VIDEO ADAPTATION OVER HETEROGENEOUS NETWORKS

Ph.D Thesis

MARÍA ALEJANDRA PIMENTEL NIÑO

ADVISOR: DR. M. A. VÁZQUEZ-CASTRO

UAB
Universitat Autònoma
de Barcelona

PH.D PROGRAMME IN TELECOMMUNICATIONS AND SYSTEMS ENGINEERING
Department of Telecommunications and Systems Engineering
Universitat Autónoma de Barcelona
2014, July

PHD THESIS: VIDEO ADAPTATION OVER HETEROGENEOUS NETWORKS

PHD CANDIDATE: MARÍA ALEJANDRA PIMENTEL NIÑO

ADVISOR: DR. MARÍA ÁNGELES VÁZQUEZ-CASTRO

PH.D PROGRAMME IN TELECOMMUNICATIONS AND SYSTEMS ENGINEERING
Department of Telecommunications and Systems Engineering
Universitat Autónoma de Barcelona
Bellaterra, July 25th, 2014

ABSTRACT

Video services have become part of everyday interactions and contribute to a major portion of network traffic. Their broader usage includes out-of-the-ordinary scenarios as aid in emergencies, or telemedicine. Moreover, user demands of such services in terms of overall user experience continue to increase, leading to more specialized Quality of Experience (QoE) requirements. Guaranteeing a level of satisfaction to the user in challenging scenarios where the alternative networks are heterogeneous in nature continues to be an open issue.

The main objectives of this thesis have been to: 1) propose a framework for heterogeneous networking that allows for a seamless delivery of video content along diverse heterogeneous networks, 2) propose a user-centric framework for video transmission in line with heterogeneous networking, and 3) design a complete model and solution to provide video adaptation in heterogeneous networks such that it meets the requirements for user satisfaction.

The contributions of this thesis, such that the three objectives are met are as follows.

First, we propose to model heterogeneous networks with a holistic approach. The methodology of this holistic system design is based on two novel concepts. On one hand, to provide a framework by which heterogeneous network instances can be modeled to guarantee generality and robustness. On the other hand, to uniquely characterize the network instances via their min-cut. The strength of this framework is its usage as an underlying system model that can guarantee seamless content delivery regardless of the network instance. The latter is possible by formulating a general cross-layer optimization for content delivery, coherent to information-centric networking philosophy.

Second, we propose a QoE-driven adaptive video framework, based on a cross-layer optimization formulation. The derived adaptive video algorithm for time-variant networks is delay-driven, hence contemplates the constraints of long-delayed networks and the challenges of establishing a feedback loop to enable the network adaptability. The framework is evaluated systematically, in both an emulation and a fully implemented experimental environment.

Third, we propose the main contribution of this thesis: a complete model to provide user-centric video services in heterogeneous networks. The problem of combined erasures and congestion in best effort network is decoupled to match the specific degrading effects on the video. This allows for two separate QoE driven optimization approaches in time (freezes) and space (artifacts) domain. The complete solution offers a feasible dynamic streaming adaptation that suits constraint heterogeneous networks such as satellite. The performance is evaluated through a novel QoE three-dimensional analysis. The overall solution contemplates a novel semantical dimension, in line with information-centric networking, with an unexplored take on semantics that intends to reflect on the perceptual needs of the end user. We prove the strength of our design for the situation awareness scenario, where heterogeneous networks are often used, and show substantial gain in terms of mitigation of the effects of congestion and erasures while improving QoE and achieving the expected user's perceptual demands.

ACKNOWLEDGEMENTS

The path leading to the presentation of this dissertation has been challenging, frightening at times, but ultimately an invaluable lesson. I would like to acknowledge the persons that in one way or another have helped me in accomplishing this major milestone.

I would like to express my gratitude to my thesis advisor, María Ángeles Vázquez-Castro, for her support and expert guidance at every stage of the Ph.D process. The experience of being supervised by her and her research expertise have been immensely valuable. Her tireless dedication and efforts have been crucial in successfully reaching this final phase.

Throughout the course of my Ph.D studies, I was very fortunate to work for AnsuR Technologies. I am grateful to Harald Skinnemoen for the opportunity given to address practical problems in challenging communications scenarios. Many thanks to my former colleagues at Ansur, and in particular to Kristoffer Furuheim for all the fruitful technical discussions regarding the practical side of implementing video communications.

During the first two years as a Ph.D student I was given the great opportunity of working in the GEO-PICTURES project, which inspired some of the study cases in this thesis. I am thankful to the partners of the consortium for their collaborative work and enriching multidisciplinary insights.

I am glad to have spent most part of this journey sharing odd-working hours, coffees and discussions with my research group colleagues at UAB, Smrati, Paresh and Ricard. Thank you Smrati and Paresh for always having the time to discuss technical and non-technical problems. Wish you all the best in your future endeavors!

Along the road, I was lucky to have dear friends who, some from the distance, never stopped cheering, comforting, and ultimately just being there. To all of them I am indebted. A special thanks to Lilian and Octavio: we have come a long way, our “support group” chat made a big difference.

Living away from my home country has prevented me from spending much time with my family, but the words of encouragement and love from my big family (grandmothers, uncles, aunts, cousins) have never stopped arriving. I always hold you dear and thank you for keeping me in your thoughts. To my sister, Camila, you are my biggest cheerleader!

Most importantly, to my parents, who have been there every step of the way, with infinite patience and kindness, and always with wise words of advice. Mom, Dad, you have given me the strength and all the support during the most demanding moments. I cannot thank you enough.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

APP	APPlication layer
ARQ	Automatic Repeat Request
BEATLES	BGAN Enhanced Alphasat Technology for L-band Extended Spectrum
BGAN	Broadband Global Area Network
CIF	Common Intermediate Format (352 x 288 pixels)
DASH	Dynamic Adaptive Streaming over HTTP
DVB	Digital Video Broadcasting
DVB-RCS	DVB-Return Channel via Satellite
DVB-S2	DVB-Satellite 2
ETSI	European Telecommunicaitons Standards Institute
FEC	Forward Error Correction
GEO	Geostationary Earth Orbit
GEO-PICTURES	GMES-and Earth Observation with Combined Position based Image and sensor Communications for Universal Rescue, Emergency and Surveillance
GMES	Global Monitoring for Environment and Security
GoP	Group of Pictures
HEVC	High Efficiency Video Coding
HTTP	HyperText Transfer Protocol
ICN	Information-Centric Networking
IHP	International Humanitarian Partnership
IoT	Internet of Things
ITU	International Telecommunications Union
LEO	Low Earth Orbit
LTE	Long Term Evolution
MASERATTI	Mejora de la Atención Sanitaria en Entornos Rurales mediante Aplicaciones de Telemedicina sobre Tecnologías Inalámbricas
MDS	Maximum Distance Separable
MOS	Mean Opinion Score
MSS	Mobile Satelllite Services
NC	Network Coding
NCC	Network Control Center
NUM	Network Utility Maximization
ONC	Overlapping Network Coding
PEP	Performance Enhancement Proxy
PSNR	Peak Signal to Noise Ratio
QCIF	Quarter Common Intermediate Format (176 x 144 pixels)
QoE	Quality of Experience
QoS	Quality of Service
RCST	Return Channel via Satellite Terminal
RNC	Random (linear) Network Coding
RTCP	Real Time Control Protocol
RTCWEB	Real time Communications for the Web
RTP	Real Time Protocol
SRNC	Systematic RNC
SSIM	Structured Similarity Index Metric
TCP	Transmission Control Protocol

TFRC	TCP Friendly Rate Control
UDP	User Datagram Protocol
UEP	Unequal Error Protection
UMTS	Universal Mobile Telecommunications System
UN	United Nations
UTRAN	UMTS Terrestrial Access Network
VoIP	Voice over IP
VSAT	Very Small Aperture Terminals

1 INTRODUCTION

1.1 MOTIVATION

Current and future trends in network convergence create challenges of interoperability as well as integration obstacles among different kinds of networks. Such hybrid network configurations are diverse and in need of a model characterization. Circumstances are particularly challenging for solutions that combine terrestrial and satellite networks for specialized uses, such as emergency operations or telemedicine in remote areas.

The convergence towards an IP-waistband in networking, with the "everything over IP/IP over everything" paradigm has helped integrate next generation networks under the IP protocol stack premise. However, given the diversity in the performance and configuration of hybrid networks, and the imminent shift to an information-centric Internet of Things (IoT), a gap remains as to how these network instances are to be modeled.

These heterogeneous networks may offer limited or no Quality of Service (QoS) guarantees, may suffer from failures, be bandwidth limited, and shared by many users. Meanwhile, video services have become part of everyday interactions and contribute to a major portion of network traffic. The afore-mentioned characteristics of heterogeneous networks influence the performance of video streaming services, damaging the Quality of Experience (QoE) of the end-user.

Optimization of video services is primarily tackled from two perspectives. On one hand, at the application layer of the protocol stack, from an image processing perspective, with efficient, error-resilient video (source) coding techniques (HEVC, VP8) and intelligent stream generation and encapsulation. On the other hand, at lower layers of the protocol stack, from the communications and networking perspective. The latter includes cross-layer alternatives, relevant for wireless access networks, to exploit both perspectives for a higher performance.

While the literature is broad, solutions are often solving partially the challenges faced when the aim is to guarantee user satisfaction when delivering video in heterogeneous networks. Some are QoS oriented rather than user-centric, others lack a global understanding of heterogeneous networking for a generalized optimization and adoption. Cross layer techniques in particular have been broadly adopted for all sorts of QoS resource management with success. Nevertheless, most of the approaches are network-specific. As interoperability with standards and co-habitance becomes an asset, offering closed specialized solutions may not be attractive. Two main trends, internet-oriented and supported by standardization tracks, can be currently distinguished in video streaming, differentiating real-time services [1], from server-client video on demand solutions [2].

User demands of video services have grown and will continue to do so in the near future as broadband penetration increases, leading to more specialized QoE requirements depending on the service/scenario. Video is becoming a commodity, penetrating a wide range of usage, proving to be more than entertainment. This is true and more palpable in urban areas with fixed or wireless broadband access reaching the tens of Mbps with e.g. 4th generation mobile technologies. This reachability/availability is pushing the users to demand higher levels of QoE, also in less common and more challenging scenarios and for a broader range of applications other than the purely recreational. Guaranteeing a level of satisfaction to the user in such scenarios is a difficult task that has been scarcely explored. Moreover, those scenarios are a typical example where heterogenous networks are used.

In this thesis, we target the case of video for challenging heterogeneous networking scenarios beyond purely recreational, driven by the user's QoE demand. The goal is to propose models and feasible solutions that tackle the aforementioned challenges that are inherent to these kind of scenarios. We aim for an approach that is novel, user-driven and in line with the evolving information-centric user-driven networking.

A practical illustrative application is the case of video for situation awareness provision. Typical scenarios are emergency operations, monitoring, among others. Video has been proven to aid in the cognitive processes that lead to the acquisition of situation awareness, but its relationship to QoE, and how limited QoS provision in a video transmission hinders the use of video at all, has not been assessed.

1.2 OBJECTIVES

Based on the challenges described above, the main objectives of this thesis are the following:

Objective 1: Propose a framework of heteregeneous networking.

This framework shall be a novel contribution to the state-of-the-art, coherent with upcoming information-centric networking and focus on video services. It shall be coherent with information-centric approaches, where heterogeneous networks can be characterized regardless of their diversity in access network, end-devices, or network architectures. Moreover, the framework shall enable a suitable basis for optimized video services.

Objective 2: Propose a user-centric performance framework for adaptive video transmission.

This framework shall be in line with information-centric networking, and be aware of the QoS/QoE interplay. It shall contribute to the state-of-the-art by providing practically feasible QoE driven solution, aware of the interplay of QoS and the constraints inherent to heterogeneous networks. This solution shall be designed such that fits into the heterogeneous networking framework of Objective (1).

Objective 3: Develop a complete model and solution for video adaptation over heterogeneous networks.

The solution should build upon our proposed frameworks from Objectives (1) and (2), to provide novel modeling and analysis of the QoS/QoE interplay and propose alternatives to enhance user's satisfaction for video services beyond recreational purposes, in coherence with information-centric-networking.

As study case, the situation awareness application is chosen for performance analysis for its relevance in some heterogeneous networks.

1.3 CONTEXT OF THIS DISSERTATION

Part of the work leading to the results of this thesis was carried out within international projects with very practical user-oriented approaches. This doctoral thesis has taken the ideas developed in these projects to an academic level, with the proper modeling and generalizations.

GEO-PICTURES (acronym for GMES and Earth Observation combined with Position based Image and sensor Communications Technology for Universal Rescue, Emergency and Surveillance) [3], part of the European Commission's Seventh Framework Programme (FP7), was a user driven project focused on integrating different real time sources of image communications in near realtime for the purpose of mitigating damages in humanitarian and environmental emergencies. It involved different partners, with experts in emergency and rescue operations from United Nations and the European Civil Protection, as well as a multidisciplinary research teams.

The project BEATLES (acronym for BGAN Enhanced Alphasat Technology for L-band Extended Spectrum)[4], partly funded by the European Space Agency (ESA), was part of ESA's call for applications to the recently launched Alphasat satellite. The objective of BEATLES was to offer improved user experience in image and video delivery over band limited Mobile Satellite Services in the L-band. The target users were disaster and emergency relief organizations in need of situation awareness. Inmarsat's satellite constellations and its BGAN network offered the Mobile Satellite infrastructure.

The project entitled "Mejora de la Atención Sanitaria en Entornos Rurales mediante Aplicaciones de Telemedicina sobre Tecnologías Inalámbricas" (MASERATTI) [5] brought academics and field practitioners together, to define suitable wireless technologies in rural areas, develop a number of telemedicine services for remote medical attention and study the feasibility of deploying such technologies in remote areas of Latin America.

1.4 STRUCTURE OF THE DISSERTATION

The structure of this dissertation is based on the three main objectives addressed, covering one chapter per objective. The contribution per chapter is as follows:

1. Chapter 1: The current chapter introduces this doctoral thesis. It presents the motivation for this work, its context and main objectives. Further it describes the overall structure of this thesis and lists the contributions related to the work presented.

2. Chapter 2: This chapter provides the context and preliminaries to the work and contributions of this dissertation. We describe the preliminaries on heteroge-

neous networks and the study case of networks for emergencies and telemedicine in remote areas (a contribution to [6]). We contextualize video adaptation and its various approaches, followed by preliminaries on information-centric networking. Finally, we provide preliminaries for QoE for video, with definitions, mappings and metrics. As a study case for user's QoE in video services beyond recreational we present the concept of situation awareness and its context for video.

3. Chapter 3: This chapter contributes to the achievement of Objective (1). It presents the main contribution towards this objective with a holistic approach to model heterogeneous networks. With this framework, generality and robustness are guaranteed regardless of the network instance. The chapter explains the methodology applied, how the network instances can be characterized via the min-cut, and how this framework enables a general cross-layer optimization for multimedia content delivery. To show its usefulness, several realistic instances are presented, all within the context of emergency and humanitarian response. Further, the results of field testing show the real challenges in the heterogeneity of the network instances. Finally, we present a case of system design for a hybrid network for environmental monitoring. The work from this chapter was presented in [7] as well as in several deliverable reports from the project GEO-PICTURES [3].

4. Chapter 4: This chapter contributes to the achievement of Objective (2) by presenting a user-centric framework for video transmissions with a QoE-driven cross-layer optimized framework. First it presents the theoretical formulation, with roots in the general cross-layer optimization from Chapter 3. Second, it shows how it can be implemented. Performance of the resulting algorithm is presented in both emulation and experimental phases. Further, the potential of combining the QoE-driven adaptive video together with content aware overlapping network coding for erasure protection is assessed. The work from this chapter was presented in [8, 9], and used as a starting point for [10]. Part of this work was implemented for [4].

