chikkerur, V. Sundaram, M. Reisslein, and L.J. Karam. Objective video quality assessment methods: A classification, review, and performance comparison. *Broadcasting, IEEE Transactions on*, 57(2):165–182, 2011. ISSN 0018-9316. doi: 10.1109/TBC.2011.2104671.

Dmitriy Vatolin, Alexey Moskvin, Oleg Petrov, Sergey Putilin, Sergey Grishin, and Arsaev Marat. Video quality measurement tool, 2011.

ISO/IEC. Information technology — coding of audio-visual objects — part 14: Mp4 file format, 11 2003.

Thomas Clausen, Philippe Jacquet, Cédric Adjih, Anis Laouiti, Pascale Minet, Paul Muhlethaler, Amir Qayyum, Laurent Viennot, et al. Optimized link state routing protocol (olsr). 2003.

Rouzbeh Razavi, Martin Fleury, and Mohammed Ghanbari. Adaptive packet-level interleaved fec for wireless priority-encoded video streaming. *Adv. MultiMedia*, 2009:3:1–3:14, January 2009. ISSN 1687-5680. doi: 10.1155/2009/982867. URL `http://dx.doi.org/10.1155/2009/982867`.

Leif Wilhelmsson and Laurence B Milstein. On the effect of imperfect interleaving for the gilbert-elliott channel. *Communications, IEEE Transactions on*, 47(5): 681–688, 1999.

Gerhard Haßlinger and Oliver Hohlfeld. The gilbert-elliott model for packet loss in real time services on the internet. In *Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB), 2008 14th GI/ITG Conference-*, pages 1–15. VDE, 2008.

Nattapat Attiratanasunthron and Jittat Fakcharoenphol. A running time analysis of an ant colony optimization algorithm for shortest paths in directed acyclic graphs. *Information Processing Letters*, 105(3):88 – 92, 2008. ISSN 0020-0190. doi: http://dx.doi.org/10.1016/j.ipl.2007.08.013. URL `http://www.sciencedirect.com/science/article/pii/S0020019007002190`.

J. Lacan. Reed-solomon forward error correction (fec) schemes, April 2009.

Appendix A

Publications

This Appendix presents all the published papers during the development of this thesis.

Adaptive Motion-aware FEC-based Mechanism to Ensure Video Transmission

Roger Immich[1], Pedro Borges
University of Coimbra
Coimbra, Portugal
immich@dei.uc.pt,
pborges@student.dei.uc.pt

Eduardo Cerqueira[1]
Federal University of Para - Belem, Brazil
University of California Los Angeles (UCLA)
Los Angeles, USA
cerqueira@upfa.br

Marilia Curado
University of Coimbra
Coimbra, Portugal
marilia@dei.uc.pt

***Abstract*—Video transmission over wireless networks has shown a great increase in recent years and it is becoming part of our daily life. Meanwhile, several difficulties can impair the success of the transmission, such as limited network resources, high error rates and fluctuating signal strength that may lead to variable bandwidth. Therefore there is the need for adaptive mechanisms that can provide a good video transmission. Adaptive Forward Error Correction (FEC) techniques which assure Quality of Experience (QoE) are a convenient means of delivering video data to wireless users in dynamic and error prone networks, while taking into account the content of the transmitted data. This paper proposes an adaptive content-aware and Random Neural Network (RNN) based mechanism to provide protection of real-time video streams against packet loss in wireless networks, improving user experience and optimising network resources. The benefits of the proposed mechanism are demonstrated through simulations and assessed with QoE metrics.**

***Index Terms*—Motion Vectors (MV); Forward Error Correction (FEC); Video-aware FEC; QoE; Neural networks; Unequal Error Protection (UEP)**

I. Introduction

In the last few years the use of real-time video services in wireless and mobile networks has seen a dramatical increase [1], namely in Europe where the growth in mobile video views was of 162% from 2011 to 2012. The outlook for growth, according to Cisco, is even greater and will represent over 90% of the global IP traffic by 2015 [2]. This viewpoint can be explained by the substantial amount of new forms of entertainment and information, which include news websites, social networking communities, e-learning as well as large amounts of user-generated content.

Due to the video traffic growth, it becomes critical to improve its transmission quality. However, the network resources are not unlimited, and therefore, many factors, such as propagation loss, congestion, and channel noise, can hinder the transmission. This may cause packet loss which in turn degrades video quality and negatively impacts the end user experience. The quality of the video should be assessed in terms of Quality of Experience (QoE), because contrary to Quality of Service (QoS), it assesses the quality of the video as perceived by the end user. Therefore, QoE must be considered in the process of building a mechanism that can adapt to the video and network characteristics.

An adaptive mechanism is needed to combat the information loss on networks with different characteristics. This mechanism should improve the video transmission to ensure that the viewing experience of the users is not negatively affected by the impairing factors of the wireless environment. Retransmission and Forward Error Correction (FEC) are two commonly used techniques to handle network limitations. Retransmission mechanisms are suitable for systems where delivery must be guaranteed. However, in a real-time video transmission environment, what is most important is the timely delivery of the content, because if a frame arrives after its decoding deadline it can no longer be displayed. On the other hand, FEC mechanisms have delay-constraints. These mechanisms add redundancy to the original data so that it is possible to correct, without retransmission, eventual errors or losses caused due to the characteristics of the network. The problem is that many times they are non-adaptive and strict. This means that the mechanism only offers a fixed amount of protection/redundancy to the data to be transmitted, not taking into account the video characteristics. This can lead to a poor utilization of network resources and it can cause network congestion due to unnecessary overhead.

Taking these problems into account, this paper proposes a novel adaptive Video-aware Random Neural Networks (RNN) based mechanism (neuralFEC). It aims at overcoming the limitations of non-adaptive schemes, such as the inability to take into consideration the video's motion intensity, which is crucial to a high QoE. An efficient way to quantify the motion intensity is through Motion Vectors (MV). These vectors play a key part in the video compression process, allowing to store changes from adjacent frames, including both previous and future frames. Therefore it is possible to quantify the motion intensity of a given frame using the information inside his MV. The proposed mechanism mitigates these problems by adaptively selecting the amount of redundancy given to individual frames, through the analysis of their type and their motion characteristics through a RNN [3]. Neural networks (NN) are computational models inspired by biological central nervous systems, which are able to go through the process of machine learning and pattern recognition. They can be trained by feed-

[1]CNPq Fellow - Brazil

ing them with learning patterns and letting them change the weights according to some learning rule. The choice of RNN is related to their particular characteristics that enable their success in pattern recognition and classification problems [4].

Another important feature of neuralFEC is the use of a Unequal Error Protection (UEP) scheme. This allows the use of different amounts of redundancy to different parts of the data. The proposed UEP scheme protects frames with a greater intensity of movement with a larger amount of redundancy opposed to those with lower intensities of movement. This scheme reduces network overhead through the adaptive and selective use of redundancy while also improving or at least maintaining the level of QoE of the transmitted videos when compared to non-adaptive FEC mechanisms.

The remainder of this paper is structured as follows. The related work is presented in Section II. Section III describes the neuralFEC and its evaluation is presented in Section IV. Conclusions and Future Work are summarized in Section V.

II. Related Work

Adaptive FEC mechanisms have been gaining an increased interest because they allow for the redundant data to be used according to the video and network characteristics. The APB-FEC scheme aims to solve the problem of conventional packet level FEC by the use of a smaller packet length, while increasing the FEC block length [5]. By using feedback information from the receiver regarding the correct reception of the packets, the mechanism adapts the video in the streaming buffers to the network conditions. The use of buffers is not optimal and can increase the delay. Also, relying on information from the receiver can be problematic due to the fact that if the communication is hindered it is very probable that the feedback information will not reach the sender.