5. Chapter 5: This chapter presents the main contribution of this thesis with the fulfillment of Objective (3). It presents a complete solution for adaptive video over heterogeneous networks with the main novelties of space/time decoupling of QoE/QoS interplay and the use of perceptual semantics. The chapter first defines the scenario and challenges. Second, it presents a time-space decoupling approach that allows for a complete solution with two independent optimizations, one tackling congestion (time), the other one erasures (space). Te overall solution contemplates the use of random linear network coding for erasures. The proposed mapping is based on perceptual features in QoE and system level influential factors. Experimental results of the time-space decoupling for congestion and erasures are presented using novel three-dimensional analysis. Further, we introduce perceptual semantics, and analyze the study case of the situation awareness scenario. The solution is mapped to future information-centric networks. The work presented in this chapter has been presented in [11, 12].

6. Chapter 6: This chapter draws the final conclusions and future lines of research.

1.5 LIST OF CONTRIBUTIONS

The work leading to this thesis has been presented in different scientific publications as well as in a number of project deliverable reports. Following are the listed contributions.

JOURNAL

1. M. A. Pimentel-Niño, P. Saxena and M. A. Vázquez-Castro. "Multimedia Delivery for Situation Awareness Provision over Satellite". Submitted to IEEE Transactions on Wireless Communications, April 2014.
2. I. Hernáez-Corres, M. A. Pimentel-Niño, M. A. Vázquez-Castro. "Perceptual Semantics over Adaptive Video". Submitted to IEEE Transactions on Wireless Communications, 2014.

CONFERENCE

1. M. A Pimentel-Niño, M. A. Vázquez-Castro, and I. Hernáez-Corres, "Perceptual Semantics for Video in Situation Awareness", accepted to ICSNC 2014, Nice France. October 2014
2. M. A. Pimentel-Niño, P. Saxena and M. A. Vázquez-Castro. "QoE-driven Adaptive Video with Overlapping Network Coding for Best Effort Erasure Satellite Links". 31st AIAA International Communications Satellite Systems Conference, Florence, Oct. 2013.
3. S. Gupta, M. A Pimentel-Niño, M. A. Vázquez-Castro, "Joint Network Coded cross layer optimized video streaming over relay satellite channel", 3rd International Conference on Wireless Communications and Mobile Computing (MIC-WCMC), Valencia, June 2013.
4. M. A Pimentel-Niño, M. A. Vázquez-Castro, and H. Skinnemoen, "Optimized ASMIRA – Advanced QoE Video Streaming for Mobile Satellite Communications Systems", 30th AIAA International Communications Satellite Systems Conference, 2012.
5. M. A Pimentel-Niño, M. A. Vázquez-Castro. Cross layer content delivery optimization for holistic network design in disaster preparedness and recovery scenarios. In: International Conference on Multimedia Computing and Systems/International Conference on Multimedia and Expo.; 2011. p. 1-6.

BOOK CHAPTER

1. Vázquez-Castro, M. A., Pimentel-Niño, M. A. and Alegre-Godoy, R. Las Redes de Comunicación basadas en Satélite. In: Tecnologías de la Información y las Comunicaciones para zonas rurales: Aplicación a la atención de salud en países en desarrollo. CYTED; 2011. ISBN: 978-84-15413-09-7.

CONTRIBUTIONS TO DELIVERABLE REPORTS FROM INTERNATIONAL PROJECTS

European Commission's Seventh Framework Programme (FP7). Project GEO-PICTURES (2010-2012) [3]

1. D3: 3.1 Technology Research, 3.2 Development and 3.3 Tests.
2. D2: 2.1 System Design and 2.2 Requirements.
3. D4: 4.1 Trials and 4.2 Evaluation.

European Space Agency's ARTES Programme. Project BEATLES (2012-2013) [4]

1. Technical specifications.
2. Pilot system architecture report.
3. Design justification file.

iberoamerican Programme of Science and Technology for Development CYTED. Thematic network MASERATTI: "Mejora de la Atención Sanitaria en Entornos Rurales mediante Aplicaciones de Telemedicina sobre tecnologías Inalámbricas"(2010-2012) [5]

1. Vázquez-Castro, M. A., Pimentel-Niño, M. A. and Alegre-Godoy, R. Las Redes de Comunicación basadas en Satélite. In: Tecnologías de la Información y las Comunicaciones para zonas rurales: Aplicación a la atención de salud en países en desarrollo. CYTED; 2011. ISBN: 978-84-15413-09-7.

2 PRELIMINARIES

This chapter introduces the preliminary aspects related to the state-of-the-art in video services for heterogenous networks.

Section 2.1 discusses the preliminaries on heterogeneous networks, Moreover, it presents practical aspects with a study case on emergencies and telemedicine scenarios. Preliminaries on video adaptation follow in Section 2.2, with and introduction to cross-layer optimization. The notions of information-centric networking are introduced in Section 2.3 as a paradigm shift affecting video service provision. Finally, the background on QoE for video is presented in Section2.4, with definitions, metrics and related literature in particular for video. Moreover the concept of situation awareness for video is introduced.

The link between how we meet the main objectives of the thesis and the state-of-the-art will be discussed at the beginning of Chapters 3, 4, and 5, in order to more easily highlight the novelties of the contributions of this thesis.

2.1 PRELIMINARIES ON HETEROGENEOUS NETWORKS

Here we present an introduction to heterogeneous networks, the challenges faced and examples in the literature of modeling alternatives.

Further, we will address a particular study case where the use of heterogeneous networks is common, highlight the challenges and solutions for these scenarios.

2.1.1 Modeling heterogeneous networks

Heterogeneous networks are interconnecting a diverse variety of components, through possibly diverse links. The nature of the heterogeneity can be determined by different classifications.

Heterogeneity can be due to the following aspects. It can be that different communications modalities are part of the network, i.e., terrestrial wireless, terrestrial mobile, satellite, fixed-line. It can reflect the possibility of different network topologies. It can be due to different access networks, core architectures or new and legacy technologies trying to coexist. Heterogeneity can also manifest in the different protocols used at each of the layers of the protocol stack suite.

Interoperability of different network architectures and the demand for ubiquitous access to the internet has contributed to the convergence towards a protocol stack where the network layer protocol is the common base. The IP waistband paradigm has been very successful in accomplishing this, but as the trend towards an ubiquitous internetworking has lead to the possibility of diverse devices connecting to the internet, leading to the Internet of Things, challenges in terms of interoperability still persist.

Difficulties of heterogeneous networks studied in the literature are, among others, on-they-fly vertical handover, QoS management, interoperability, performance, security, and interference management [13].

An example of heterogeneous networks occurs within the 4th generation of mobile terrestrial standardization and beyond [14],where different cellular topologies are exploited to simultaneously coexist and increase overall spectral efficiency.

2.1.2 Identification of practical aspects in heterogeneous networks: study case for emergencies and telemedicine purposes

We discuss in this section examples of realistic scenarios where heterogeneous networks are the norm. We highlight the challenges faced in terms of networking, and the common solutions used to provide connectivity.

2.1.2.1 *Networks in humanitarian and environmental disasters*

Timely and effective action taking when natural hazards events occur, is key in mitigating the potential devastating effects to the population as well as to the environment. A rapid, focused response to events is possible when accurate and reliable information about the situation is available to the decision makers.

CHALLENGES Network availability for in-situ communications is one of the challenges faced. Network infrastructure is often damaged during emergencies or unevenly deployed in the first place. Regular fixed line communications or any other terrestrial-based network (including mobile) may be partially damaged. Furthermore, even if the network is partially available, access to the network may be restricted due to congestion.

When trying to establish fast/reliable communications from the field to an operations control center, different kinds of network may exist, hence heterogeneity is another challenge. Each one of these access networks will have its own inherent channel constraints. Moreover, disasters are location-agnostic. This means not only that heterogeneity in types of networks is characteristic but that these networks may perform differently in different locations.

COMMON OPTIONS

Satellite Communications Satellite communications stand out as suitable network in emergency communications for its ubiquity capacity as well as for not being affected by any terrestrial hazard [15] The nature of the operation, will define the needs for a short/medium term deployment of a satellite system or an immediate establishment of communications means without any needed infrastructure in a nomadic type of operation [ETSI TR 102.41 REF]. Examples for the former case are the very small aperture terminals (VSAT) systems providing more bandwidth, while satellite phones, the Thuraya network or the BGAN network by Inmarsat [16] can make the latter case possible.

The advantages of satellite communications services can be summarized as [17]:

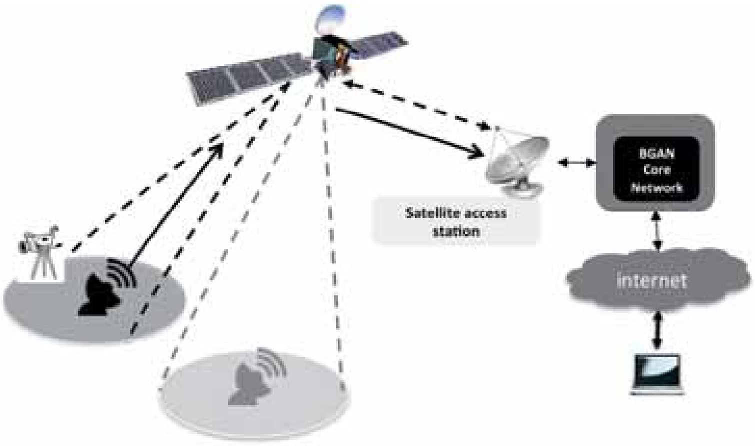


Figure 2.1: MSS network architecture.

- Global reach: It is a service not subject to the density of users. The area of coverage guarantees reaching a user regardless of its geographical location.

- Flexibility: easy installation with the possibility of using portable terminals such as VSAT. It requires only a small simple infrastructure. Due to this fact it is also an ideal solution to temporary services.

- It is a self-sufficient solution that does not depend on the availability and operability of the terrestrial networks.

- The aggregated cost of the intrinsic geography of the terrain and its location is non-existent, compared to terrestrial solutions.

- In emergency and disaster situations, telemedicine gets support from satellite networks to reach isolated and geographically difficult areas, without the need of a telecommunications infrastructure operating.

Mobile satellite services (MSS) offer very attractive advantages to the disaster response coordination community, with small portable terminals and no additional infrastructure. BGAN is an example of mobile or nomadic satellite services, operating in the L-band with global coverage except in the poles. Fig. 2.1 shows the network architecture of a MSS service like BGAN.

Two modes are offered to the user of the BGAN network: “streaming IP” or guaranteed service, and “standard IP”, or best effort. In emergency scenarios, the guaranteed service may not be available, as more users may opt for the services offered by the BGAN-like networks

Nevertheless, satellite networks may have the problems inherent to any wireless channel, namely erasures, and if many users are using the resources, the available bandwidth will reduce due to congestion. Furthermore, shadowing effects blocking the Line of Sight in the satellite link [18] during transmission as well as heavy rainfall at certain frequency bands may results in bursty erasures at packet level.

In addition, while VSAT networks may offer uplink rates in the order of megabits per second (Mbps), MSS services offer limited bandwidth, in the order of the hundreds of kilobits per second (kbps).

Mobile terrestrial networks Mobile terrestrial network should not be discarded as means of communication in emergency communications as they have by far overtaken the fixed line coverage, and reached areas, specially in developing countries, where they become the only available terrestrial network [1]. By 2017, 85% of the population will have third generation 3G mobile network coverage and 50% will have fourth generation mobile networks coverage, according to the report from the leading network technologies provider Ericsson [19].

Nevertheless, heterogeneity of the deployment of mobile terrestrial networks results in varying service provision across countries and regions, with prioritization of urban areas for the highest QoS provision, and limited access to broadband in sub-urban and rural scenarios. Further, as a shared medium, congestion at different times of day may occur.

PRACTICAL ASPECTS In practice, emergency response follows a number of protocols at regional and global level, aimed at organizing effective and rapid actions. Emergency response activities require a high degree of coordination and involve interactions among people with different background expertise.

Several organizations provide guidelines with the aim of standardizing this process and define the technical requirements for assessment missions, including the type of communications means available to first-responders. The International Humanitarian Partnership (IHP) [20] specifies the support modules available for different types of teams involved in first response, with a list of devices for communications means. Target users such as the United Nations (UN) at global level, and the European Community Mechanism for Civil Protection at regional level, follow these recommendations and define their own standard assessment kits [21]. The assessment mission kits (AMK) may include laptops, mobile phones, portable satellite terminals for broadband access to the BGAN network, and terminals for voice communications through satellite.

Middle to longer term response to emergencies may follow guidelines from international standardization bodies in the field of communications, like the ETSI, which outline possibilities of hybrid solutions for an emergency response network. It is based on temporary terrestrial communications (3G, Wifi) linked to permanent satellite networks.

2.1.2.2 *Networks for telemedicine purposes in rural areas*

We consider in this section the case of telemedicine as a medium to long term solution, aimed at providing a certain degree of medical assistance to isolated areas.

The case of medical assistance in emergency scenarios offers similar alternatives in terms of available networks as discussed in Section 2.1.2.1. Hence, it is worth discussing different possible network architectures suitable for telemedicine provision.

For a specialized discussion on the matter, [22] discusses at length, all aspects involved in the provision of telemedicine in rural areas. More details on satellite communications for telemedicine scenario are available in [6].

---

[1] see http://www.itu.int/osg/spu/ni/mobileovertakes/

CHALLENGES Telemedicine appears as an alternative to provide professional medical services to communities that lack in-situ medical professionals and are geographically difficult to reach.

The challenges to deploy telemedicine networks from the networking point of view are among others availability of network infrastructure, and geographical reachability. Typically there is no fixed network infrastructure and there might be limited to no mobile terrestrial connectivity. This motivates the use of alternative and ad-hoc solutions that combine heterogeneous networks to provide communications link. The geographical conditions, reachability and characteristics of the terrain also dictate the alternative means of communications that are feasible given the scenario.

COMMON OPTIONS Satellite communications and hybrid solutions are a feasible alternative for e-health in distant rural areas As with the case of humanitarian and environmental disasters, satellite communications can offer the flexibility and global reach needed to provide e-health to distant rural areas.

Network architectures that can be advantageous include the use of satellite technologies based on satellite communications standards for the return channel, like DVB-RCS, to provide broadband services to remote areas. The Guidelines for connectivity in rural areas from the International Telecommunications Union (ITU) [23], highlight the use of VSAT systems using standards as DVB-RCS together with DVB-S2.

Remote areas often challenge the deployment of communications systems due geographical peculiarities, accessibility among others. In particular, the concrete case of the Amazonian region has been studied in [24] due to the impact of heavy rainfall in the satellite link in the Ka frequency band. The use of DVB-RCS/S2 for Voice over IP (VoIP) services in remote areas was proposed in [25], with a hybrid configuration using WiMAX wireless links and satellite links. [26] presents an application of DVB-RCS to telemedicine.

Hybrid architectures such as the one proposed in [27][28](, (See Fig. 2.2), are a feasible option for telemedicine provision, that can cover wider extension using a combination of wireless terrestrial and satellite topology. Broadband access in isolated regions is guaranteed through a wireless local WiMAX network connected to a satellite DVB-S2/DVB-RCS that serves as gateway to the internet.

Alternatives that explore other types of wireless architectures in similar challenging settings are described in [22].

PRACTICAL INITIATIVES Following we present a summary of initiatives for the use of satellite technologies for telemedicine that have been studied and deployed in remote areas of developing countries and how they aid in providing medical services to isolated communities.

AFRICA Organizations such as the European Union, The European Space Agency (ESA) and the World Health Organization (WHO), have been actively involved in projects that support the use of satellite networks for health improvement.

[29] details the opportunities laying ahead for the use of satellite technologies in telemedicine services in the sub-Saharan region. Some the the pilot cases presented include projects in Mali, Uganda and Nigeria. Cases like Rwanda are worth highlight-

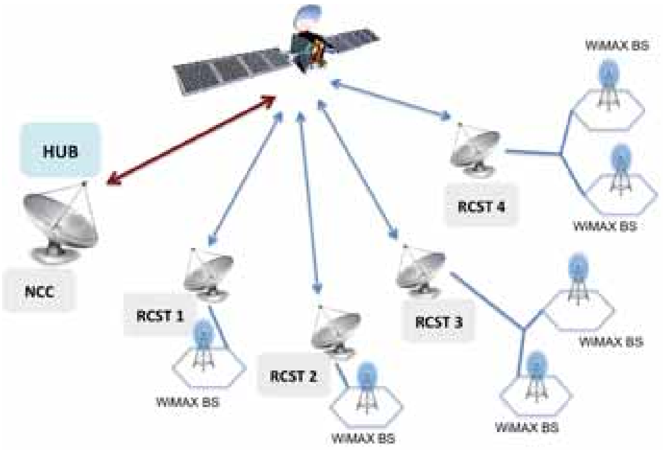


Figure 2.2: Hybrid network to provide broadband to isolated areas.

Figure 2.3: Architecture of the T@HIS network.

ing: a satellite network is vital due to the fact that Rwanda is not connected to the international fiber optics network.

ASIA The Indian Space Research Organization, ISRO, has launched pilot projects focused on the application of satellite technologies for health and education [30]. [31] presents the possibility of having a dedicated satellite to cover the needs of rural regions of Asia and Africa in order to provide telemedicine services.

LATIN AMERICA The AmerHis system [32], enables broadband provision to the regions covered by the Amazonas satellite operated by Hispasat. This system is totally compatible with the satellite standards DVB-RCS and DVB-S, therefore enabling the provision of IP services.

Fig. 2.3 shows the architecture proposed by the project T@HIS [33], focused on telemedicine service provision in rural areas with difficult access in Latin America. The provision of developed a medical networks via satellite in rural areas of the peruvian amazonian has been proposed in [34].

2.2 PRELIMINARIES ON VIDEO ADAPTATION AND CROSS-AYER OPTIMIZATION

Video applications are populating everyday network usage. Heterogeneous networks may not fully guarantee QoS, as they are product of converged networks of different kinds.

In particular, video is occupying most of the traffic on the internet, the classic best effort network. Video traffic, contrary to so-called elastic traffic, is highly affected by the delay and throughput in a network. The need to adapt the video traffic to the time variations of the network becomes crucial in order to guarantee QoE to the user.

Following we give an overview of the main trend in video adaptation, dealing in particular with erasures and congestion.

2.2.1 Adaptive Video for congestion/erasures

2.2.1.1 *Congestion*

APPLICATION LAYER Source adaptation techniques vary often from application to application and have evolved together with the advances on video coding algorithms. Source adaptation is performed in general terms either through variations in temporal (frame rate), spatial (resolution), or fidelity (quality) descriptors of the video. These can be performed for a single stream (i.e. H.264/AVC codec) or in multiple layers (i.e. Scalable Video Coding in H.264/AVC, or H.265). Multiple description coding is another approach to video coding available, where unlike SVC, results in streams that can be independently decoded, without the aid of other streams.

The choice of source adaptation schemes in video streaming has been linked to the application, delay-budget constraints and transport protocol used. In UDP streaming on-they-fly bit rate control adaptation is common for real-time applications, best effort networks [35]. In TCP scenarios, commonly chosen for non-real time such as Video on Demand (VoD) over broadband networks, stream switching is performed. These server-client solutions keep several versions of the video coded with different profiles, and use standards such as Dynamic Adaptive Streaming over HTTP, DASH, [2] to switch between pre-encoded streams as source adaptation, in a solution that meets have way between streaming and progressive downloading.

TRANSPORT LAYER Video streaming is growing in popularity in best effort internet with over the top services that use the most common transport protocol: TCP. Although TCP has an advantage in terms of practical deployment, it does not suit all scenarios. On the one hand, TCP has low performance for streaming in satcom scenarios: the long round trip times and packet erasures decrease the TCP throughput, leading to long startup delays in video playback. Moreover, even with long delays at the start-up (in the order of several seconds), the video source rate cannot be of more than half TCP throughput [36], which also decreases QoE. WIth regards to live streaming of user generated content, HTTP options rely on intermediate servers to prepare content for the clients e.g. in order to use DASH. This option again affects QoE with longer startup delays and freezes. Improvements to general purpose TCP in satellite communications, includes performance enhancement proxy (PEP) solutions, that alter the system architecture.

A variety of TCP-friendly congestion control schemes have been studied for video streaming [37], [38], [39], [40] mainly focusing on the fairness of the schemes rather than on the impact on QoE. Further, they rely on heavy feedback from the receiver, a problematic issue in long-delayed networks. Real-time applications, on the other hand, opt for Real Time Protocol (RTP) and use ad-hoc congestion control schemes [35],[41]. They offer the flexibility our scenario needs at transport layer, but rarely focus on QoE assessment nor on band limited long-delayed networks.

With respect to standardization, two distinct lines are traced, namely video conferencing applications are under the umbrella of RTCWEB for over-the-top video conferencing applications [1], mainly ver UDP, while video on demand applications are following DASH [2], mainly over TCP . Further, a a work group of the IETF is studying congestion control algorithms for interactive real-time multimedia [42].

2.2.1.2 *Erasures*

Wireless systems in general, and in particular satellite systems, also suffer from packet erasures due to poor wireless reception conditions and to the presence of channel fading in wireless links, which can not be coped with by adaptivity at the physical layer. This thesis does not focus on coding techniques for erasure protection, but it does study how online video adaptation can be integrated with forward erasure correction techniques in optimal ways to improve QoE. Hence we provide a short overview of the different approaches to increase reliability in video transmission.

Two paths can be differentiated, namely add robustness within the video codec's internal structure, or implement it externally, at packet level or below in the protocol stack.

Error concealment is part of state-of -art-codecs, with strategies that try to cope with loss of video frames. Nevertheless, error concealment overcomes short temporal error propagation but it may not be sufficient to handle a more severe one [43]. Therefore external robustness means should be used according to the type of application.

Furthermore, the goodness of such error concealment methods relies not only on the codec's concealment algorithms but on the level of integration of the codecs into the transport protocols, following payload guidelines dependent on each codec (as in e.g. [44]).

Existing erasure recovery strategies mainly include retransmissions or use of redundant packets. Retransmission based schemes (e.g. ARQ, TCP) may provide perfect recovery, however, the increase in delay and overhead due to per packet feedback may decrease the throughput, specially in case of wireless systems with when round trip times are high.

Forward error correction (FEC) solutions that add redundancy to the data to be delivered are either media-unaware, agnostic to the type of content, or media-aware. Cross layer strategies for video robustness:[45], or Unequal Error Protection (UEP) are part of the media aware strategies. [46] combines FEC with a feedback loop for acknowledgments for random losses and bursty losses in the network. Other approaches involve multi-path streaming and content-aware FEC.

Emerging topics in communications, such as Network Coding, have also been studied to add robustness in video streaming applications. Random Linear Network Coding (RNC) [47],[48] allows mixing of packets to obtain a fixed amount of redundancy such that enough protection is guaranteed without the need of feedback . The use of RNC for reliable communication in wireless networks was first studied in [49] where RNC is shown to be capacity-achieving for both unicast and multicast communication with packet erasures. [50],[51],[52] also discuss the use of RNC for reliability.

2.2.2 Cross-layer optimization

The traditional layered paradigm of the Open Systems Interconnection (OSI) protocol stack in communications networks provides simple layer-independent solutions to optimize the performance of the network. Each layer in the stack has specific tasks to fulfill, allowing for a modular, clean approach.

Nevertheless, in real-time applications and transmission of multimedia content, this layered approach results in suboptimal performance. Furthermore, in wireless scenarios the already limited resources are not able to adapt to the content requirements. All layers are working separately at trying to optimize their processes disregarding each other's mechanisms to i.e. adapt to bad channel conditions at lower layers or improve video/voice quality at higher layers.

An alternative to optimize the network performance is to exchange information among layers, and in that way lower layers can adapt to the multimedia content requirements while upper layers also adjust the content to the signaling from the lower layers. There are different ways to design cross layer solutions, with different levels of interactions among the layers in order to obtain an optimal strategy [45] that e.g. minimizes power consumption, maximizes the throughput, guarantees a certain level of QoS or QoE.

The cross layer queuing and scheduling architecture for a QoS aware transmission of layered video in [53] is an example of a full cross-layer design in satellite communications scenarios. The cross layer signaling in this design involves all the layers in the protocol stack. Other examples of cross-layer design in satellite and hybrid networks [54] for reliability, and [55] for VoIP.

2.3 PRELIMINARIES ON INFORMATION-CENTRIC NETWORKING

Networking approaches are shifting from the traditional host-centric towards information-centric architectures. The shift is driven by the user demands for content delivery, and their new role of both producers and consumers of information in the Internet environment. The aim of this approach is to integrate content delivery as a native network feature, where focus is not on the network as an enabler of communication links but as a platform for information dissemination.

Approaches in information-centric networking (ICN) propose fundamental changes in networking architectures where the addressing is linked to the information rather than to the host, the location of the information in the network is not relevant. [56, 57] offer surveys of a currently very active research topic. Different approaches in ICN are examined, some offering new, clean slate architectures completely independent of the current IP protocol stack networking, while others propose architectures that cohabit with the very successful TCP/IP to which practically all kinds of heterogeneous network architectures (regardless of their access medium) have converged to.

Some of the features of interest in ICN research (mostly from the Internet perspective as content distributor) are: optimization of content distribution by means of e.g. in-network caching, reduction of replicated data, decoupling of time (asynchronous) and space in content distribution (dissociation of provider/consumer as they do not need to be aware of each other).

ICN is also influenced by the advent of plethora of end devices that are becoming part of the internet landscape into what has been called the Internet of Things (IoT) [58].

With respect to video, while video-on-demand is a clear beneficiary in ICN, other services will coexist under this new paradigm. [59] ponders if ICN architectures are ready for other sorts of video services, and shows how the capabilities of different ICN proposed architectures can help, and what challenges remain for video services provision.

Information-centric networking is particularly relevant to applications that intend to deliver video content and cohabit in the future Internet of Things.

2.4 PRELIMINARIES ON QoE FOR VIDEO

Quality of Service (QoS) has been the standard measure of performance in a network. Its purpose tends towards representing the network provider's benefits for an optimal use of available resources rather than the end-user satisfaction with such services. This paradigm is now shifting towards a user-centric evaluation of quality, from which Quality of Experience (QoE) has emerged.

Efforts in characterizing what QoE means and how to model it have attracted the research community, towards a more global understanding of user experience and satisfaction with a service [60]. In particular for video, the complexity of this media has been a challenge to model. The traditional notions of Quality of Video (QoV) were in terms of fidelity metrics such as distortion [61], coming from image quality assessment.

2.4.1 QoE definitions and literature

QoE is a multidisciplinary field aiming to understand the degree of human satisfaction towards an application or service. General QoE models thus try to integrate psychology, technology, business, context aspects, among others, into a holistic view of QoE [62]. A thorough review of general purpose QoE models and QoE management for wireless networks can be found in [63].

2.4.1.1 *Definitions*

Definition 1. Quality of Experience QoE is the degree of delight or annoyance of the user of an application or service. [64].

Human, system and context are influential factors affecting QoE [64]:

- Human factors consider emotional, sociological backgrounds among others. Examples are user needs and expectations
- System influential factors relate to the technical aspects affecting quality of the application or service, including media capture, coding, transmission, playback among others.
- Context factors describe when the application is being used, with respect to temporal and location characteristics, as well as economic or social ones, describing how the application is being experienced. Examples include location related to the access network or time of day of usage

2.4.1.2 *Mappings of QoE vs QoS*

Towards a global understanding of how QoS affects QoE, general behaviors have been studied, [65] [66][67][68], leading the way to a broader characterization of the mapping between QoS and QoE.

Studies have led to mainly two generic quantitative relationships. The logarithmic relationship [68], is based on psycophysics studies on human perception and reaction to variations in the stimulus, and can be stated as:

$$\partial QoS = k \cdot QoS \cdot \partial QoE \tag{2–1}$$

where ∂ is the symbol for partial derivative, and k is a constant.

On the other hadnd, the exponential relationship [66] claims a user-centric approach where variations in QoE are not only dependent on variations in QoS but on the current level of QoE the user is experiencing

$$\frac{\partial QoE}{\partial QoS} = -k_1(QoE - k_2) \tag{2–2}$$

An example of what Eq. 2–2 represents is the fact that variations in QoS in terms of increasing packet loss are more noticeable for the user if the current level of QoE is high, compared to the case when the current level of QoE is low. This relationship has been confirmed in [69], and is also known as the IQX hypothesis

One such example of logarithmic relationship in Equation 2–1, is the one proposed in [70] and [67]for video quality in terms of increasing goodput for controllable use of the codec resources, i.e. increase in source rate.

2.4.1.3 *QoE and video*

QoE for video services has been analyzed from different angles and to different video applications.

The dominant factor that affect QoE in IPTV services have been studied and ranked in [71] according to subjective testing. Further, a study of the impairments in video streaming over HTTP in the internet,[72], discusses the user's tolerance towards start-up times (initial delay) and interruptions (freezes) in the playback.With respect to video conferencing applications, [73] performed subjective testing to study the effect of video coding parameters on QoE, in controlled scenarios.

Classification	
level	media, packet, hybrid
reference	reduced reference (RR), full reference (FR), no reference (NR)
domain	spatial, temporal, spatio-temporal

Table 1: Classification of QoE objective metrics

Considering the above, several features have been studied to improve user's experience in streaming to cope with network constraints, e.g. video coding parameters [71], or temporal impairments [72, 74]. Such solutions focus commonly on tackling one of the network constraints, i.e. on erasure protection solution in lossy networks or dynamic rate adaptation for best effort cases [75] [76].

2.4.2 Metrics

The challenge of understanding the user's degree of satisfaction comes hand-in-hand with the question of how to measure it.

The purpose of QoE assessment varies and may determine the most suitable measurement methodology. Examples of QoE assessment scenarios are: comparison between video codecs, monitoring QoE in a streaming service, or offline evaluation of the effects of QoS in end to end video QoE.

Subjective metrics aim for a close interaction of the user in the process of QoE assessment. Scientific assessment using Mean Opinion Score (MOS) methodologies for subjective QoE has been standardized by the International Telecommunications Union (ITU), i.e. [77]. The subjective metrics approach asks the user directly, in a very controlled environment, to rate his/her level of satisfaction, following the strict guidelines from the ITU. MOS measures "the likely level of satisfaction of a service or product as appreciated by an average user in a typical case",

Given the complexity of performing subjective metrics for all applications, its counterpart, objective metrics, provide aim for practicality. The target has been to develop objective metrics that could estimate the subjective metrics with high correlation.

QoE objective metrics for video can be classified as shown in Table 1. Objective metrics can be measured at media level, i.e. frame by frame basis, at packet level, i.e. using metadata information and QoS metrics, or a combination of both media and packet information.

Metrics may need, along with the received video, the original (reference) video (full reference), only partial information about the original video (reduced reference), or no information at all about the original video (no reference).

Further, the metrics may measure only spatial, temporal, or combined spatio-temporal features of the video.

The ITU has issued recommendations for FR and RR metrics for video that combine both media level and packet level measurements, and are usually linked to particular video codecs.[78]

An evaluation of several QoE metrics for video can be found in [79] or [63]. Following we define the objective metrics that will be used in the performance assessment of the work presented in this thesis. Although the literature offers metrics that may have better correlation to subjective metrics, the following metrics have the advantage of low complexity and are still widely used in the community, making them tractable. We note that the models proposed in this thesis are not dependent on these metrics.

2.4.2.1 *Full reference (FR)*

The FR metrics for video at media level compare the original and received video in a frame-by-frame basis.

PSNR The Peak Signal to Noise Ratio, PSNR, is based on the mean square error in a pixel-by pixel, frame-by-frame basis, between the original video and the received one.

PSNR is still of common usage in the scientific community although it does not correlate well with subject metrics. The main reason of its popularity is the low complexity in its calculation.

The PSNR value between two frames X_i and Y_i part of the original video sequence **X** and the received video sequence**Y** respectively is defined as:

$$PSNR_{X_i,Y_i} = 10 \log_{10} \left(\frac{L^2}{MSE_{X_i,Y_i}} \right) \tag{2–3}$$

where L is the dynamic range of pixel values (i.e. if the number of bits per pixel is 8, $L = 2^8 - 1 = 255$), and MSE_{X_i,Y_i} is the mean squared error between the two frames X_i and Y_i in a pixel-by-pixel basis.

The PSNR value between the two videos **X** and **Y** is calculated as the mean value of (2–3).