The ViewFEC module-based mechanism [6] uses a database module with motion and complexity information from several videos that is built prior to video distribution. It also uses another module to assess the length of the Group of Pictures (GoP) which has a greater influence in the received video quality if packets are lost during transmission. The motion intensity classification adopts a heuristic based comparison with the database, which may lead to non-optimal results.

Another FEC mechanism uses a layer-based approach which processes each GoP and divides it into layers of different utility to adequately distribute FEC among them [7]. This process is very time consuming and takes more time with larger GoP sizes, which in turn adds delay to the transmission. If the layer-division is not already processed, this mechanism requires a great amount of processing which means it cannot be used in real-time video transmission.

Also related to the use of GoP information is the optimized cross layer FEC mechanism which assigns different priorities to the GoPs according to the commutative mean squared error of the entire GoP [8]. The GoPs are encoded and cyclic redundancy check bits are added to detect coding errors. Afterwards the FEC codes are optimized with different parameters for different situations. This mechanism has several optimization phases, for each frame, that are very time consuming, increasing the delay and degrading the QoE.

III. Adaptive mechanism

The issues presented above, together with the need to further optimize the use of network resources, make way for the novel mechanism proposed in this paper. The neuralFEC goal is to use an optimized UEP scheme according to the video frame's motion intensity characteristics. Making it possible to better protect the most QoE-sensitive frames, therefore providing both, the reduction on the impact of packet loss on the video quality and the use of only the necessary amount of redundancy. Through this, it is possible to distribute the redundancy in a way that will improve the QoE for the end user while sparing precious network resources.

Figure 1 depicts the overall operation of the neuralFEC mechanism. First of all, in the offline process, an exploratory analysis using hierarchical clustering was carried out to train the RNN. Through the human experience about the intrinsic video characteristics and several simulation experiments the RNN was also validated. After this, the RNN can be used in real-time. The offline process is very important because it leads to accurate results in real-time. Then the decision making process conducted by the RNN determines a specific amount of redundancy needed by each frame. This allows the neuralFEC to shield only the QoE-sensitive data against packet loss, resulting on better video quality as perceived by the end-users while saving network resources. A detailed explanation of the proposed mechanism is presented afterwards.

In order to perform the classification of each frame according to its motion intensity a RNN was employed. As other NNs, this model has the capability for learning and generalization, but excels in pattern recognition and classification problems [4]. By training the network successfully with an adequate range of video samples, it can be used in real-time to classify a given video sequence according to the intensity of the movement. This is achieved by attributing a specific value to determine frames with different motion intensities. After that, the neuralFEC is able to select in real-time the appropriate amount of redundancy to be transmitted so that the network overhead is minimized and QoE is maximized.

The RNN structure consists of three input nodes, seven hidden layer nodes and one output. The three input nodes represent each frame's characteristics specifically frame size, frame type, and MV ratio (total number of motion vectors divided by the distance described by them). The MV in a video sequence was adopted from the classical mechanics and vector-oriented model of motion used by them. This is a simple model that defines the movement of objects as the progression of small translations on a plane. A MV ratio was used because a certain frame can have several vectors pointing to a close distance while other frames can have less vectors pointing further away however, and consequently defining a situation of higher motion intensity. Finally, the output node provides the motion intensity classification value, computed by the network from the given inputs. Through these parameters it is possible

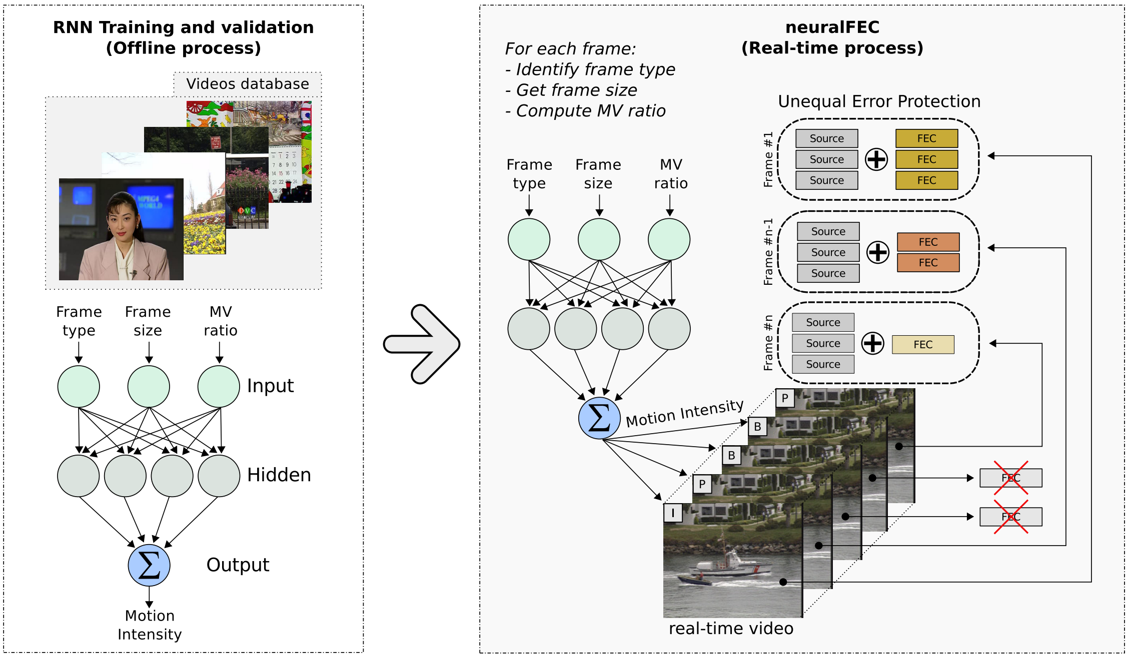

Fig. 1. neuralFEC mechanism

to characterize the video motion intensity and choose the optimal amount of redundancy on a frame-by-frame basis.

To adequately train the RNN, an exploratory hierarchical cluster analysis using Ward's method [9] was performed to categorize selected video sequences which represent different types of movement. The video sequences were selected according to the recommendations of the Video Quality Experts Group (VQEG) [10] and the International Telecommunications Union (ITU) [11], and represent sequences that cover different distortions and content that are commonly seen in online videos. A set of 15 videos was selected to perform the hierarchical cluster analysis. Each video was broken down into three parameters, namely about the frame size and type, and the total number of motion vectors divided by the distance described by these vectors.

Using the exploratory analysis results, the video sequences were classified into three categories of motion intensity, namely low, median and high intensity. Afterwards, two videos of each motion intensity category were randomly selected to train the RNN. The training of the RNN consisted in feeding the information of this set of selected videos to the inputs of the network for about 600 iterations which was the point at which the Minimum Mean Squared Error (MSE) stabilized. After the training of the RNN was completed, the RNN was validated with a different set of video sequences, the remaining 9 videos from the exploratory analysis, which also cover all three motion intensity characteristics. The results obtained proved to be correct classifications of the video sequences present in the validation set.

After the training and validation phases, the RNN can be used in the real-time process. Using cross-layer techniques, the neuralFEC is able to obtain important information about several video characteristics, namely frame type and size, as well as the number of motion vectors and the Euclidean distance pointed by them. All these details are fed to the RNN which in turn provides, in real-time, an accurate motion intensity value for each frame. After the video frame is classified it is encoded by the Reed-Solomon (RS) [12] with the amount of redundancy selected by the RNN. This erasure code has low complexity, and therefore provides better performance for real-time services [13]. The RS is a linear block code that has a rule to convert of source bits s, of length k, into a transmitted sequence t of length n bits. To add redundancy, n is made greater than k. In a linear block code, the extra $(n-k)$ bits are linear functions of the original k bits, which are called parity-check bits. It allows us to use different block sizes represented by each individual frame. Through this procedure, a precise UEP amount can be assigned to each frame, where only the QoE-sensitive data will be protected. In turn, this results in better video quality while reducing the amount of redundancy data needed, thus not adding unnecessary the network overhead. This reduction is very important because it can be the source of serious interference problems. In this way, not adding unnecessary redundancy will allow more users to access services with better QoE, improving the overall system performance.