SSIM The Structural Similarity Index, SSIM, metric was initially defined for still image assessment but its use has been extended to video with a frame-by-frame analysis between the reference and the distorted video [80]. The SSIM metric identifies artifacts in the video image due to packet losses, compression rates, and other factors degrading video quality. Its range of values is [0,1] where the highest value (1) will represent identical sent and received video.

SSIM has shown higher correlation to subjective metrics, than e.g. Mean Squared Error, or Peak Signal to Noise Ratio[80]. The Structural Similarity approach of SSIM metrics is more closely related to the Human Visual System. More computationally complex SSIM-family metrics for video include temporal and spatial aspects performing closer to MOS scores.

SSIM is based on the comparison of luminance, contrast and structural similarity between the frames of the received and original videos.

Given the received video **Y** and the original video **X**, the SSIM index for two windows of image signals **x** and **y** components of a frame in **X** and **Y** respectively,

$$SSIM_{xy} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{2–4}$$

where μ_x and μ_y are the mean luminance intensities of **x** and **y** , σ_x and σ_y are the standard deviations of the luminances, and σ_{xy} is the covariance of the luminances. C_1, C_2 are stabilizing constants. The SSIM index of the entire videos **Y** and **X** will the the mean value of (2–4) over all the frames in the video sequence.

SSIM spread The Interquartile Range (IQR) value in statistical modeling is used to measure *SSIM spread*. The IQR provides the spread of a sample from its 25th percentile to its 75th percentile (the 50% of data surrounding the median value of the sample). A narrower spread (low value) would indicate that the majority of frames in the video sequence share very similar quality, while a broad spread indicates how uneven the quality is.

Spatio-temporal QoE QoE_{ST} Since both PSNR and SSIM are averaged values of quality for an entire video, the variations of these metrics throughout i.e. a streaming session are not reflected.

QoE_{ST}, is the spatio-temporal full-reference media level QoE metric, considered as in [81], where it has been shown to have good correlation to subjective QoE metrics. It is defined as

$$QoE_{ST} = \mu(\boldsymbol{\theta}) - w{\cdot}\sigma(\boldsymbol{\theta}) \tag{2–5}$$

where $\boldsymbol{\theta}$ is, in our case, the vector with frame-by-frame full reference video quality metric SSIM from each experiment, $\mu(\cdot)$ indicates the mean value function, $\sigma(\cdot)$ is the standard deviation and $w > 0$ is a weight value. This metric considers the variability of quality throughout the streaming session, hence it is able to represent the impact of time variations in the network.

No reference (NR)

Flow continuity, or temporal QoE QoE_T Flow continuity, or QoE_T is a highly valued user metric to guarantee QoE. Bad flow continuity would mean the user is perceiving constant freezes in the video. QoE_T measures at frame level the difference $\Delta_i = t_i^R - t_{i-1}^R$ between availability of two new consecutive frames at the receiver. QoE_T is the percentage of $\Delta_i < \gamma$, $\forall\, i$ in the streaming session, where γ is the tolerated threshold. $QoE_T \in [0, 1]$, where 1 would represent no freezes in the video flow.

$$QoE_T = P\{\Delta_i < \gamma\} \tag{2–6}$$

2.4.3 Situation Awareness study case

2.4.3.1 *Definition*

Situation (or situational) awareness enables good decision-making [82] and hence it is a major asset in e.g. emergency operations.

A broadly accepted definition of situation awareness is the following:

Definition 2. *Situation Awareness*. The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future. [83].

A three-level model is thus inferred, with the following stages: perception, comprehension, and projection.

Other models that have drawn attention to design completely autonomous “cyber” situation awareness are found in [82]. The work presented in this dissertation refers to non-computer-aided situation awareness.

2.4.3.2 *Video for situation awareness*

Video as a specific source of information for situation awareness can both give accurate accounts and contemplate the temporal aspects of dynamically changing situations. The possibility of capturing dynamic scenes, gives video an advantage in assessing temporary evolving events.

Examples of provision of situation awareness using video in the literature can be found in [84] for healthcare, or [85, 86] in emergency response. It can be a "people as sensors" video capture, or an unmanned capturing device as in surveillance or monitoring systems, whether static or moving. Expert practitioners in disaster management and humanitarian operations mentioned words like immediacy, rapidness, fast response, and the possibility of seeing actions in the make, as the advantages of video [3].

Video can be used in different scenarios characterized by their temporal immediacy [85]. In live mode, a dynamic sequence is streamed at the same time a particular event is happening to the pertinent control center authorities. In near live mode, a short lag in time lies between the event and the moment the control center is viewing the video material. In the scheduled mode, the video clips are viewed at specific moments during the assessment and coordination processes, to inform broader audience of current events. Finally, the video is a valuable archiving material for any post-event analysis.

Studies on the relevance of video in situational awareness suggest that video best supports the comprehension level in the situational awareness model, compared to other media types like photos or audio [86]. The argument lies in the broader and richer information video can convey, while it also enhances memory and recalling related activities/events claimed to be directly linked to better comprehension and therefore situational awareness.

To our knowledge, QoE in the context of situation awareness, and the negative effects of network impairments when transmitting the video for this purpose have not been studied so far.

2.5 CONCLUSIONS

This chapter has introduced the context and preliminaries necessary prior to presenting the main contributions of this thesis. The aim has been to introduce to the reader the key context and and definitions such that the derivation of our proposed frameworks is sufficiently understood and justified.

We have discussed the characteristics of heterogeneous networks, the approaches towards video adaptation and how cross layer optimization can be of use. Furthermore, we have presented the preliminaries on QoE for video, with the mappings and necessary metrics. We have introduced the concept of situation awareness part of our case studies are related to this scenario. Finally we have introduced the new paradigm of information-centric networking.

3 HOLISTIC MODELING FOR HETEROGENEOUS NETWORKS

3.1 OBJECTIVES

This chapter presents the holistic system design proposed to tackle the challenges of network convergence in heterogeneous networks.

The work presented in this chapter was framed within a project focused on integrating different real time sources of image communications in near real time for the purpose of mitigating damages in humanitarian and environmental emergencies[3]. The type of scenarios addressed were very dissimilar in nature, which consequently led to propose holistic ways for a system design.

3.1.1 Link to thesis objectives

This chapter presents the contribution that meets Objective (1) of this thesis, as defined in Section 1.2:

Objective 1: Propose a framework of heterogeneous networking.

As stated in Section 1.2 these models shall tackle heterogeneity, interoperability aspects in network convergence and provide a suitable framework for optimized video content delivery.

Moreover, the resulting holistic system design is coherent with the content-oriented approach of information-centric-networks. This is possible since under the holistic umbrella, dissimilar network instances seamlessly enable content delivery

3.1.2 Link to state-of-the-art

Our contribution differs from current state-of-the-art as follows. In [87] a holistic view is used in vehicular networks, with the target of improving scalability of data dissemination along the nodes . In [88] they have a layered approach to optimize multi-layer networks, where an alternative formulation of the min-cut concept is used. Our approach, does not involve layered networks and our holistic perspective is of the overall system design, in particular the network instantiations. Regarding optimization methods, the model for video streaming in wireless networks in [89] allows different instantiations to use different models for the cost function and the network constraints. Instead of their approach, in which each instantiation will solve a different optimization problem, we propose a general optimization formulation. We argue that our general optimization model can be used with different network instantiations by means of the min-cut characterization.

3.1.3 Contributions

- We propose a holistic framework to model heterogeneous network configurations considering. This framework is applied to the context of disaster recovery scenarios.
- Our methodology proposes two main novelties compared to state-of-the-art. First, we propose to model the communication network following a holistic approach. This framework guarantees generality and robustness of the overall design since any network can be considered as part of the different instantiations of a disaster scenario.
- Second, we characterize the particular network instantiations via their min-cut (well known parameter from graph theory), which can be used as underlying system model to formulate a cross-layer optimization for content delivery.
- This holistic approach allows us to formulate a general general cross layer optimization problem that will guarantee a seamless transmission for all possible network instantiations while it complies with user experience requirements. The network instantiations are represented with the min-cut. This is the basis for the QoE adaptive video model presented in Chapter 4.
- We show with realistic emergency examples how this model can be applied to dissimilar instances of humanitarian/environmental emergencies.
- Field trials in realistic scenarios showed the complexity of using heterogeneous networks for video transmission, and how the holistic approach can be useful.

3.1.4 Structure of chapter

The outline of this chapter is designed as follows. The holistic framework is presented in Section 3.2, which includes the system design, the characterization of network instances and how this framework enables a general a cross-layer optimization for video content delivery. Section 3.3 presents examples of realistic network instances and how they are modeled using the holistic design. Section 3.4 draws the final conclusions and the related publications are listed in Section 3.5.

3.2 HOLISTIC FRAMEWORK FOR HETEROGENEOUS NETWORKS

The inspiration of a holistic approach comes from the study of realistic scenarios for disaster preparedness, as introduced in Section 2.1.2.1.

Disaster preparedness and recovery operations can benefit considerably from integrated service platforms for disaster management operations in which communication, navigation, and observation data are provided. This integration includes the use of near real time access to in-situ multimedia content for proper disaster/emergency assessment. The complexity inherent to these operations, heterogenous in nature, makes guaranteeing operability a major challenge. Moreover, permanent infrastructure may be compromised and therefore unreliable. New approaches to the problem

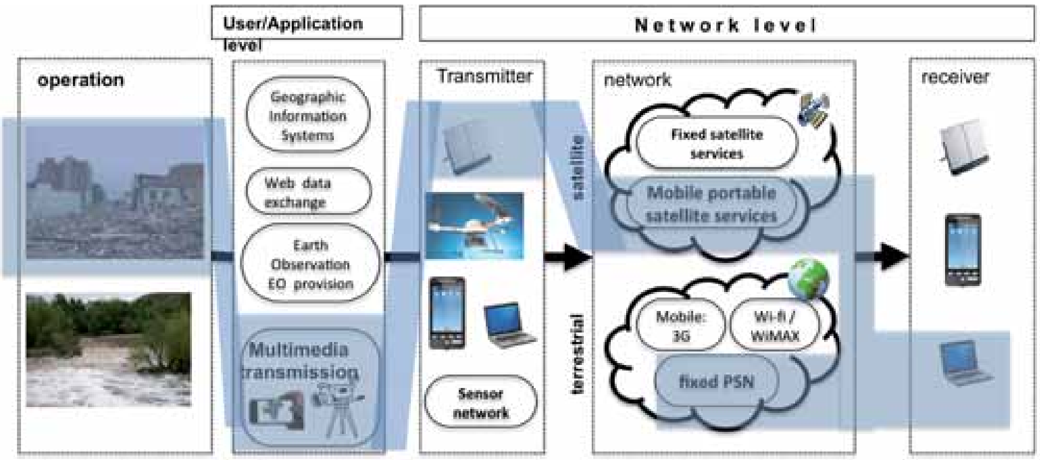


Figure 3.1: Holistic model representation for disaster recovery operations.

as a whole can be beneficial to guarantee a service to the humanitarian relief community, in which it can take advantage of these new integrated technologies.

From the communications point of view, different network configurations and terminals may be used depending on the operability of infrastructure as well as the availability of equipment from the assessment mission. In order to achieve an efficient design and optimization within the above described scenario, these multiple instantiations of the platform for disaster management call for a structured definition. A suitable approach for such a structured definition is to endow it with a holistic design approach. A holistic design simply means to take the system as a whole, considering all possible instantiations, to design its parts.

3.2.1 System modular design

The holistic framework adopted to characterize our system aims at allowing a modular approach that can represent different system instantiations.

We define three levels of interest in our design, *Network*, *Protocol* and *User/Application*. Each level contains a set of modules, each one having particular descriptors. An instantiation of an integrated service platform for disaster recovery will comprise a given set of module instances for each level. Fig. 3.1 shows a representation of our holistic proposal, with possible instantiations at each design level.

Our main focus in this paper is at the *Network* level. The holistic network model is characterized by the descriptors in Fig. 3.2. Our approach aims at expressing all network instantiations with these modules and ensuring a robust solution that takes into consideration its adaptability to different instantiations. The examples correspond to actual terminals and networks available to assessment teams of the humanitarian relief community.

Henceforth, we limit the instantiations at *Protocol* and *User/Application* level to IP-based and multimedia content delivery respectively.

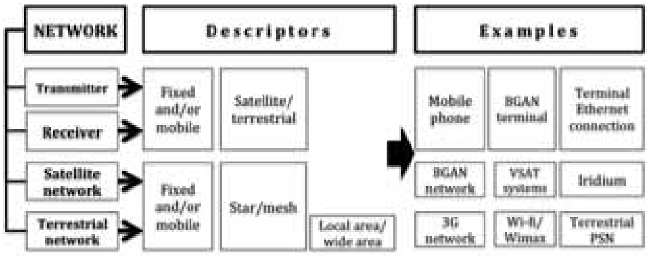


Figure 3.2: Holistic model. Modules and their descriptor for the *Network* level.

3.2.2 Characterization of network instances

In order to provide a robust solution that takes into account heterogenous network instantiations, we propose to identify the bottleneck that limits the throughput performance of each instantiation, and from all bottlenecks identify the one instantiation which poses the stringent constraints.

We model a network instantiation, as a directed graph $G = (V, E)$, with transmitter S and receiver R. V is set of network nodes (vertices), and $E \subseteq V \times V$ is the set of network links (edges). Each link $e_i \in E$ has an associated capacity w_i.

A cut C is defined as a set of links whose removal disconnects S from R. The cut value is given by the sum of the capacities of the links in the cut. The min-cut will be given by the cut with the minimum value.

The Min-Cut Max-Flow theorem (see [90]), states that the value of the maximum flow in a network corresponds to its min-cut. In other words, the maximum throughput of a network is given by the slowest link in the network path, and the only way to maximize the throughput of a network is by assuring maximum flow at its bottleneck. This theorem is the basis for our representation of the bottleneck of a network and enables the characterization of our network instantiations.

3.2.3 Cross layer content delivery optimization for a holistic network model

In particular, the holistic approach dictates that the transmission of multimedia should work seamlessly regardless of the technical differences in each instantiation. In order to do so, we will optimize the latency and bandwidth usage for QoS and QoE guarantees, by using cross layer design.

DESIGN CONSIDERATIONS IN A HOLISTIC APPROACH Two main considerations in our holistic framework are the guidelines for the cross layer design.

Given the heterogenous nature of the network instantiation, solutions tailored for specific network configurations are not possible. Common signaling exchange from lower layers of the protocol stack is not feasible either, i.e. a satellite service network instantiation [91] has a different configuration at protocol level than a terrestrial mobile or fixed wireless [92, 93, 94]. Moreover, some of the most common mobile satellite services used in disaster management have a proprietary protocol stack (i.e. Inmarsat's BGAN). However, it is reasonable to assume all modules are IP based, and cross layer design in which upper layers share information is feasible. Therefore, like in [25, 95], our design is based on control messaging from transport layer to the appli-

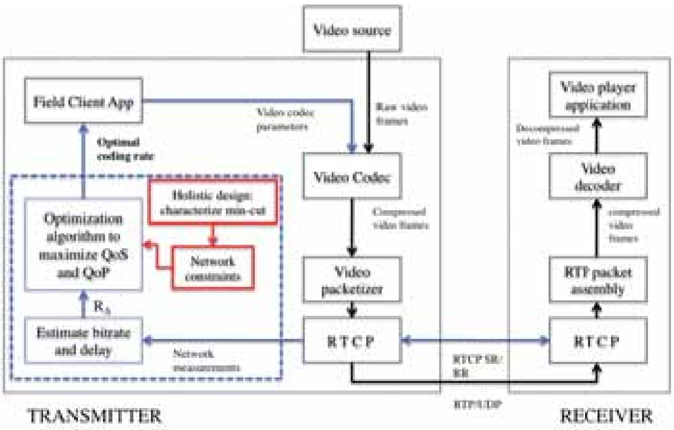


Figure 3.3: Block diagram of the implementation of our cross-layer design strategy with a holistic network design

cation layer. Information on network conditions can be extracted from the sender and receiver reports in the RTCP protocol built over RTP/UDP [96].

A common design assumption in multimedia delivery is a QoS guarantee from the network [89], but in disaster operations the limited availability of in-situ infrastructure may not allow this. Moreover, this type of premium service can be costly, as in some cases even Best Effort (BE) provision already is (i.e. in mobile satellite service). Therefore, we focus on optimizing for BE provision, where network conditions for the user are constantly changing, and the design cannot rely on bandwidth guarantee. In our design, the network measurements provided by the cross layer flow will help cope with these changing network conditions on the fly. The application layer will adapt the multimedia content to meet performance requirements.

Fig. 3.3 shows the block diagram of our proposed cross layer strategy. Feedback on network conditions is provided by the sender and receiver reports of the RTCP protocol, enabling the cross layer algorithm to update the optimal parameters given the delay τ and receiver packet rate R_Δ. The network constraints are calculated from the min-cut characterization in our holistic network design.

Optimization formulation After presenting our holistic approach and min-cut representation of the network in Sections 3.2.1 and 3.2.2, we present the formulation of a general general cross layer optimization problem that will guarantee a seamless transmission while complying with user experience requirements. In order to do this, we select the following concave utility function.

The concave utility function $U : \mathbb{R}^+ \rightarrow \{0, 100\}$ represents the QoE/QoS performance metric to be maximized. The set of constraints $C = \{C_1, C_2, \ldots, C_m\}$ is given by the min-cut bottleneck instantiation. The application layer parameters $A = \{A_1, A_2, \ldots, A_n\}$ will be adapted to maximize U. The general optimization formulation is set as follows:

$$\max_{A} U \text{ s.t. } C \leqslant 0 \tag{3–1}$$

we assume C and U are such that they pose a convex problem which can be solved with Lagrangian multipliers. The next section will show the final step of our 4-step methodology, and a solution for this problem for a particular cost function will be shown.

Chapter 4 will further dwell into the optimization formulation and how to solve it in practice.

3.3 FIELD/PRACTICAL STUDY CASES

3.3.1 Network Instantiations in realistic scenarios

In this section we present three instances of possible scenarios where the system for disaster management could be used and how they are modeled through our design.

3.3.1.1 *Selected instances*

INSTANTIATION 1. Disaster recovery

This instance in Fig. 3.1 describes a humanitarian disaster where the communications infrastructure has also been damaged. An organization like the UN would be a target user, as it is the main hub for global disaster management. In order to assess the dimension of the damage caused a person in situ, a field user part of a humanitarian organization, explores the area, and wants to send live footage of the current status to the headquarters of his/her organization for high level decision making. Permanent network infrastructure has been damaged but the assessment teams have BGAN terminals available in their emergency kits for emergency response. All modules of the network instance are thus identified, and the particular application we focus on limits o multimedia content delivery.

INSTANTIATION 2. Disaster preparedness:

In this case, Fig. 3.4 we consider flooding at regional level, where the European Union Civil Protection is a major actor in coordination. An in situ observer is monitoring evolving environmental changes,(e.g. river level, flood, oil spread). Video capture is suitable to capture the temporary evolution of such event. The observer has access to a terrestrial mobile network and can capture video and stream it through a 3G network to environmental organizations or Civil Protection. In this case the network instance has a combination of terrestrial wireless and fixed networks.

INSTANTIATION 3: Environmental monitoring of river levels

This instance in Fig. 3.5 represents a more complex network topology, in which a sensor network is monitoring several environmental changes in the anticipation of possible floods. This network consists of fixed stations, each station with weather, hydrological sensor, as well as visual sensor to capture images/video. The information captured is reporting to regional civil protection organizations. We consider the case of monitoring in the Amazon Basin, a region which faces high risk of potential environmental emergencies like flood and draught risks. The sensor network covers a

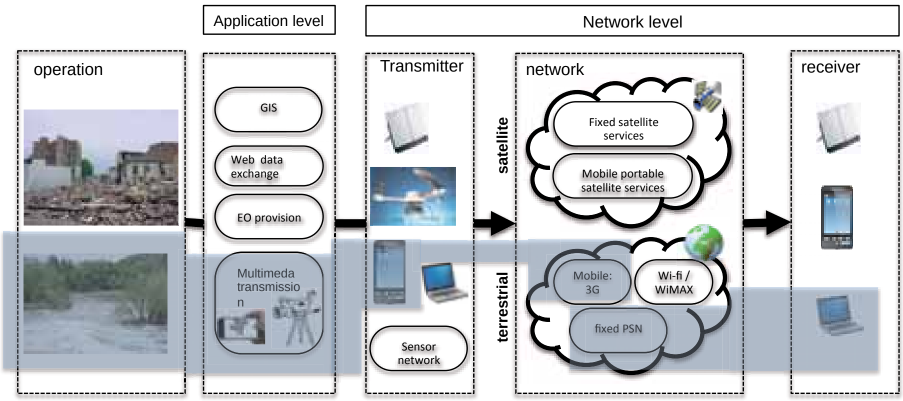


Figure 3.4: Instantiation 2. Civil protection use case at regional level.

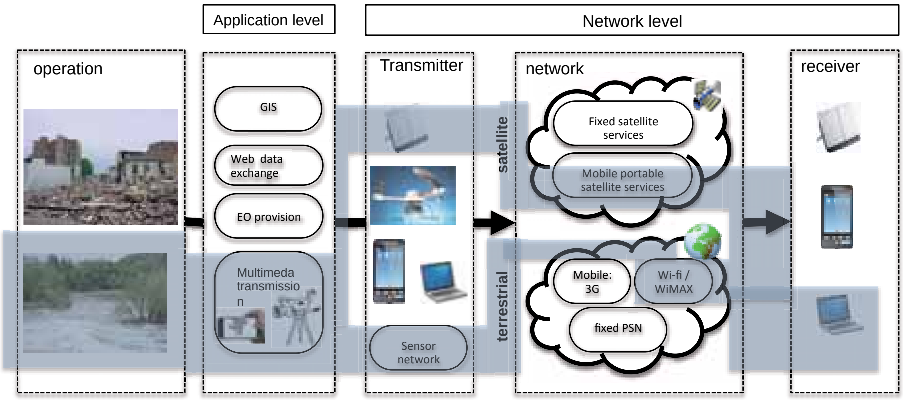


Figure 3.5: Instantiation 3. Environmental monitoring of rivel level.

Instance	example	Target user group	role	Communications equipment available to supporting teams
1	Earthquake, Tsunami	United Nations (UN). UN technical assistance cooperation (UNDAC)	main organization managing global disaster and damage assessment	ICT module for INDAC emergency response team: laptops (3-4), mobile phones (3), BGAN and Iridium/Thuraya terminals. At control center: wireless router, laptops, mobile phones, BGAN terminal
2	Flooding, regional level emergency	EU Civil protection mechanisms	Facilitate cooperation and civil protection assistance at regional levels	EU Assessment Mission kits (AMK): 1 BGAN terminal, 2 mobile phones
3	prevention: Monitoring of river levels	Civil protection	Facilitate cooperation and civil protection assistance at local/regional levels	Ad-hoc sensor network

Table 2: Realistic specifications for each instance in disaster management

region surrounding the city of Manaus, capital of the state of Amazonas in Brazil and its aim is to provide near real-time information on the water levels.

Table 2 shows the mapping of each instance in a realistic scenario, and the real communications equipment available to the different actors from the United Nations or European Civil Protection, as explained in Section 2.1.2.1.

3.3.1.2 *Min-cut identification in network instances*

We proceed to identify the min-cut of Instantiation1 and 2. Fig. 3.6 shows a simple diagram of the network architecture for both. Following we show the methodology used in identifying the min-cut for the case of the UMTS and BGAN satellite network instantiations.

We analyze the cuts of the graph in the path from transmitter S to receiver R in Fig. 3.6. For both cases we focus on the access network, assuming the core and backbone networks do not impose a restriction in the performance of the transmission. In the case of UMTS, if the links interconnecting the Radio Network Controllers (RNCs) are dimensioned correctly to handle the traffic from their base stations, the value of cuts with these links will be higher, compared to the cuts with radio access links. Therefore the cuts with minimum value in the graph are the ones including links between base stations and the transmitter S. The UMTS (Release 99) standard provides up to 384kbps uplink.

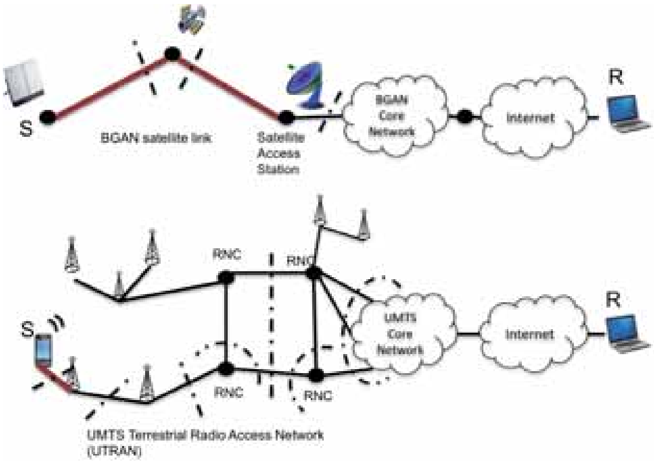


Figure 3.6: Characterization of the bottleneck for the two instantiations studied.

Access Network	type	Theoretical min-cut in the uplink
BGAN	satellite	492 kbps
3G->UMTS (R99)	mobile terrestrial	384 kbps
4G->LTE (R11)	mobile terrestrial	75 Mbps

Table 3: Summary of min-cut values for different commercial access networks.

A similar analysis can be performed in the case of the BGAN satellite network configuration. The min-cut is imposed by the satellite link, due to its low transmission bit rate in the L-band (a maximum of 492kbps) as well as by the inherent long propagation delay (250ms).

Table3 lists the summary of different commercial wireless access network technologies and their (uplink) min-cut. Note these are theoretical values as from the released standards, and in particular with mobile terrestrial, values may vary drastically depending on the deployment. Further, these rates are assumed without transport layer rate or flow control (no TCP), hence they would match UDP traffic.

3.3.2 Video using heterogenous networks in realistic scenarios

Field trials carried out within the GEO-PICTURES project [3] during October 2011 were a suitable opportunity to validate our proposed holistic approach for a network model.

The overall objective of the trials in the context of [3] was to test the different components of the integrated platform for disaster management in a close-to-real scenario with target users. Experts on European Civil Protection and disaster management operations carried out a civil protection training course in Brazil. The first course was carried out in Manaus, in the heart of the Brazilian Amazonian rainforest, and the second one in Rio de Janeiro.

An evaluation of the communications networks during the trials highlighted different aspects that affect performance in particular due to local network capabilities.

	Network instances	Locations	Network Characteristics
Scenarios	• BGAN-3G • 3G-3G • BGAN-BGAN	• Suburban/tropical (Manaus) • Urban (Rio de Janeiro)	• BGAN: bandlimited, geo-satellite communications • 3G: mobile terrestrial network

Table 4: Network instances tested in trials.

The heterogeneous characteristic of the two geographical locations reflected the complexity in network instances in which a solution for video streaming in disaster or environmental hazards may be used. On one hand, the particular location of the Amazon region where the communications channel is affected by heavy rainfall. On the other hand, a pure urban scenario in Rio de Janeiro, where the challenge is network congestion. Thus, our objective was the evaluation of the network operational aspects of the solution and the validity of the proposed holistic approach.

3.3.2.1 *Data collection methodology*

A video streaming application developed was tested from a laptop computer, having as a video source the data captured by an external USB camera. The streaming receiver module not only displayed the retrieved frames but it also collected network measurements, for offline analysis. The network-collected data, all at transport layer of the protocol stack, was processed offline to infer network statistics like packet loss, arrival rate, and inter packet delay.

Indoor and outdoor configurations were tested when possible, to simulate field to control center, and field to field connections. Different times of day were considered when testing in the urban scenario of Rio de Janeiro.

The traces collected were used to emulate network behavior in our QoE driven automatic video streaming, as shown in Section 4.3.1.

3.3.2.2 *Performance evaluation*

Table 4 summarizes the different scenarios considered throughout the trials, with the setup as shown in Fig. 3.7.

The performances was assessed in terms of the factors key in converged networks.

INTEROPERABILITY Interoperability between different networks was successfully accomplished. The choice for IP-based networks allows for such transparency, and interchangeable receiver and sender terminals with different access networks were in this case possible.

UBIQUITY On this matter several aspects are to be taken into account. From the field tests we have identified the following parameters an ubiquitous solution must

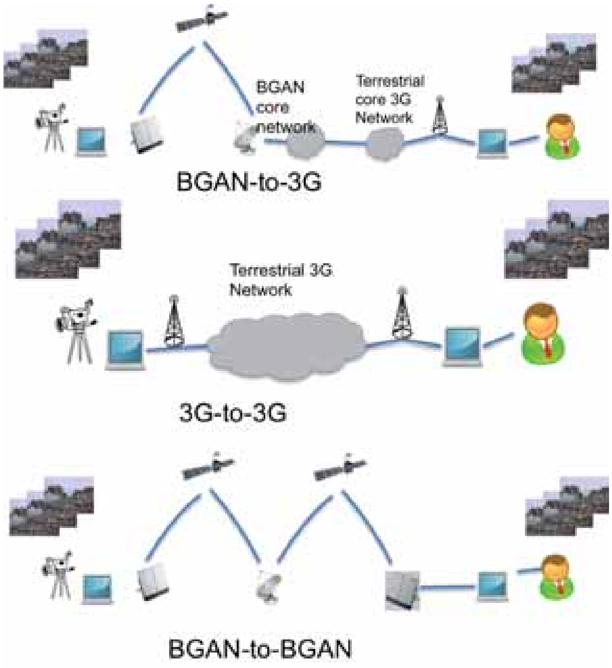


Figure 3.7: Network instances considered during the tests.

consider. A ubiquitous solution from the networking point of view must consider limitations of the physical channel, network availability, geographical location.

LOCATION The geographical location of Manaus (close to the equator) guaranteed easy connectivity from the terminal to the BGAN satellite constellation, since the Inmarsat's satellites are geostationary. The physical channel restriction hints to the type of scenario a specific network is most suitable for.

A determining factor in 3G networks, correlated to the geographical location, is the heterogeneous behavior of the network. This factor has much to do with the dimensioning and deployment of local 3G networks and the coexistance of legacy technologies with state-of-the-art deployment.

Particularly in Brazil, there is a big division in the network deployment per regions given the vast extension of territory. Furthermore, Brazil is now the market with biggest growth in the mobile broadband arena to compensate for the lack of fixed broadband infrastructure. Unfortunately, the real network capacity and coverage is lagging behind compared to the demand, highly increasing network congestion and poor performance. As a result, a video streaming application experiences very poor video, with high packet drops, leading to impossible video frame recovery and frozen frames.

The BGAN network, on the other hand, was in overall a very reliable communications solution, independent of local providers and hardly affected by packet loss (under 2%)

TIME OF DAY Additional factors affecting performance were the traffic behavior dependent on the time of day in urban areas, and climatological aspects affecting tests. Congested times in Rio de Janeiro account for a packet error rate of 26% for a 150kbps streaming session, compared to 1.75% during a non-congested time. It is worth noting that the frequency band in which the BGAN network works, the L band, in general is not susceptible to snow or rain precipitation.

3.4 CONCLUSIONS

In this chapter we have presented a novel approach to model heterogenous networks, considering factors such as interoperability and heterogeneity of network instances. With the proposed holistic system design, all network instances can be encompassed under the same model. Further, this system design enables a general optimization for content delivery. This approach is in line with information-centric networks as the nature of particular instances becomes transparent for the content delivery.

The methodology followed to build this framework allows to 1) guarantee a design that suits all the network instances considered, 2) formulate a general cross-layer optimization for content delivery by means of the min-cut characterization of each instance.

The holistic system design is applied to the possible instances in a disaster and emergency response scenario, and we show this with realistic instances. Further, field trials show the challenges of network instances when used for video transmission.

The model used in this chapter is the starting point for the adaptive video approach we propose in Chapter (4).