Figure 2 shows the pseudo-code portraying the neuralFEC real-time operation. All procedures are performed inside a for-loop, at line 01, which will go through all the frames in the video sequence. At line 02, the frame type is identified to be used in the selection control mechanism (if statement) at line 03. This allows the change in the control flow according to neuralFEC needs, which is, to assign a tailored redundancy amount to I- and P-Frames, and send B-Frames without additional data. At lines 05, 06, 07, and 08 it is possible to observe the identification of the frame size, the computation of the MV ratio, the classification of the video frame motion intensity using the RNN, and the assignment of an unequal amount of redundancy to the most QoE-sensitive data, respectively. Lines 09 and 11 are responsible for sending the frame with or without redundancy.

```
01 for each Frame
02  FT=getFrameType(Frame)
03  if(FT equal (I- or P-Frame))
04   then
05    FS=getFrameSize(Frame)
06    MVratio=calculateRatio(getMV(Frame))
07    MotionIntensity=RNN(FT, FS, MVratio)
08    addRedundancy(RS(MotionIntensity))
09    sendFrame(Frame+Redundancy)
10   else
11    sendFrame(Frame)
12  end if
13 end for
```

Fig. 2. neuralFEC pseudo-code

IV. Performance Evaluation and Results

The main objective of the neuralFEC is to reduce the network overhead by not adding unnecessary redundancy while having a light improvement or at least maintain the same video quality. In order to assess the performance of the proposed mechanism in wireless networks, several experiments were performed using Network-Simulator 3 (NS-3) [14]. The scenario for evaluation is comprised of 25 nodes in a grid disposition (5x5), separated by 50 meters. The Optimized Link State Routing Protocol (OLSR) [15] was used as the routing protocol. Ten video sequences were used in this scenario, namely Bowing, Coastguard, Container, Crew, Foreman, Hall, Harbour, Mother and Daughter, News and Soccer. These particular sequences were selected in order to have a great variety of motion intensities. They are in Common Intermediate Format (CIF) with a resolution of (352x288) and coded with the H.264 codec. The GoP size was set to 19 and after each I- or P-frames, come two B-frames.

A two-state discrete-time Markov chain model was implemented following a simplified Gilbert-Elliot packet-loss model [16], which approximates the behaviour of a wireless network. It produces simulation results which are closely related to those of burst loss patterns of wireless channels [17]. The simplified Gilbert-Elliot is shown in Figure 3, where the probability of packet loss in the Good state (G) was set at 0, which means no losses, and the probability of packet loss in the Bad state (B) was set at 1, where all packets are lost. The Packet Loss Rate (PLR) can be obtained by Equation 1, where P_{BG} represents the probability of transitioning from the Bad state to the Good state and vice-versa with P_{GB}.

$$PLR = \frac{P_{BG}}{P_{BG} + P_{GB}} \qquad (1)$$

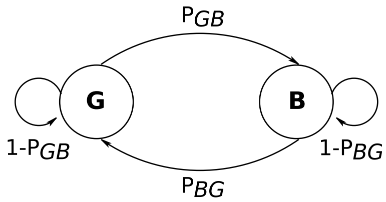

Fig. 3. Gilbert-Elliot simplified model

To validate and compare the results, three experiments with different schemes were performed. The first experiment serves as a baseline as there was no FEC mechanism in use. The second experiment was performed with a non-adaptive video-aware FEC mechanism (Video-aware FEC), where a fixed amount of 38% of redundancy was added only to I- and P-frames. This amount of redundancy was selected according to an extensive set of experiments, which showed the best video quality situation taking into consideration the characteristics of the scenario defined for the experiment. The last experiment is the proposed adaptive mechanism with RNN and UEP (neuralFEC). Each of the three experiments was simulated 10 times with an error rate of 20% representing an average loss [18] obtained through the simplified Gilbert-Elliot model.

The video quality of each evaluation scenario was assessed through an objective measurement, namely the Structural Similarity (SSIM) Index. This objective evaluation metric is a method based on the analysis of luminance, contrast and structural similarity of images. SSIM is one of the most commonly used metrics for objective evaluation of QoE [19]. The objective quality assessment of the video sequences was performed with EvalVid [20] and the MSU Video Quality Measurement Tool (VQMT) [21].

A set of ten video sequences different from those used to train the RNN was used. They represent distinctive situations with a broad type of motion intensity characteristics. Three different mechanisms were used to perform the experiments on the mentioned sequences, namely without protection, with the non-adaptive Video-aware FEC mechanism and finally with the adaptive neuralFEC mechanism.

Figure 4 shows the results in terms of overhead of the experiments. While using the non-adaptive Video-aware FEC mechanism the network overhead added was between 35% and 43%. On the other hand, when the neuralFEC mechanism was employed the amount of overhead remained between 13% and 24%. This means that the average redundancy added by the non-adaptive mechanism was around 38% on contrast to only 19% added by neuralFEC. It is also clear that the proposed mechanism can assess the importance of frames according to motion intensity. This assessment is performed by the RNN, which attributes a higher classification for frames with a great amount of movement, and a lower classification for frames with less amount of movement. In doing that, a greater amount of redundancy was attributed to video sequences such as Crew, Soccer, Harbour and Coastguard. On

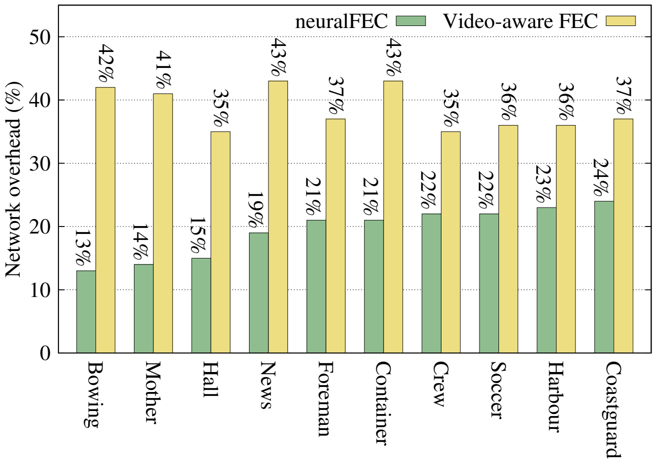

Fig. 4. Network Overhead

the contrary, video sequences which are classified as being of lesser motion intensity, such as Bowing, Mother and Hall are given less redundancy. These results show that the neuralFEC mechanism performs better than the non-adaptive Video-aware FEC mechanism in terms of overhead, by reducing in average a half of the redundancy needed to protect the data.