3.5 PUBLICATIONS

The framework and results presented in this chapter are part of the following publications:

1. M. A Pimentel-Niño, M. A. Vázquez-Castro. Cross layer content delivery optimization for holistic network design in disaster preparedness and recovery scenarios. In: International Conference on Multimedia Computing and Systems/International Conference on Multimedia and Expo.; 2011. p. 1-6.
2. Several deliverable reports for the FP7 project GEO-PICTURES [3]:
   a) D3: 3.1 Technology Research, 3.2 Development and 3.3 Tests.
   b) D2: 2.1 System Design and 2.2 Requirements.
   c) D4: 4.1 Trials and 4.2 Evaluation.

4 CROSS-LAYER OPTIMIZATION OF ADAPTIVE VIDEO

4.1 OBJECTIVES

This chapter proposes a QoE-driven cross-layer optimization for adaptive video in time varying network conditions. Benefits of this mechanism are assessed in two stages, namely emulation and experimental. Moreover, we evaluate the potential of combining this solution to a content-aware erasure protection method and quantify the advantages of following this approach to tackle different network constraints.

Both emulation and experimental tests are designed to consider satellite networks as one common network instance with very heterogenous and challenging QoS performance.

4.1.1 Link to thesis objectives

This chapter presents the strategy for QoE-driven adaptive video that accomplishes Objective (2) of this thesis, defined in Section 1.2:

Model a user-centric performance framework for adaptive video transmission.

As discussed in Section 1.2, this strategy:

- shall be QoE-driven.
- suit heterogeneous networks and their potential limited QoS guarantees.
- be in line with information-centric networking.
- should provide a feasible solution for practical adoption.

The motivation for this work is formed by the potentially time-varying nature and constrained nature of heterogeneous networks, and how this ultimately affects QoE of the transmitted video. Chapter 3 successfully modeled these networks with a holistic system design, which allowed for a formulation of a general cross-layer optimization for content delivery. This chapter picks up on this contribution to build a user-centric framework for adaptive video transmission.

4.1.2 Link to state-of-the-art

Recent rate adaptive approaches can be divided into solutions for interactive/conversational video [35], or server-client based for HTTP dynamic streaming [75], however they are not designed for satellite networks. The latter would be deeply affected by propagation delay and would require higher available bandwidth to meet QoE, while the former is not QoE driven. Further, both alternatives rely on heavy feedback

In this work, we propose congestion avoidance over RTP, specifically for QoE in satellite networks, using a utility-based optimization approach. Cross layer design [97] has been explored for satellite networks [91] and used in multimedia communications for video[53] and VoIP[55] . These approaches are driven by QoS objectives. In our setting, the advantage of our formulation lies in the flexibility of deployment in any IP-centric satellite network. Other approaches use parametrizations from standardized quality of video to obtain video quality adaptive algorithms [98]. Finally, mappings of subjective QoE metrics are also used as optimization functions [76]. The drawback of the aforementioned approaches is the heterogeneity in choices for mapping and scenarios, that may not be reproducible or generalized for broader adoption. Further long roundtrip times are not addressed, hence limited towards the heterogeneity of networks such as satellite.

A close solution to the combined approach we propose, [38] considers adaptation of video flows with Forward Erasure Correction (FEC), but it is not adaptive streaming, nor QoE-driven. The video source rate is fixed, hence the possibility of improving QoE through higher video goodput is not considered.

4.1.3 Contributions

The summarized contributions for this chapter

1. Optimized video transmission that is QoE driven, and use QoE/QoS correlation models utility function in line with QoE/QoS correlation models.

2. It addresses heterogeneous networks in best effort mode (lack of QoS guarantee), in a shared medium. In particular, long delayed networks and lossy channels are considered in the approach, resulting in a delay-driven adaptation algorithm that is not affected with the losses of wireless channels. Networks with long propagation delays are particularly considered in the adaptation and feedback mechanisms.

3. We provide a feasible solution that contemplates the feedback constraints of long-delayed networks. The solution is tested systematically, first in an emulation phase, second in a full implementation, and experimental results prove it

4.1.4 Structure of chapter

This chapter is structured as follows. The formulation of the QoE-driven cross-layer optimization is presented in Section 4.2, from its theoretical framework to its practical considerations. Section 4.3 presents the performance evaluation in its two phases, namely at emulation and experimental level. Further, the feasibility of using content-aware erasure protection is shown in this section. Finally, conclusions are drawn in Section 4.4 an the related list of publications is presented in Section 4.5.

4.2 QOE-DRIVEN CROSS-LAYER OPTIMIZATION

Let us recall the general formulation proposed for the optimization of content delivery in Section 3.2.3. The general optimization model in (3–1) optimizes the utility function U which is a function of the QoE with the QoS restrictions. We will proceed to instantiate such optimization for video in time varying networks.

4.2.1 Video model

We consider a video source modeled as the current state-of-the-art video codecs, such as H.264/AVC, VP8 or the latest H.265 [43] ,[99] .

Coded frames are grouped into Group of pictures (GOPs). Each GOP has three types of frames namely I, P and B. The Intra-frame (I-frame) is the first frame in a GOP followed by a series of Prediction frames (P-frames) and B-frames. For instance, a GOP can be given in form by IPPP, IBPBP, IBBPBBP, etc. An I-frame being intra-coded is encoded and decoded by itself. P-frames are inter-coded and depend on the previous frame (I-frame or P-frame) while B-frames are encoded by referring to both preceding and following I-frame and P-frame. I-frames are usually bigger in size than the other frames, and due to the interdependency of P- and B-frames an error in I-frames propagates through its entire GOP. Further, the correct decoding of P-frames is needed to decode correctly the following P or B frames. Finally, none of the frames depend on the B-frame and hence its loss does not effect the decoding of any other frames.

The use of different frames are usually application specific. Applications with a baseline profile (only I and P frames) are designed for low delay, low complexity applications [100]. In this work, we focus on the applications which consider only I and P frames.

We consider each GoP has a fixed number of frames N_{frame}. The frame rate r_{frame} is also constant such that the codec outputs each GoP in a fixed time $T_{GoP} = \frac{N_{frame}}{r_{frame}}$. We denote N_{GoP} as the total number of GoP's output by the codec during the whole streaming session such that $N_{GoP} \times T_{GoP} = T$.

The codec outputs the n^{th} GoP, coded at application layer rate r_{APP}, in bits per second, for $n \in \{1, 2, \dots, N_{GoP}\}$ at time $t_n \in \{0, T_{GoP}, 2T_{GoP}, \dots, (N_{GoP} - 1)T_{GoP}\}$. Although N_{frame} is fixed, frame sizes vary depending on the r_{APP}. For the n^{th} GoP, each frame is fragmented into multiple packets of equal length l_{packet}, and delivered for end-to-end delivery. We denote $K(n) = \left\lceil \frac{r_{APP} \times T_{GoP}}{l_{packet}} \right\rceil$ as the total number of packets from the n^{th} GoP. We drop the index n for simplicity in formulation, however, as the source r_{APP} is time varying, coding parameters depending on r_{APP} may also vary from one GoP to another GoP.

As noted in the model, we are assuming a video codec working in constant-bit-rate (CBR) mode, at the application rate r_{APP} .If variable bitrate (VBR) mode was used, the output rate of the codec could not be controlled externally. Additionally, our model expresses the video coding in terms of rate, rather than on quantization parameters, in order to extrapolate the codec parameters to a networking framework.

Finally, this simplified model is good approximation to the streaming scenarios depicted in this thesis, where video may be encoded in real-time and generated by the users, in contrast to server-client solutions where the video content has been encoded offline.

4.2.2 Theoretical framework

Let us consider the best effort wireless scenario with a network varying over time t.

The general formulation of our objective optimization problem is presented in (4–1), where the utility function U is dependent on the QoS parameters. The QoS parameters considered are the transmission rate R, delay τ and congestion-induced erasures ϵ_c, all of them varying with time t[1]. r_{av}^{max} is the upper bound on maximum rate offered by the network.

$$R^* = \arg\max_R U(R, \tau, \epsilon_c) \quad \text{s.t} \qquad R \leqslant r_{av}^{max}, \tag{4–1}$$

The utility U can be further specified considering the interplay between QoE and QoS. Let us consider an additive model, where the utility is composed of two functions, namely one representing QoE's improvement with increasing assignment of network resources, and a second one representing the dynamics degrading the network in the best effort scenario.

$$U(R, \tau, \epsilon_c) = U_{QoE}(R) - U_{QoS}(R, \tau, \epsilon_c), \tag{4–2}$$

$U_{QoE}(R)$ is a concave function, defined in (4–3) based on the logarithmic mappings from QoS to QoE. Studies have shown that if rate is increased in a controllable fashion (e.g. by increasing the application layer rate r_{APP} of the video, such that $R = r_{APP}$), QoE is benefited with a logarithmic relationship [67].

$$U_{QoE}(R) = \kappa \cdot \log(R), \ \kappa > 0. \tag{4–3}$$

(4–3) is in line with the QoE/QoS models described in Section 2.4.1.2, in particular with (2–1).

$U_{QoS}(R, \tau, \epsilon_c)$, on the other hand, expresses the penalizing effect of a congested network scenario, where injecting higher rate than the one available for the user translates into accumulating delay τ, and eventually an overflow of network buffers leading to packet losses. Hence, we formulate it as a bilinear function of τ and R in (4–4)

$$U_{QoS}(R, \tau, \epsilon_c) = \gamma(\tau, \epsilon_c(\tau)) \cdot \tau \cdot R. \tag{4–4}$$

Notice we define the function $\gamma(\cdot) > 0$ to strengthen or weaken the effect of U_{QoS} in the overall optimization depending on the level of congestion perceived, as proposed in [101, 102] for flow control applications.

4.2.3 Implementation as dynamic rate adaptation

4.2.3.1 *Solution to the optimization problem*

Proposition 3. *The optimization problem stated in (4–1) where the utility function U has been defined as in (4–2) is solved by using the discrete rate update algorithm (4–5), to find the value of R at time t_{k+1}. for $k \in \mathbb{N}$, where $T_{samp} = t_{k+1} - t_k$ is the network sampling time, δ is the step size and $\nabla_R \cdot$ is the gradient with respect to R.*

$$R(t_{k+1}) = \quad R(t_k) + \delta[\nabla_R U_{QoE}|_{t=t_k} - \nabla_R U_{QoS}|_{t=t_k}] \ . \tag{4–5}$$

---

[1] R,τ, an ϵ_c are function of time t. For clarity in stating the optimization problem we have dropped t in (4–1)-(4–4).

Proof. First we proof U is concave and hence an optimal value R^* that solves (4–1) exists. The function U_{QoE} from (4–3) is strictly concave increasing with R, while $-U_{QoS}$ is concave, decreasing with R. The sum of concave functions is concave, hence so is U and an optimal $R^* \leqslant r_{av}^{max}$ that solves (4–1) exists. Further, the gradient ascent method can be used to find the optimal R^*, where R is varying over time in the direction of the positive gradient of U: $dR/dt = \nabla_R U$. In practice, rate updates happen in discrete time, and if we consider sampling time $T_{samp} = t_{k+1} - t_k$, the rate control update is expressed as in (4–5). □

Notice that $\nabla_R U_{QoS}$ is changing with the current network conditions in time t, and hence knowledge of QoS levels at transport layer are needed in order to solve the optimization problem. Such knowledge of the network is based on feedback from the receiver end. If we consider feedback delay, the measurements represent network performance at delayed points. This is especially true in the case of long delay networks, e.g. satellite networks, where propagation delay is noticeable.

Proposition 4. *In the case of a delayed network, with propagation delay* τ_D, *the algorithm in* (4–5) *that solves* (4–1), *with* U *defined from* (4–2)(4–3)(4–4), *is expressed as in* (4–6)

$$R(t_{k+1}) = R(t_k) + \delta\left[\frac{\kappa}{R(t_k)} - \gamma(t_k - \tau_D)\tau(t_k - \tau_D)\right]. \tag{4–6}$$

Proof. The algorithm is triggered when new network measurements are available at the sender side, at time instant t_k, but those are measurements corresponding to the network state at time $t_k - \tau_D$. Further, if we consider the rate control update with sampling time greater than the network's roundtrip time: $T_{samp} > T_{RTT}$, we can assume the receiver is able to report on network changes related to the last rate control action from the sender of time t_k. Bearing this in mind we can express (4–5) as

$$R(t_{k+1}) = R(t_k) + \delta[\nabla_R U_{QoE}|_{t=t_k} - \nabla_R U_{QoS}|_{t=t_k-\tau_D}] \tag{4–7}$$

where the gradient of U_{QoE} is evaluated at time t_k and the gradient of U_{QoS} is evaluated at time $t_k - \tau_D$. Substituting U_{QoS} and U_{QoE} for (4–4) and (4–3) we obtain (4–6). □

The function $\gamma(\cdot)$ is chosen such that the response of the adaptation reacts faster to increasing delay constraints and packet loss as described by (4–8), where $t_i = t_k - \tau_D$. Note that $\gamma(t_i)$ responds to increases in ϵ_c that are accompanied with increases in delay τ. Hence other sources of packet erasures not related to congestion will not trigger a change in the rate control update.

$$\gamma(t_i) = \begin{cases} \gamma(t_{i-1}), & \gamma(t_{i-1}) = \gamma(t_0),\ \tau(t_i) \leqslant \tau^{max} \\ \lambda\gamma(t_{i-1}), & \tau(t_i) > \tau^{max},\ \epsilon_c(t_i) > \epsilon_c^{max} \\ \gamma(t_{i-1}) - \frac{1}{\gamma(t_{i-1})}, & \gamma(t_{i-1}) > \gamma(t_0),\ \tau(t_i) \leqslant \tau^{max} \end{cases}. \tag{4–8}$$

A major drawback in streaming over TCP is the lack of smoothness in congestion control that is unpleasant to the user's QoE [75]. The delay-driven rate update obtained from (4–5) using the value of $\gamma(\cdot)$ according to (4–8) provides a smoother output that is also capable of reacting fast to severe degradations in QoS. $\gamma(t_0)$ and $\lambda > 1$ are chosen for a desired response time, while τ^{max} and ϵ_c^{max} correspond to upper bound limits to τ and ϵ_c, set according to application requirements.

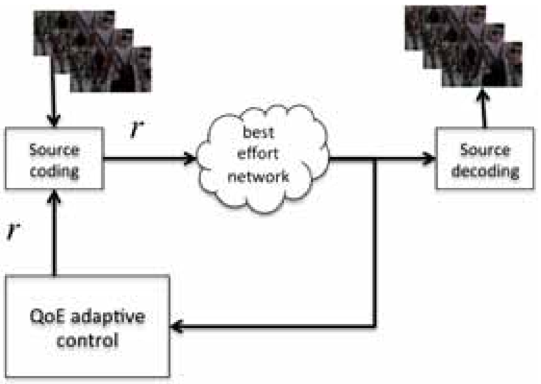


Figure 4.1: Block Diagram for online QoE adaptive video.

4.2.4 Cross-layer aspects and practical issues

The control algorithm for adaptive video requires a feedback loop from the receiver, as shown in Fig. 4.1. This signaling delivers information from the lower layers in the protocol stack to be used at the application layer, such that the application layer rate r_{APP} matches R^*. On the fly adaptation to network conditions can be thus guaranteed. In our solution we collect this information from the transport layer, hence assuring deployment regardless of the particular lower layer architecture of each satellite system's protocol stack. In this way lower layers are left untouched.

With the APP/transport cross-layer design based on the Real Time Control Protocol (RTCP), the video payload is transmitted using RTP over UDP [96], and the QoE adaptive video framework put in place with the RTCP signaling adds TCP-like flow control in best effort scenarios where it might be contending with other TCP flows. Moreover, as video streaming applications can tolerate losses rather than additional delay due to retransmissions for reliable data transfer, the overhead of using RTCP signaling is minimal.

The frequency of RTCP reporting is $1/T_{samp}$. In order to obtain ϵ_c and τ to estimate R^*, the following fields from both Sender and Receiver RTCP reports (SR and RR) are required (following RFC 3550 standard [96]): *fraction lost, Delay Since Last Report (DLSR), and Last Sent Report (LSR), and RR timestamp* $RR_{timestamp}$: $RTT = RR_{timestamp} - DLSR - LSR$.

In the discretized control algorithm of (4–5), the variation of delay is calculated as the difference between the current measurement available of RTT and the minimum RTT measurement received so far: $\tau(k) = \triangle RTT = RTT - RTT_{min}$.

The algorithm we propose in this section results in high granularity rate adaptation. Therefore it requires a video codec capable of performing on-the-fly encoding with fine granularity. The standard codec H.264/AVC offers such features, with possibility of adaptation of its quantization parameters (QP) at encoding time. The VP8 codec also offers such capabilities [103], with the option of real-time encoding with on-the-fly re-configuration of application layer rate.

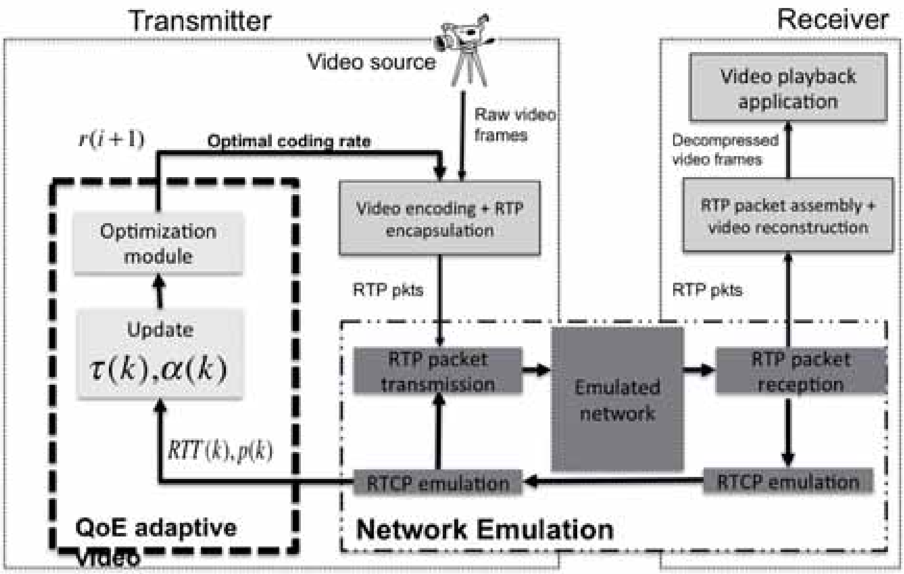


Figure 4.2: Block diagram of emulator platform.

4.3 PERFORMANCE EVALUATION

4.3.1 Emulation Phase

4.3.1.1 *Description of emulation platform*

In this section we present the experimental platform implemented to test QoE adaptive video solution. The platform was developed in C++, designed to be easily upgraded for deployment and future testing in the field. It uses video coding and packetizing modules to allow for a video streaming session of VP8-coded video packets. The packetizer outputs RTP packets at the sender and at the receiver reassembles them into coded frames appropriate for decoding. Fig. (4.2) represents the block diagram of the emulator platform.

OPTIMIZATION MODULE The QoE adaptation algorithm described in Section III is implemented in this module. The optimization module receives network statistics from the transport layer through the RTCP emulation and outputs the new rate, which the codec in the application layer must set. The process is performed when a new measurement from the RTCP reports is available.

NETWORK EMULATION: The network emulation module is in charge of simulating the transmission and reception of RTP packets as well as generating all the packet statistics necessary to emulate RTCP measurements. The Network emulation module receives the video coded into RTP packets and outputs the transmitted RTP packets to the receiver modules. This module emulates transmission of RTP packets through a network, which consists on network data collected from real channel traces. Other synthetic traces can be used as network input to the emulator. All packet statistics and time stamping are recorded in this module in order to emulate the RTCP reporting. Recording of all packet statistics is also a job of this module, enabling offline analysis of the system. Time stamping at the emulator is key in keeping track of statistics collected, contemplating propagation times and feedback delay of RTCP reports.

VIDEO CODING We use the open source VP8 codec [103]. VP8 is based on Discrete Cosine Transform compression and provides different configuration for streaming mode, including the possibility of on-the-fly reconfiguration of parameters such as the target codec bitrate. The codec is configured in Constant Bitrate (CBR) mode, and a new target bit rate $r_{APP} = R^*$ value is set on the fly whenever the optimization algorithm outputs a new rate update value. Other parameters of the video such as frame rate and resolution are not part of the adaptation but can be manually adjusted by the user.

4.3.1.2 *Performance metrics*

In order to evaluate the solution, we consider the following metrics.

ADAPTATION RATE: For Adaptation rate, the Granular mean absolute error adaptation rate μ_{ar} was defined. We take into consideration the granularity of RTCP measurement when calculating μ_{ar} and measure the mean absolute error (MAE) between the mean network available rate μ_{net} and the mean video coding rate μ_R during an RTCP period. A small value of μ_{ar} means the video adaptation is better.

QOE METRICS We consider Flow Continuity, PSNR, and SSIM metrics as defined in Section 2.4.2.

4.3.1.3 *Experimental setup*

VIDEO SEQUENCES The video sequences used are Foreman, Coastguard, and Crew, commonly used by the research community. In order to obtain larger video files to accurately assess the adaptation over time, the sequences were concatenated to obtain longer streams. Coastguard and Crew are considered to be "fast" sequences with more movement and complexity while Foreman has an intermediate rating.

NETWORK TRACES The channel traces were generated by collecting packet arrival statistics from a real-time video streaming application, in a similar configuration as in Fig. 2.1. The video sources were encoded at a fixed constant rate and transmitted over the BGAN satellite network. Packet arrival statistics, including inter-arrival delays and other packet metadata were collected into the traces at the receiver. With this data, the network traces for the emulation were constructed. The traces were taken at different geographical locations within the BGAN coverage Synthetic traces were also used for testing response to step increase/decrease in available bitrate.

4.3.1.4 *Numerical results*

The tests were carried out to compare QoE adaptive video with a non-adaptive solution. Synthetic network traces were used for the first test case, while real network traces were used for the second test case. The Performance metrics defined in Section 4.3.1.2 are used to evaluate the solution. The range of values in network availability considered, were chosen to match real scenarios when using a MSS services such as

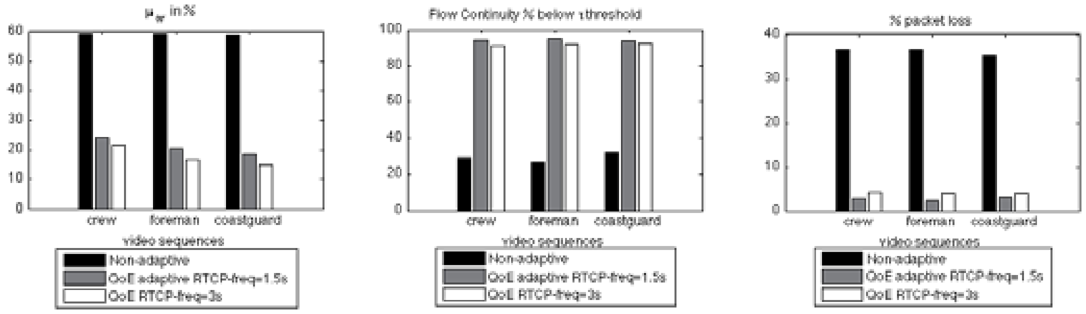


Figure 4.3: QoS metrics for Test case 1.

BGAN network. and from Eq. (4) were optimized offline through extensive simulations.

TEST CASE 1. SYNTHETIC NETWORK TRACES This test case considers the effect on QoE of video when a drop in network available rate occurs. This is emulated using a synthetic network trace, which for the first 50 seconds has a constant available bitrate of 350kbps, followed, by a drop of 150kbps for the rest of the transmission session.

Two options are considered: a non-adaptive solution and our QoE adaptive one. The non-adaptive solution configures the codec for a constant bitrate of 300kbps throughout the whole session.

The QoE adaptive starts with the same configuration as the non-adaptive one, and thanks to the feedback from the receiver through RTCP reporting is aware of the variations in the network. Two RTCP frequencies of reporting are considered: 1.5s and 3s.

The evaluation of performance metrics for this test case is summarized in Fig. 4.3. The difference between the QoE adaptive rate and the available network rate is evidently lower than the in the non-adaptive case, with a range from 15% to 24% compared to almost 60%,. This indicates how QoE adaptive solution tracks the availability of the network. High Flow continuity above 90% and packet loss less than 4% also indicates the improvements which will be reflected in overall better user experience and improved quality. In comparison, the non-adaptive solution is unable to control this effects and Flow continuity drops to 30% and packet loss goes up to 36%.

Fig. 4.3 also shows the results for two RTCP frequencies of reporting. As the RTCP interval increases, the control of flow continuity degrades and increase in packet loss is unavoidable. For an RTCP frequency of 3s packet loss increases up to 48% compared to the frequency of 1.5s, in the worst case for Crew video.

On the contrary, adaptation rate is better for a longer RTCP interval. An increase in RTCP frequency is thought to allow better accuracy on current network status and therefore better performance in the adaptation. Nevertheless, in our case, the better performance of a longer RTCP interval for is due to an added factor affecting adaptation response. The codec's delay in the response to a change in target rate has shown in our experiments to be such that it affects the algorithm's control update. An increase or decrease in 30% of target rate will take 6 seconds for the codec to reach. Therefore, a longer RTCP interval will provide a feedback with more trustful measurements since by that time the codec has been able to reach the target rate set by the adaptive algorithm. In order to counteract these effects, a mechanism that

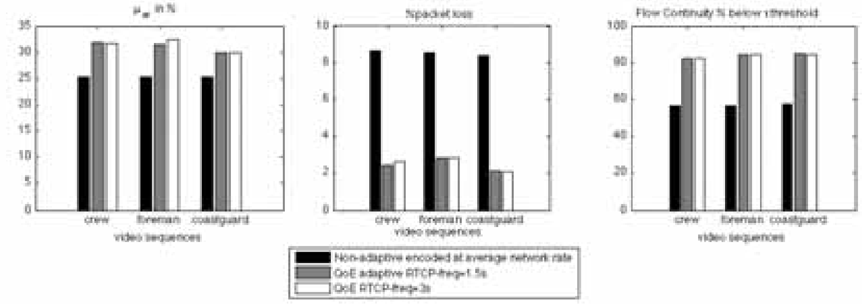


Figure 4.4: QoS metrics for Test case 2, Trace 1.

contemplates the delay in the response of the codec is necessary for the QoE adaptive solution to output more accurate rate updates.

TEST CASE 2. REAL NETWORK TRACES This test considers the variation over time of the network. Two network traces are presented here. Trace 1 is a time varying trace of 300 seconds, while Trace 2 is 40 seconds long. Two options are considered in this case:

1. Non-adaptive: video is transmitted at constant rate X. We assume the coding rate of the video to be the first measurement of the network trace, this would be a best case scenario of non-adaptive, since in reality, the non-adaptive solution does not have this information at hand.
2. Adaptive: video is transmitted starting at rate X.

The two traces, behaving differently will show how both longer streaming sessions and short ones can benefit from the adaptation. For this case, as well as in the previous one, we consider two RTCP intervals, 1.5s and 3s.

Fig. 4.4 shows the QoS related metrics for the streaming session using Trace 1. Although the non-adaptive solution shows to have 20% better adaptation rate, packet loss is 4 times higher and flow continuity is 30% lower. This shows how fluctuations in the network are not tracked with a non-adaptive solution.

Video Quality metrics in Fig. 4.5, especially the SSIM metric, also reflect the improvements shown by packet loss metric and flow continuity. Coastguard and Crew video, being the videos with more complexity and movement than Foreman show lower values in PSNR and SSIM, but the ratio of improvement compared to the non-adaptive solution is maintained.

Fig. 4.6 shows the quality metrics for the case of Trace 2. For the case of Coastguard and Crew videos, the quality is lower than for Foreman, due to the high complexity of the sequences, while an average sequence like Foreman has higher SSIM.

Worth of notice is the fact that the QoE adaptive solution shows more improvements in terms of quality for the fast videos, compared to the non-adaptive solution.

While in the previous case for Trace 1,there is hardly a difference in the metrics for the two RTCP frequencies considered, the case for Trace2 shows how higher frequency improves video quality and adaptation rate. This is especially visible for the case of the fast video sequences., which do benefit of higher frequency in the feedback to

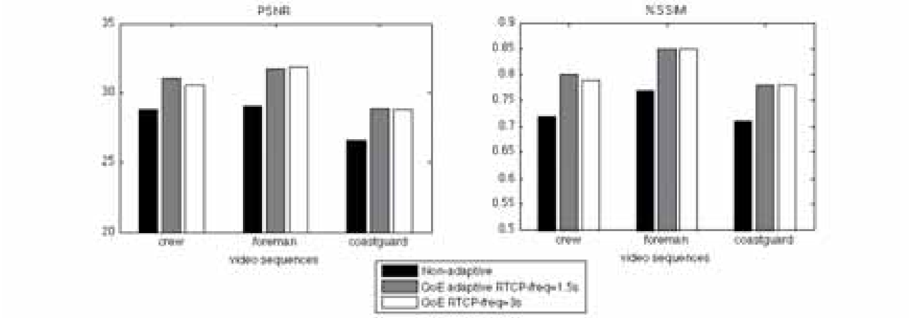


Figure 4.5: QoE metrics for Test case 2, Trace 1.

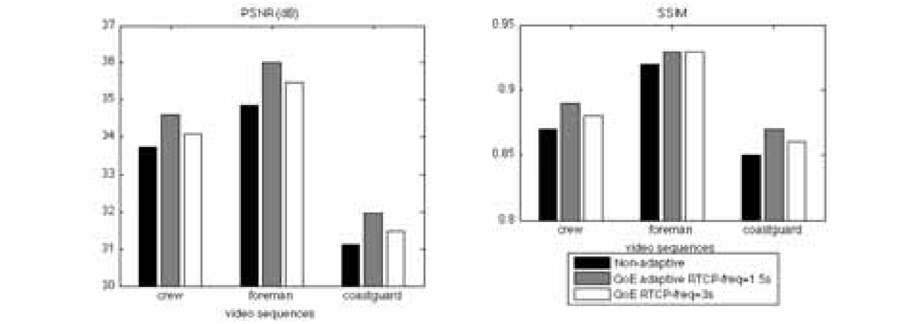


Figure 4.6: QoE metrics for Test case 2, Trace 2.

attain higher QoE. Nevertheless, the use case of Trace 2 represents a time interval ten times smaller than Trace 1, and therefore the effects of not having a frequent feedback is more notorious. From both cases for Trace 1 and Trace 2, we can infer that on a long time window, doubling the feedback frequency does not improve proportionally the QoE metrics. A higher frequency implies more overhead, which can be saved.

4.3.2 Experimental Phase

Following we present experimental results of QoE adaptation and the feasibility of joining both the QoE adaptive framework for best effort satellite links and the content-aware protection for bursty erasures. Performance is analyzed using QoE metrics.

We use a point-to-point configurable streaming application that uses RTP/RTCP at transport layer, and have implemented the necessary modules for RTCP feedback between the two ends. The main difference with the emulation phase previously described in Section 4.3.1 is that in the experimental Phase the RTCP protocol and feedback has been fully deployed.

Further, we use the VP8 [103], an open source video codec very popular in both interactive and client-server streaming applications [1]. VP8 is configured for low delay, low complexity real-time coding, following the source model in Section 4.2.1, and its output rate can be re-configured on-the-fly. Several standard video sequences were tested, corresponding to different types of content and video complexity [104]. Each sequence was concatenated into a loop to ensure a constant streaming input for long simulations.

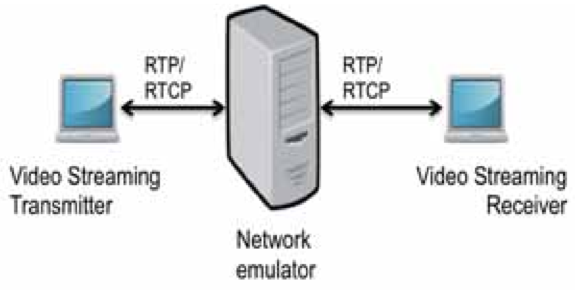


Figure 4.7: Experimental setup for performance evaluation..

4.3.2.1 *Experimental setup*

NETWORK EMULATION SETUP Receiver and sender are connected through an emulated network using NetEm emulator [105] installed in a server machine configured as a network bridge, as shown in Fig. 4.7.