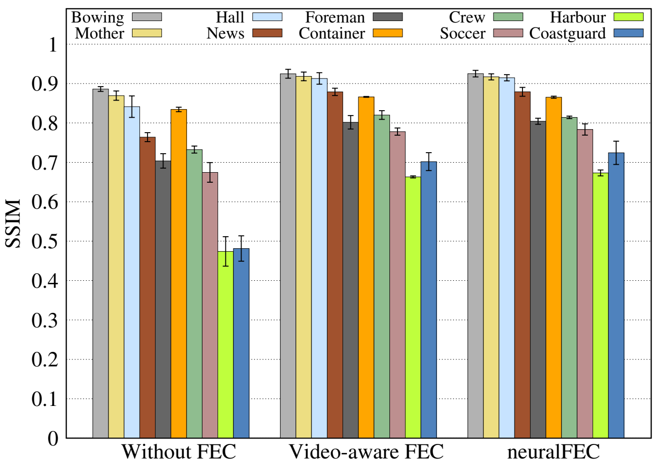

Fig. 5. Objective QoE assessment (SSIM)

Besides saving the already scarce network resources by not adding unnecessary redundancy, it is also important to provide good video quality. In order to verify this situation, a set of assessments were performed using the SSIM metric. Figure 5 depicts the SSIM values for each video sequence while using the three aforementioned protection schemes. The results show the neuralFEC mechanism obtained an average SSIM value of 0,831 against a value of 0,819 for the video-aware FEC mechanism and 0,726 for the mechanism that did not use any type of protection. This represents a slight improvement of almost 1,5% , on average, in terms of SSIM value for the adaptive neuralFEC mechanism in comparison to the non-adaptive video-aware FEC mechanism. In further detail, the SSIM score achieved by neuralFEC for the Harbour video sequence was of 0,675 against 0,662 for the video-aware mechanism and 0,485 for the mechanism without FEC. Although all video were transmitted with the same PLR, the SSIM score obtained by the same three mechanisms for the Bowing sequence was of 0,915, 0,914 and 0,920 respectively. This can be explained by the different characteristics of these two sequences. The Harbour video sequence has a greater amount of motion compared to the Bowing sequence, meaning that packet loss has a greater effect on this type of sequences. This results in lower SSIM scores for sequences with a higher degree of motion intensity, and also shows that videos with a lower degree of motion intensity have greater resilience to packet loss. Due to this, it is important to employ adaptive FEC mechanisms, such as neuralFEC to protect the contents of the video taking into account its motion intensity characteristics.

TABLE I
AVERAGE SSIM AND NETWORK OVERHEAD

	neuralFEC	**Video-aware FEC**	**Without FEC**
SSIM	0,831	0,819	0,726
Overhead	19,334%	38,460%	–

Table I summarizes the results presenting the average SSIM and network overhead for all video sequences. It demonstrates that the proposed neuralFEC mechanism had a slightly improved video quality. Most importantly, it was able to do so while drastically reducing the network overhead by not adding unnecessary redundancy. This is of great importance in wireless networks, due to the limited nature of the wireless channel resources, which can be aggravated by packet loss due to interference from concurrent transmissions and network congestion.

The results showed that the neuralFEC mechanism, through an accurate motion intensity classification of video sequences with distinct characteristics, is able to add a precise amount of protection. In doing that, it can offer less overhead during transmission in a wireless mesh network setting while providing as good video quality as non-adaptive FEC mechanisms.

V. CONCLUSION AND FUTURE WORKS

The growth of the online video transmission over wireless networks calls for adaptive QoE-aware mechanism to ensure the video quality. To fill this gap, the neuralFEC provides the possibility to shield the video transmission in wireless networks, protecting only the most QoE-sensitive data, maximizing the video quality while saving network resources by not sending unnecessary redundancy. This is important to better use the already scarce wireless resources. Both impact and advantages of the neuralFEC approach were demonstrated through a set of experiments using real video sequences.

The experimental simulation results showed that neuralFEC was able to highly reduce the amount of network overhead by 50% while maintaining or even improving the QoE for the end user. This is a great enhancement over non-adaptive FEC mechanisms and also reinforces the importance of using adaptive FEC mechanisms which take into account motion intensity when protecting a video stream with varying characteristics. Future work should emphasize further refinement of the mechanism by taking into account a larger set of video sequences. Additionally, other adaptive FEC mechanisms will be used to evaluate the performance of neuralFEC. Different scenarios should be explored by introducing mobility and cross-traffic to assess the resilience of the mechanism in such conditions.

Acknowledgment

This work was funded by the Brazilian National Counsel of Technological and Scientific Development (CNPq) and also supported by the Intelligent Computing in the Internet of Services (iCIS) project (CENTRO-07-ST24-FEDER-002003), co-financed by QREN, in the scope of the Mais Centro Program.

References

[1] comScore, "More than 200 billion online videos viewed globally in october,," comScore inc., www.comscore.com/Press Events/Press Releases/2011/12/More than 200 Billion Online Videos Viewed Globally in October, Tech. Rep., 2011.

[2] Cisco, "Cisco visual networking index: Forecast and methodology 2010-2015," Cisco, Tech. Rep., 2011.

[3] A. Abraham, "Artificial neural networks," *handbook of measuring system design*, 2005.

[4] S. Mohamed and G. Rubino, "A study of real-time packet video quality using random neural networks," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, no. 12, pp. 1071–1083, 2002.

[5] M.-F. Tsai, N. Chilamkurti, and C.-K. Shieh, "An adaptive packet and block length forward error correction for video streaming over wireless networks," *Wireless Personal Communications*, vol. 56, no. 3, pp. 435–446, 2011.

[6] R. Immich, E. Cerqueira, and M. Curado, "Cross-layer fec-based mechanism for packet loss resilient video transmission," in *Data Traffic Monitoring and Analysis*, ser. Lecture Notes in Computer Science, E. Biersack, C. Callegari, and M. Matijasevic, Eds. Springer Berlin Heidelberg, 2013, vol. 7754, pp. 320–336.

[7] V. Lecuire, "Unequal error protection under bitrate constraint for video streaming over internet," *Computer Communications*, vol. 35, no. 3, pp. 287 – 297, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366411003161

[8] A. Talari, S. Kumar, N. Rahnavard, S. Paluri, and J. Matyjas, "Optimized cross-layer forward error correction coding for h.264 avc video transmission over wireless channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–13, 2013.

[9] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.

[10] N. Staelens, I. Sedano, M. Barkowsky, L. Janowski, K. Brunnstrom, and P. Le Callet, "Standardized toolchain and model development for video quality assessmentthe mission of the joint effort group in vqeg," in *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*. IEEE, 2011, pp. 61–66.

[11] T. S. S. O. ITU, *Objective perceptual multimedia video quality measurement in the presence of a full reference*, ITU-T Std. J.247, August 2008.

[12] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, June 1960.

[13] J. Neckebroek, M. Moeneclaey, and E. Magli, "Comparison of reed-solomon and raptor codes for the protection of video on-demand on the erasure channel," in *Information Theory and its Applications (ISITA), 2010 International Symposium on*, 2010, pp. 856–860.

[14] G. F. Riley and T. R. Henderson, "The ns-3 network simulator modeling and tools for network simulation," in *Modeling and Tools for Network Simulation*, K. Wehrle, M. Güneş, and J. Gross, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, ch. 2, pp. 15–34.

[15] T. Clausen, P. Jacquet, C. Adjih, A. Laouiti, P. Minet, P. Muhlethaler, A. Qayyum, L. Viennot *et al.*, "Optimized link state routing protocol (olsr)," 2003.

[16] R. Razavi, M. Fleury, and M. Ghanbari, "Adaptive packet-level interleaved fec for wireless priority-encoded video streaming," *Adv. MultiMedia*, vol. 2009, pp. 3:1–3:14, Jan. 2009. [Online]. Available: http://dx.doi.org/10.1155/2009/982867

[17] L. Wilhelmsson and L. B. Milstein, "On the effect of imperfect interleaving for the gilbert-elliott channel," *Communications, IEEE Transactions on*, vol. 47, no. 5, pp. 681–688, 1999.

[18] R. Immich, E. Cerqueira, and M. Curado, "Adaptive video-aware fec-based mechanism with unequal error protection scheme," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, ser. SAC '13. New York, NY, USA: ACM, 2013, pp. 981–988. [Online]. Available: http://doi.acm.org/10.1145/2480362.2480550

[19] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 165–182, 2011.

[20] J. Klaue, B. Rathke, and A. Wolisz, "Evalvid - a framework for video transmission and quality evaluation," in *In Proc. of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, 2003, pp. 255–272.

[21] D. Vatolin, A. Moskvin, O. Petrov, S. Putilin, S. Grishin, and A. Marat. (2011) Video quality measurement tool. Moscow State University.

AntMind: Enhancing Error Protection for Video Streaming in Wireless Networks

Roger Immich[1], Pedro Borges
University of Coimbra
Coimbra, Portugal
immich@dei.uc.pt,
pborges@student.dei.uc.pt

Eduardo Cerqueira[1]
Federal University of Para - Belem, Brazil
University of California Los Angeles (UCLA)
Los Angeles, USA
cerqueira@upfa.br

Marilia Curado
University of Coimbra
Coimbra, Portugal
marilia@dei.uc.pt

***Abstract*—On-line video services are becoming a large part of the daily routines of people all over the world, where most of the content is accessed through wireless networks. Therefore, it is of ever growing importance that the negative aspects of these types of error prone networks are lessened in order to ensure adequate quality of the delivered video streams. Forward Error Correction (FEC) techniques allow the stream to be protected with an amount of redundancy to preserve the video quality during transmission. Nevertheless, some FEC schemes do not make an efficient usage of the available network resources due to unnecessary use of redundancy as a result of video-unawareness. The adaptive FEC mechanism proposed in this paper uses the motion intensity characteristics of the video and the network loss state to deliver the video streaming with adequate Quality of Experience (QoE), while keeping the use of network resources to a minimum level. It does so from a combined use of a Random Neural Network (RNN) for motion intensity classification and an Ant Colony Optimization (ACO) scheme for dynamic redundancy allocation. QoE metrics are used to assess the performance of the mechanism showing its advantages over adaptive and non-adaptive protection schemes.**

***Index Terms*—Motion Vectors (MV); Forward Error Correction (FEC); Video-aware FEC; QoE; Neural networks; Unequal Error Protection (UEP); Ant Colony Optimization**

I. Introduction

The usage of real-time video services on the go has grown in recent years [1]. This leap can be explained by the technological advances in mobile devices and the increasing popularity of content sources which flood the wireless systems every day. As video traffic increases, the probability of errors derived from network congestion and interference rises, especially in a wireless network environment. Such factors will negatively impact the video stream quality, leading to diminished Quality of Experience (QoE) for the end user.

Video sequences have different characteristics which describe a different Motion Intensity (MI), such as steady camera sequences (low MI), and panning and zooming sequences (high MI). When losses occur on a high MI frame, video quality will be more degraded than on a frame with less movement. Therefore, videos with a greater MI need better protection than those with a lower one. To preserve the video quality while maximizing the QoE, it is of utmost importance to employ a protection mechanism that can adapt to the video and network's characteristics. Hence, in order to assess the impact of such characteristics in the perceived video quality by humans, QoE metrics must be used.

Two commonly used techniques for error recovery are Retransmission and Forward Error Correction (FEC). Retransmission schemes, such as Automatic Repeat Request (ARQ), require a data reception acknowledgement in order to retransmit lost data. This ensures complete delivery of the data but introduces delay which is unsuitable for real-time video transmission over wireless scenarios. If a frame is received after its playout deadline it will be discarded, thus leading to degradation of the video sequence. FEC mechanisms avoid such situations using redundancy to account for errors during transmission. In this way QoE is ensured without the need for retransmission. Nevertheless, FEC mechanisms can be non-adaptive, meaning that the amount of redundancy allocated is fixed and does not account for the variations in the MI characteristics of the video and in the network state. This is a greedy utilization of the already scarce wireless network resources ultimately leading to an aggravation of network traffic congestion due to a disorderly use of redundancy.

On the other hand, to surpass the aforementioned limitations, adaptive FEC mechanisms can be used. These perform Unequal Error Protection (UEP) by allocating different amounts of redundancy to different units of data according to a given rule in their design. Random Neural Networks (RNN) [2] and Ant Colony Optimization (ACO) [3] are dynamic and adaptive mechanisms which enable the use of UEP. Therefore, an adaptive video- and loss-aware mechanism based on RNN and ACO is proposed. The main goal is to improve overall QoE for the end user through the analysis of the video MI parameters and network loss state to adequately allocate an amount of redundancy which minimizes overhead without degrading video quality from the human point-of-view.

The Motion Vectors (MV) are extracted from a video sequence to infer about the MI. They are a crucial part of the compression process of the video, containing information related to nearby frames. By analysing the number of MVs and their Euclidean distance, it is possible to attribute a MI category to the video sequence, hereby allowing the FEC mechanism to use this information during redundancy allocation. The proposed mechanism uses a combination of MV,

[1]CNPq Fellow - Brazil

frame size, and frame type information to categorize motion intensity through a RNN. This is then combined with further information from network losses in a ACO scheme to produce a tailored amount of redundancy suited for each situation.

Neural Networks (NN) are information processing paradigms based on the way biological central nervous systems (e.g., the brain) process data. They can be configured for specific applications through a learning process, comprised of pattern reading and connection weight adjustment following determined rules to produce a desired output. Some common uses of such type of systems, specifically RNNs, are pattern recognition and data classification [4] [5].

The ACO system is a metaheuristic technique which is based on the behaviour of ants in the presence of pheromones. This technique is used to dynamically solve computational problems that can be expressed as path-finding in a graph. From a given starting point on the graph, ants traverse the paths between each node to reach a solution. During each trip, a pheromone marker is deposited. This aids the construction of a solution because the path which has the greater amount of deposited pheromone is a solution for the problem.

The AntMind mechanism processes information on a frame-by-frame basis to adaptively perform QoE-aware data protection, thus being a video-aware UEP scheme. A frame will have a larger amount of redundancy if it is categorized of higher importance and vice-versa. This helps to minimize the impact of overhead on the wireless network as well as preserving the quality of the video, ultimately improving QoE when compared to adaptive and non-adaptive FEC mechanisms.

The remainder of this paper is structured as follows. The related work is presented in Section II. Section III describes the AntMind and its evaluation is presented in Section IV. Conclusions and Future Work are summarized in Section V.

II. Related Work

There has been a growing interest in adaptive FEC mechanisms to improve video quality over wireless networks. This is due to their tailored redundancy allocation from the analysis of characteristics of the video/network without added delay.

Optimized Cross-Layer FEC (OCLFEC) is a Group of Pictures-based (GOP) mechanism which attributes a priority value to the GoP for transmission from the mean squared error of each frame [6]. It uses two error correcting codes, Luby Trasform (LT) and Rate-compatible Parity Check (RCPC). The data is first encoded with LT codes and then redundancy check bits are added to prevent coding errors. These codes are optimized with different parameters for specific situations. The only metric used to assess the performance of the mechanism is Peak Signal-to-Noise Ratio (PSNR) which by itself does not tell much information about QoE. This mechanism does not take into account the MI of the video. This is problematic because the mechanism may have a good Quality of Service (QoS) performance, but this does not mean that QoE is guaranteed. Also the several phases of the mechanism are time consuming which increases delay reducing the overall QoE.

"Transport Audiovisuel avec Protection Ingale des Objets et Contrle d'Admission" (TAPIOCA) is a non MI-based mechanism which divides each GoP by layers and attributes different priority values to them [7]. This enables the distribution of redundancy through the layers of most importance. To assess the performance of the mechanism two metrics were devised, specifically the Decodable Frame Rate (DFR) and the Protection System Efficiency (PSE). These are so unique that they do not provide much information about the QoE performance of the scheme. The process of dividing each GoP into layers is both computationally heavy and time-dependent, making this scheme unsuitable for real-time use.

Adaptive Packet and Block length FEC (APB-FEC) is a non MI-based mechanism which uses packet lengths that are smaller than usual in order to increase the size of the FEC block [8]. It uses packet loss rate feedback information to adapt the video to the network loss characteristics. The metrics used to assess its performance are the network overhead, the effective packet loss rate, and PSNR. However, these do not provide adequate information to evaluate the QoE.

The Video Aware (viewFEC) mechanism is a module-based scheme which analyses MI [9]. It uses an off-line database of several video sequences with MI and complexity information. A heuristic comparison is performed between the GoP and the database to assess the MI of the video being transmitted. The metrics used to assess the performance of the mechanism are suited for a QoE evaluation, but relying too heavily on the database may not provide optimal results when using videos different than those present in it. Also, this mechanism does not rely on packet loss rate information to adapt itself to the current network state while keeping the video QoE high.

The neuralFEC mechanism is a MI based mechanism [10] which performs the categorization of each frame of a video sequence according to its characteristics. It employs a RNN to use the video characteristics to compute an adequate redundancy value in real-time. The metric used to evaluate the performance of the mechanism is Structural Similarity (SSIM). It does not use any type of network loss information, which can introduce unnecessary amounts of added redundancy.

The AntMind mechanism takes into account all the above limitations and improves on them to create a novel QoE-aware adaptive system which provides adequate protection in video distribution over wireless networks.

III. AntMind mechanism

Using the AntMind Mechanism, the aforementioned limitations can be surpassed by taking into account the MI characteristics and the network loss state. Precious wireless network resources are thus spared and the impact of losses is reduced, leading to an overall QoE improvement. Figure 1 shows a general overview of the AntMind mechanism.

Firstly, the off-line training and validation of the RNN created for MI categorization purposes is performed. An exploratory hierarchical cluster analysis using Ward's method for clustering [11] assessed video characteristics and grouped video sequences according to MI. It formed clusters of similar

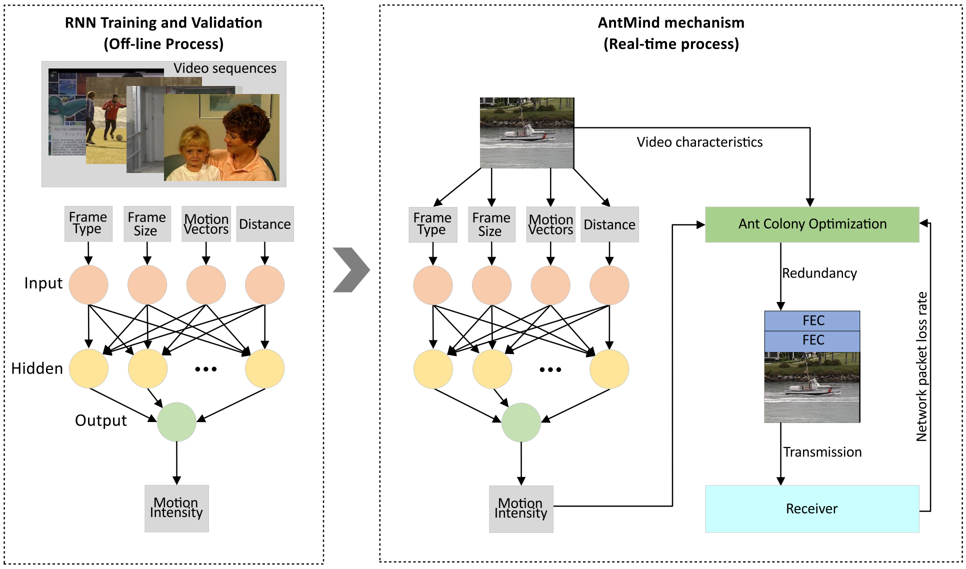

Fig. 1: AntMind mechanism

MIs (Figure 2) from a high intensity of motion to a low intensity. Video characteristics such as frame size, frame type and motion vectors were used during the analysis. From this result, a training set representing video sequences of low motion intensity (Akiyo and Silent), medium motion intensity (City and Football) and high motion intensity (Mobile and Flower) was selected. The RNN was trained using this set so that a different value can be attributed to distinct motion scenarios. After completing its training, the RNN is ready to be used to categorize the videos off-line (i.e., previous to transmission).

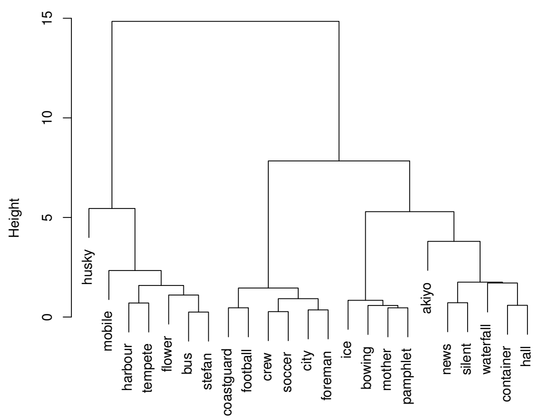

Fig. 2: Hierarchical Clustering of video sequences according to motion intensity

The RNN comprises four input nodes, seven hidden nodes and one output node. The four input nodes represent the video frame's parameters such as frame size, frame type, number of motion vectors and the distance described by those vectors. The single output node represents the computed MI value. Two MV parameters were used due to distinct situations: when a frame has a high number of MVs which have no distance value; when there are fewer MVs but these have higher distance values describing a scenario of greater movement.

Afterwards, the real-time process begins. The network loss information is retrieved from feedback packets sent from the receiver. These contain information representing the rate of erroneous packets in the previously sent frame.

After the identification of the MI characteristics by the RNN and the retrieval of network loss information, the ACO mechanism starts its operation. Based on the RNN assessment and the known characteristics of the video/network, the value representing the amount of redundancy to be used is computed in real-time for each frame. This is done in a way which minimizes wireless network overhead and maximizes QoE.

The graph used for this study, shown by Figure 3, consists of fourteen nodes depicting video and network characteristics. The first node serves as a starting point. Then, the next three nodes represent the classification given by the RNN in terms of MI, namely low, medium, and high. This classification was obtained from the hierarchical cluster analysis' results shown in Figure 2. The following two nodes represent the frame type, I- or P-. The next three nodes depict the frame size as small, medium, and large. The last five nodes represent the network loss rate for the previously transmitted frame.

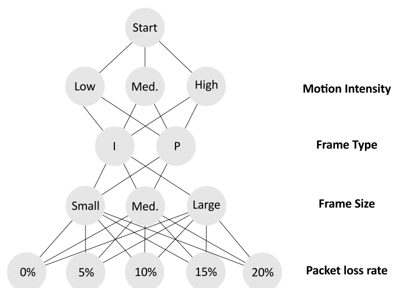

Fig. 3: Graph used in the ACO mechanism

Through the deposition of pheromone in the paths corresponding to the current situation, a preliminary path is generated for the ants. Both the number of iterations and ants is 10 and the simple ACO model was used. The amount of ants and preliminary pheromone was selected through extensive experimentation in order to obtain a plausible solution which did not worsen the processing delay of the AntMind mechanism. During runtime, the ants search the graph while leaving pheromone in the travelled path, resulting in a possible solution for the problem. The results are stored and can be re-used for similar conditions. This avoids re-running ACO again, thus allowing the use of ACO in real-time.

The value computed by the ACO mechanism represents the parameter which will be used in the Reed-Solomon (RS) algorithm [12]. This allows the addition of a specific amount of redundancy according to the video/network characteristics.

RS is of low complexity, resulting in low computational overhead making it suitable for real-time processing of streams of video. The RS code belongs to the family of linear block codes. It converts a source of s bits of an arbitrary length k , into a sequence of n bits of length. In order to protect the data, extra $(n-k)$ bits are added to the original data set. These extra bits are commonly known as parity-check bits.

Through this streamlined process, a tailored amount of protection is allocated to each frame, thus better protecting the most error-sensitive data. The overall QoE and network performance are improved from the minimization of overhead which also reduces the probability for congestion.

Figure 4 shows the AntMind mechanism pseudo-code for each frame. It first retrieves the frame type (02) and checks it against a condition to assess whether it is of type I- or P- (03), as only I- and P- frames will be protected. If a frame is of type B- it is immediately sent (13), otherwise, additional frame characteristics are retrieved, such as frame size (05), number of MVs (06) and the distance pointed by those MVs (07). Then, these characteristics are fed to the RNN for MI categorization (08). From the MI value, the information about the loss rate along with the video characteristics are fed into the ACO mechanism, which computes a value for the parameter of the RS code. This parameter governs the amount of added redundancy (09). Afterwards, RS encodes the data with the RSpar parameter (10) and the frame is sent (11). The packet loss rate is retrieved from feedback information sent from the receiver and the value is updated (15).

```
01 for each Frame
02  FT=getFrameType(Frame)
03  if(FT equal (I- or P-Frame))
04   then
05    FS=getFrameSize(Frame)
06    MV=getMotionVectors(Frame)
07    MVDist=computeDistance(MV)
08    MotionIntensity=RNN(FT, FS, MV, MVDist)
09    RSpar=ACO(MotionIntensity, FT, FS, LR)
10    addRedundancy(RS(RSpar))
11    sendFrame(Frame+Redundancy)
12   else
13    sendFrame(Frame)
14  end if
15 LR=getLossRate(ReceiverFeedback)
16 end for
```

Fig. 4: AntMind pseudo-code

From this accurate classification of the MI characteristics and network loss rate, the adaptive AntMind mechanism can minimize overhead. It does so while preserving the quality of the video. Therefore, it reduces the footprint of video transmission on the network while improving overall QoE.

IV. Performance Evaluation and Results

The main goal of AntMind is to improve the use of wireless network resources through the reduction of the network overhead derived from the allocation of redundancy by FEC mechanisms while assuring QoE. The performed evaluation aims at showing that the AntMind mechanism can effectively reduce network overhead while preserving video quality. A wireless mesh network scenario was devised, consisting of 25 nodes disposed on a 5 by 5 grid with a distance of 50 meters between them. The Network-Simulator 3 (NS-3) [13] was used to perform the evaluation of the proposed mechanism. The Optimized Link State Routing Protocol (OLSR) [14] was used as the routing protocol. The data set consisted of ten video sequences with different characteristics. These videos represent content commonly found in on-line video services. The video sequences are coded with the H.264 video codec and are in the Common Intermediate Format (CIF). Furthermore the sequences have a GoP of 19 with two B-frames after each I- and P- frames.

With the objective to describe the burst loss patterns of a wireless network [15], a two-state discrete-time Markov chain model following the Gilbert-Elliot packet-loss model [16] was implemented. The simplified Gilbert-Elliot packet-loss model consists of two nodes representing G, a Good state where no packets are lost, and B, a Bad state where all the packets are lost. These two states are connected by the probability of transitioning from G to B P_{GB} and the probability of transitioning from B to G P_{BG}. By adjusting these probabilities we can obtain different patterns of error which represent specific Packet Loss Rate (PLR) values (5%, 10%, 15%, and 20%) commonly present in wireless networks.

In order to thoroughly assess the results provided by AntMind, four distinct data protection schemes were used, namely one without protection, one with non-adaptive protection and two with adaptive protection. The first scheme is used to obtain a baseline value, to further assess the improvements of the FEC mechanisms. The second and third schemes are Video-aware Equal Error Proctecion (VaEEP) and Video-aware UEP (VaUEP) respectively. In the VaEEP mechanism a fixed amount of redundancy is added I- and P- frames. In the VaUEP mechanism, the same frames are protected with different fixed amounts of redundancy depending on the frame type. The protection of only I- and P- frames is a common practice in video transmission. The redundancy amounts were attained after a thorough set of simulation studies, where the best possible trade-off regarding the amount of redundancy and QoE in terms of video quality was found. This translates into a value of 38% average added redundancy for VaEEP and of 30% for VaUEP. The proposed mechanism (AntMind) is the last scheme used consisting in a combination of an RNN and ACO for Unequal Error Protection.

Two objective QoE metrics were employed [17], namely the Structural Similarity (SSIM) Index and the Video Quality Metric (VQM). The SSIM analyses luminance, contrast and structural similarity of images to assess the similarity/likeness of the pictures. VQM is a Discrete Cosine Transform (DCT)-based metric which uses the spatial-temporal property of the human visual system's perception to assess the distortion of the pictures. In SSIM, higher values represent better video quality, while in VQM lower values represent higher video quality. To enable the objective evaluation of video quality two tools were used, namely EvalVid [18] and the MSU Video Quality Measurement Tool (VQMT) [19].

The selected sequences are different from those used to train the RNN in order to guarantee that the mechanism will function with any other random video sequence, thus ensuring the validity of the classification performed by it.

Table I shows the results in terms of the average added overhead of all PLRs using the three FEC schemes. VaEEP's overhead ranged from 35% to 43% averaging at 38%, and VaUEP's added overhead ranged from 25% to 36% averaging at 30%. AntMind's overhead ranged from 9% to 19% averaging at 15%. This means that in some cases the reduction of added overhead reached 77% over VaEEP and 72% over VaUEP. Therefore, far less redundancy data is used by the AntMind opposed to VaEEP and VaUEP translating into reductions of 61% and 50%, respectively.

From the results, it is shown that the AntMind mechanism assesses the importance of frames according to their MI characteristics. Therefore, attributing a less redundancy to videos with a lower MI and a more redundancy to videos with a higher MI. This is clearly observable for Bowing and Mother which are categorized as low MI video sequences versus Coastguard, Harbour and Container which are categorized as high MI video sequences. Additionally, in the Mother and Soccer video sequences, the VaEEP mechanism allocates a higher amount of redundancy to Mother (41%) than to Soccer (36%). In Mother, the proportional size of I- and P-frames is higher than that of the B- frames due to its low MI. In Soccer, this is not valid because it has a higher MI. Therefore, our mechanism allocated a higher amount of redundancy to Soccer (16%) in contrast to Mother (9%).

TABLE I: Average Overhead

	VaEEP	VaUEP	AntMind
Bowing	42%	36%	10%
Coastguard	37%	27%	19%
Mother	41%	33%	9%
Soccer	36%	27%	16%
Hall	35%	27%	13%
Container	43%	35%	18%
Crew	35%	25%	16%
Foreman	37%	28%	15%
News	43%	35%	17%
Harbour	36%	26%	18%
Overall	38%	30%	15%

The SSIM and VQM metrics were used to perform the evaluation of the AntMind mechanism in order to complement the results obtained for the added overhead and to ensure that video quality was preserved. The different observable values are due to each video sequence's uniqueness, highlighting the importance of performing the experiments with video sequences with a broad array of MI levels/characteristics.

Figure 5 shows the values obtained in terms of SSIM for each of the schemes and for all of the video sequences. The average SSIM value was computed from all the video sequences for all PLRs. The mechanism without FEC averaged a value of 0,805. The VaEEP and VaUEP mechanisms obtained values of 0,881 and 0,882, respectively, which are closely matched by a value of 0,876 obtained for the adaptive AntMind mechanism.

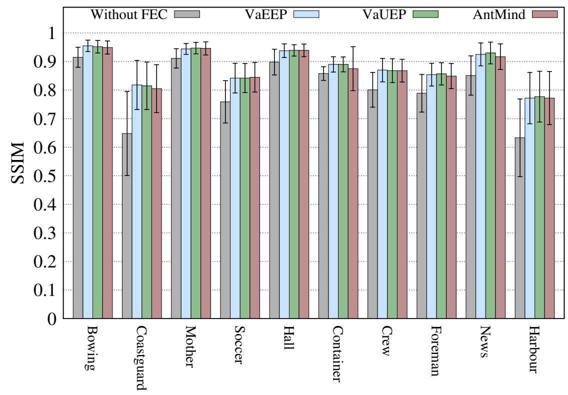

Fig. 5: Objective QoE assessment (SSIM)

Analysing in further detail the Coastguard video sequence in Table II, for a higher PLR the SSIM scores are considerably worse than for a lower PLR. The Mother video sequence does not show this kind of behaviour between distinct PLRs. Such reduction of quality is of course expected due to the increased PLR. Nevertheless, a change of such magnitude as observed for Coastguard and Harbour means that packet loss has a greater impact on video sequences which have a higher intensity of motion, contrary to Mother and Bowing. This shows that video sequences with a lower intensity of motion have a greater resilience towards loss.

TABLE II: SSIM variation through the PLR

	Without FEC		AntMind	
	Coastguard	**Mother**	**Coastguard**	**Mother**
PLR 20%	0,482	0,868	0,688	0,915
PLR 15%	0,574	0,899	0,784	0,938
PLR 10%	0,678	0,923	0,841	0,959
PLR 5%	0,859	0,955	0,906	0,971

Figure 6 shows the VQM results. The VaEEP and VaUEP mechanisms achieved an average VQM of 3,895 and 3,860, respectively, while the adaptive AntMind mechanism obtained an average of 3,940, meaning that video quality was maintained. These results follow a trend similar to that of SSIM. This helps corroborate the improvements of the proposed mechanism in overhead reduction with video quality shielding.

Table III summarizes the SSIM, VQM, and network overhead results. It shows that the proposed adaptive mechanism managed to achieve a considerable reduction in network

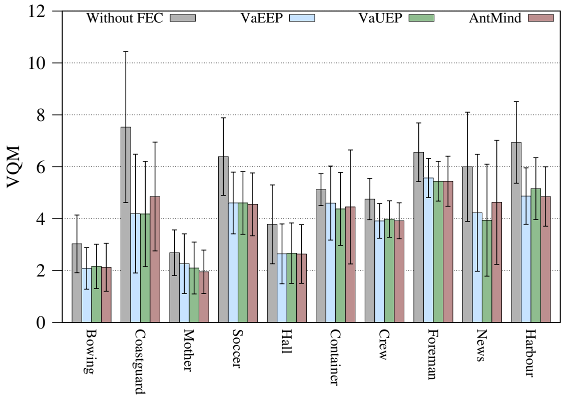

Fig. 6: Objective QoE assessment (VQM)

overhead by avoiding the addition of unnecessary redundancy. Furthermore, it did so while preserving video quality. This is of great importance in wireless network environments due to their limited network resources.

TABLE III: Average SSIM, VQM and network overhead

	AntMind	VaEEP	VaUEP	Without FEC
SSIM	0,876	0,881	0,882	0,806
VQM	3,940	3,895	3,860	5,278
Overhead	14,898%	38,460%	29.827%	–

Overall, the results showed that the AntMind mechanism can considerably reduce network overhead. It does so by performing an accurate classification of the MI components of the video. This information is used in conjunction with network loss state information to dynamically allocate redundancy on a frame-by-frame basis. By doing so, it ensures adequate protection of video sequences with different MIs while preserving video quality.

V. Conclusion and Future Works

The growing interest in video streaming on the go increases the stress upon the existing wireless networks. To make an efficient use of the resources while ensuring video quality, adaptive FEC mechanisms must be used. The AntMind mechanism performs a dynamic protection of the data that impacts QoE the most. Video quality is preserved while greatly reducing the use of network resources. The footprint and benefits of the AntMind scheme were demonstrated through a set of experiments using real video sequences.

The simulation results evidenced interesting behaviours from some video sequences, showing that those with a higher amount of MI require more protection. The AntMind mechanism uses MI and network loss to adequately shield a video stream with changing video characteristics in a lossy wireless network. It reduced network overhead by 61% on average, showing an improvement over non-adaptive and adaptive FEC mechanisms. Future improvements should further enhance of the mechanism considering other video sequences and the position of frames inside the GoP. Mobility and cross-talk scenarios should be explored to assess the performance of the mechanism under such conditions.

Acknowledgment

This work was funded by the Brazilian National Counsel of Technological and Scientific Development (CNPq) and also supported by the Intelligent Computing in the Internet of Services (iCIS) project (CENTRO-07-ST24-FEDER-002003), co-financed by QREN, Mais Centro Program.

References

[1] comScore, “2013 europe digital future in focus,” comScore, WhitePaper, March 2013.
[2] A. Abraham, “Artificial neural networks,” *handbook of measuring system design*, 2005.
[3] M. Dorigo, V. Maniezzo, and A. Colorni, “Ant system: optimization by a colony of cooperating agents,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 26, no. 1, pp. 29–41, 1996.
[4] S. Mohamed and G. Rubino, “A study of real-time packet video quality using random neural networks,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, no. 12, pp. 1071–1083, 2002.
[5] E. Aguiar, A. Riker, A. Abelém, E. Cerqueira, and M. Mu, “Video quality estimator for wireless mesh networks,” in *Quality of Service (IWQoS), 2012 IEEE 20th International Workshop on*. IEEE, 2012, pp. 1–9.
[6] A. Talari, S. Kumar, N. Rahnavard, S. Paluri, and J. Matyjas, “Optimized cross-layer forward error correction coding for h.264 avc video transmission over wireless channels,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–13, 2013.
[7] V. Lecuire, “Unequal error protection under bitrate constraint for video streaming over internet,” *Computer Communications*, vol. 35, no. 3, pp. 287 – 297, 2012.
[8] M.-F. Tsai, N. Chilamkurti, and C.-K. Shieh, “An adaptive packet and block length forward error correction for video streaming over wireless networks,” *Wireless Personal Communications*, vol. 56, no. 3, pp. 435–446, 2011.
[9] R. Immich, E. Cerqueira, and M. Curado, “Cross-layer fec-based mechanism for packet loss resilient video transmission,” in *Data Traffic Monitoring and Analysis*, ser. Lecture Notes in Computer Science, E. Biersack, C. Callegari, and M. Matijasevic, Eds. Springer Berlin Heidelberg, 2013, vol. 7754, pp. 320–336.
[10] R. Immich, P. Borges, E. Cerqueira, and M. Curado, “Adaptive motion-aware fec-based mechanism to ensure video transmission,” in *19th IEEE Symposium on Computers and Communications*, 2014.
[11] J. H. Ward Jr, “Hierarchical grouping to optimize an objective function,” *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.
[12] I. S. Reed and G. Solomon, “Polynomial codes over certain finite fields,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 300–304, June 1960.
[13] G. F. Riley and T. R. Henderson, “The ns-3 network simulator modeling and tools for network simulation,” in *Modeling and Tools for Network Simulation*, K. Wehrle, M. Güneş, and J. Gross, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, ch. 2, pp. 15–34.
[14] T. Clausen, P. Jacquet, C. Adjih, A. Laouiti, P. Minet, P. Muhlethaler, A. Qayyum, L. Viennot *et al.*, “Optimized link state routing protocol (olsr),” 2003.
[15] L. Wilhelmsson and L. B. Milstein, “On the effect of imperfect interleaving for the gilbert-elliott channel,” *Communications, IEEE Transactions on*, vol. 47, no. 5, pp. 681–688, 1999.
[16] R. Razavi, M. Fleury, and M. Ghanbari, “Adaptive packet-level interleaved fec for wireless priority-encoded video streaming,” *Adv. MultiMedia*, vol. 2009, pp. 3:1–3:14, Jan. 2009.
[17] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, “Objective video quality assessment methods: A classification, review, and performance comparison,” *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 165–182, 2011.
[18] J. Klaue, B. Rathke, and A. Wolisz, “Evalvid - a framework for video transmission and quality evaluation,” in *In Proc. of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, 2003, pp. 255–272.
[19] D. Vatolin, A. Moskvin, O. Petrov, S. Putilin, S. Grishin, and A. Marat. (2011) Video quality measurement tool. Moscow State University.