In order to emulate best effort dynamics we use Netem's features to shape network traffic and add delay, to generate sudden drops in available rate and include the satellite link propagation delay.

VIDEO QOE ASSESSMENT Video QoE assessment is performed offline using full reference objective QoE metrics. As error concealment mechanism for post-processing and video quality assessment we assume frame copy strategy. In frame copy, if at least one of the packets conforming a frame is lost, the frame is considered lost and the last frame successfully recovered replaces it.

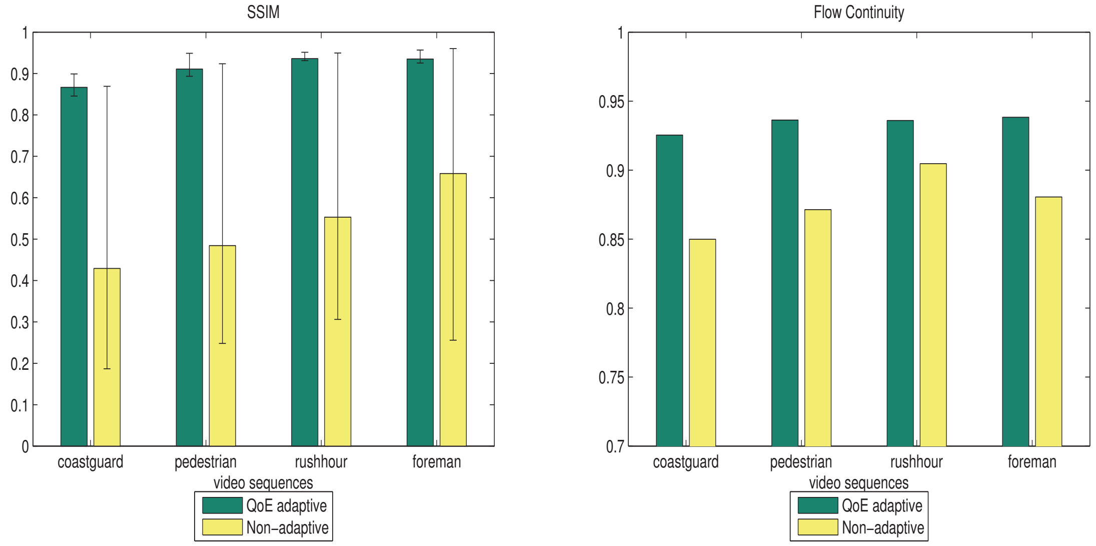
Performance is evaluated using the following QoE objective metrics, as described in Section 2.4.2: *SSIM, SSIM Spread*, and *Flow Continuity*.

4.3.2.2 *QoE adaptation performance*

In Fig. 4.8 we show the performance of the QoE adaptation algorithm from (4–6) when best effort network dynamics affect the available rate. We compare the results to a non-adaptive strategy.

The QoE adaptation algorithm has been embedded in the sender streaming application and is triggered by the RTCP cross layer feedback. The best effort satellite link is emulated in Netem such that a sudden drop to half the available rate occurs halfway through the streaming session that lasts 5 minutes. A non-adaptive strategy is unaware of this changing network dynamics, and continues to inject video packets at the same source rate.

Videos *coastguard, pedestrian, rushhour* and *foreman* were used in our tests, set at 15 frames per second (fps). *foreman* has the least video complexity among the sequences, and hence QoE for this video is least affected by the drop in available rate as shown in Fig. 4.8a. In this case adaptation improves QoE 50%. Higher complexity videos are the most damaged with the changes in the best effort network, such as *coastguard* and *pedestrian.* The advantage of using QoE adaptation is thus more noticeable for these types of videos, with over 100% improvement in SSIM quality metric. The error bars in Fig. 4.8a represent the SSIM spread, from which we can observe the minimal spread that can be obtained by using the QoE adaptive solution.



(a) SSIM video quality.

(b) Flow continuity.

Figure 4.8: QoE performance of online QoE adaptive video in best effort network.

In terms of Flow continuity, the QoE adaptive strategy improves flow continuity up to 10% for the higher complexity videos.

4.3.3 Evaluation of the potential of combining the framework with content-aware erasure protection

We evaluate the the interaction of the adaptive video solution provided in previous sections with content-aware erasure protection for lossy networks with bursty erasures.

4.3.3.1 *Unequal erasure protection using overlapping Network Coding (ONC)*

Let us consider the video model described in Section 4.2.1.

GoP consists of the IPPPP...P structure where the first frame is a I-frame and the remaining are P-frames. For the transmission, each frame is fragmented into multiple packets depending on its size. We consider each packet of equal length l_{packet}. Let us denote α as the total number of packets from the I-frame and β as the total number of packets from the remaining P-frames such that $K = \alpha + \beta$ is the total number of source packets.

Recall that in the GOP structure, I frames are needed by the other frames of the GOP for a successful frame recovery. Hence the loss of an I frame is more damaging and propagates throughout the entire GOP. We consider content-aware ONC to encode the K source packets into $N \geqslant K$ coded packets in order to decrease the probability of losing the I frames. The target is to provide better protection to I frames than a minimum distance separable (MDS) block code could provide. Recall that MDS codes guarantee full recovery of the original K packets if the erasures are less than $N - K$.

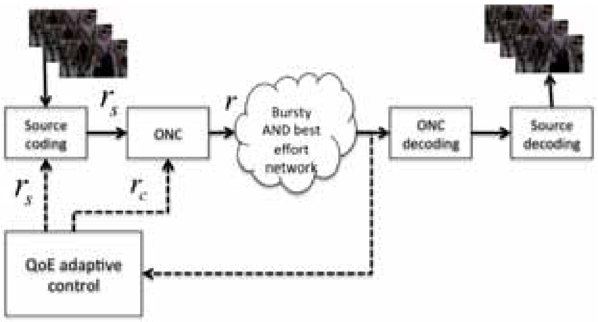


Figure 4.9: Combined ONC+QoEAdaptive.

Specifically, we can list two main advantages of ONC: i) Firstly, ONC guarantees an MDS-like performance for the I-frame when there are less than $N - K$ erasures; i.e., the I-frame is recoverable with the probability of almost one and ii) Secondly, even if there are more than $N - K$ erasures then unlike an MDS code, ONC helps to have some non-zero probability for the recovery of the I-frame. The tradeoff for ONC is that when there are less than $N - K$ erasures, then unlike an MDS code, ONC does not guarantee always the complete recovery of the P-frames.

ONC works as follows. For more details please refer to [9].

1. The K source packets are split into n overlapping generations
2. Random linear network coding (RNC) is performed to each generation i to obtain N_i coded packets, and generate a total $N = \sum_{i=1}^{n} N_i$ coded packets.
3. The overlap results into unequal protection UEP. The amount of overlapping results in different ways to obtain the total N coded packets
4. Content aware: Designed to protect more I frames

In order to integrate ONC to the QoE-driven adaptive algorithm, we consider the output R from (4–6) to be the overall rate budget to be split into source rate r_s and FEC rate r_c, such that $R = r_s + r_c$.

4.3.3.2 *Results*

SETUP In this section the experimental results show: 1) the improvement in QoE when content-aware ONC is used compared to no erasure protection, 2) the advantage of combining our solutions for best effort and bursty erasures satellite networks. Fig. 4.9 shows the block diagram of the combined solution.

For the experiments, we compare three different cases. In each case, we compare two scenarios. In the first scenario we distribute available rate between the source rate and FEC rate such that we keep $(r_c = \frac{r_s}{2})$. In the other scenarios, we use all the available rate as the source rate. The three cases are as follows: i) ($r = 150$, $r_s = 100$, $r_c = 50$) and ($r = 150$, $r_s = 150$, $r_c = 0$), ii) ($r = 300$, $r_s = 200$, $r_c = 100$) and ($r = 300$, $r_s = 300$, $r_c = 0$), and iii) ($r = 450$, $r_s = 300$, $r_c = 150$) and ($r = 450$, $r_s = 450$, $r_c = 0$). For all the three cases, corresponding values of α, β, K and N are given in Table 5.

For the packet erasure process,we consider a Gilbert-Elliot erasure channel model which is a two-state continuous time markov process $\{a_i(t)\}$, with $a_i(t) \in \{0, 1\}$. A packet transmitted is lost if the path is in bad state, i.e., $a_i(t) = 1$, and otherwise

r_s (in kbps)	r_c (in kbps)	α	β	K	N	ONC	Gilbert-Eliott Erasure channel	% I-frames lost	% P-frames lost
100	50	3	6	9	13	$(N_1, N_2) = (4, 9)$	$(\frac{1}{\lambda}, \frac{1}{\lambda_e}) = (\frac{1}{5}, \frac{1}{50})$	9.9%	10.4%
150	0	4	9	13	13	NO FEC	$(\frac{1}{\lambda}, \frac{1}{\lambda_e}) = (\frac{1}{5}, \frac{1}{50})$	15%	22.3%
200	100	5	13	18	27	$(N_1, N_2) = (8, 19)$	$(\frac{1}{\lambda}, \frac{1}{\lambda_e}) = (\frac{1}{10}, \frac{1}{100})$	9.0%	11.3%
300	0	7	20	27	27	NO FEC	$(\frac{1}{\lambda}, \frac{1}{\lambda_e}) = (\frac{1}{10}, \frac{1}{100})$	14.7%	20.9%
300	150	7	20	27	40	$(N_1, N_2) = (12, 28)$	$(\frac{1}{\lambda}, \frac{1}{\lambda_e}) = (\frac{1}{14}, \frac{1}{140})$	8.8%	13.3%
450	0	10	30	40	40	NO FEC	$(\frac{1}{\lambda}, \frac{1}{\lambda_e}) = (\frac{1}{14}, \frac{1}{140})$	11.00%	23.3%

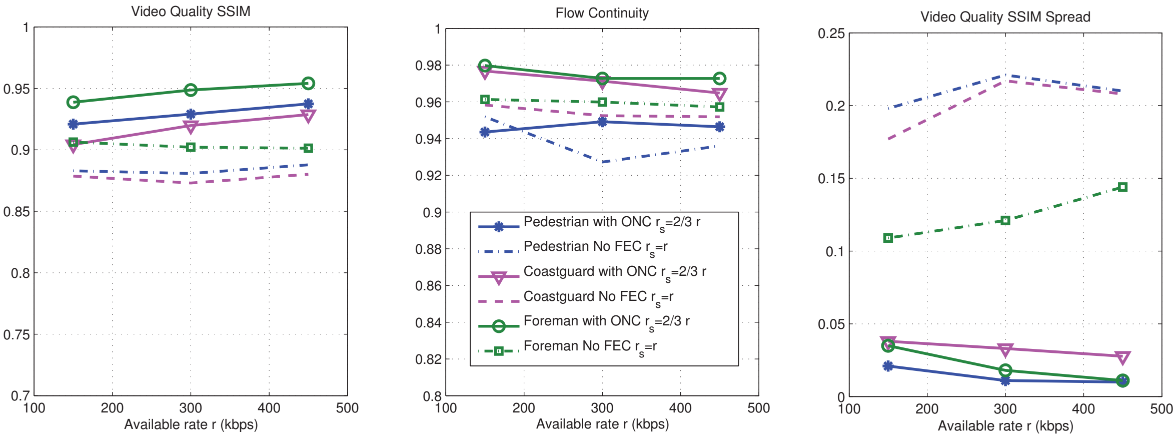
Table 5: Experimental parameters for ONC in combined solution with QoE adaptation.

received correctly in good state when $a_i(t) = 0$. The transfer probability from good to bad state is defined as λ and from bad to good state as λ_e. As long as the channel stays in the bad state, the packet erasure process is bursty where length of each burst is a geometric random variable with mean of $\frac{1}{\lambda}$ and the interval between two consecutive bursts is a geometric random variable with mean of $\frac{1}{\lambda_e}$.

We consider the Gilbert-Elliot erasure channel process based on the first scenario. In all the three cases, we choose the erasure model where on an average the lost packets per coding window is more than $N - K$; i.e., $\lambda > N - K$ and we repeat this erasure with $\lambda_e = 10\lambda$. For FEC we use ONC which helps us for content aware protection. For all the above three cases, we select the ONC parameters as given in the Table 5, conforming also to VP8's frame sizes at the corresponding values of r_s. Note that this selection is not based on any optimization.

RESULTS Table 5 provides the final percentages of I-frame lost and P-frames lost after decoding. We transmit a video at 15fps, containing 1000 GoP's with each GoP having 15 frames: 1 I-frame and 14 P-frames. For the cases, where ONC is used for erasure protection, residual erasure probability for I-frames and P-frames have been substantially reduced as shown in Table 5 as compared to the cases where ONC is not used.

The results in Fig. 4.10 show the QoE advantage of using ONC for bursty errors for different values of available rate R. Using part of the estimated value R in (4–5) to implement content-aware ONC provides and advantage in terms of mean value of SSIM of 6%. Fig. 4.10c shows that the spread of SSIM, which accounts for the impact on the temporal variations in quality, is up to ten times better when using ONC. This 90% reduction in SSIM spread results in variations below 5.5% around the value of SSIM from Fig. 4.10a, guaranteeing a smooth QoE performance throughout the entire streaming session. We can conclude that by using part of the estimated available bandwidth for erasure protection, higher overall QoE can be achieved than by using the entire available rate for video data in a bursty erasure network.



(a) Video quality SSIM. **(b)** Flow Continuity. **(c)** Video quality SSIM spread.

Figure 4.10: QoE performance for combined approach. Results for ONC considering varying available rate r.

4.4 CONCLUSIONS

We have presented in this chapter a QoE driven solution for adaptive video that successfully achieves Objective (2) of this thesis. The user-centric framework for video transmission is designed within our heterogeneous networking framework developed in Chapter 3 and is a feasible option for long delayed band limited networks.

We follow a utility optimization approach for this novel framework, where the utility function represents the QoE/QoS tradeoffs in a time-variant network. The resulting adaptation algorithm is delay-driven and not affected by other sources of packet erasures. Further, it is a feasible user-centric solution suitable for long-delayed networks that cannot afford per-packet acknowledgements.

We have evaluated the performance of the solution from a number of angles, namely QoS and QoE improvements. In a first stage, tests were carried out in an emulation platform. The second stage, where a full implementation of the solution is used in an experimental scenario. The performance evaluation through our experimental platform includes the use of different types of videos and real as well as synthetic network traces showing the capabilities in real scenarios.

Our performance evaluation shows that in terms of QoE, our solution outperforms a strategy unaware of time varying network dynamics. Moreover it is specially beneficial to higher complexity videos with more motion content. Finally it proves to be suitable for long-delayed networks, as it does not require a high feedback frequency as other schemes in the literature, in order to achieve a high performance in terms of flow continuity and better usage of the available network rate.

In addition, we have further studied the feasibility of combining the video adaptation framework to a solution that enhances reliability with a content-aware erasure protection strategy using ONC. We conclude that in a scenario where bursty erasures affect a best effort network, sharing the rate budget for adaptive video with an erasure protection scheme like ONC significantly mitigates the temporal variations of QoE due to erasures, by a factor of ten. ONC guarantees an overall time-invariant high QoE performance throughout the streaming session.

4.5 PUBLICATIONS

1. M. A. Pimentel-Niño, P. Saxena and M. A. Vázquez-Castro. “QoE driven adaptive video with Overlapping Network Coding for Best Effort Erasure Satellite Links”. 31st AIAA International Communications Satellite Systems Conference, Florence, Oct. 2013.
2. M. A Pimentel-Niño, M. A. Vázquez-Castro, and H. Skinnemoen, "Optimized ASMIRA – Advanced QoE Video Streaming for Mobile Satellite Communications Systems", 30th AIAA International Communications Satellite Systems Conference, 2012.
3. Documentation for the ESA funded project BEATLES [4]:
    - a) Technical specifications.
    - b) Pilot system architecture report.
    - c) Design justification file.

5 VIDEO ADAPTATION OVER HETEROGENEOUS NETWORKS

5.1 OBJECTIVES

This chapter presents the main contribution of this thesis. We present a complete model and solution with a QoE/QoS framework that performs a novel time/space decoupling for cross-layer optimization of video transmission in constraint networks suffering of congestion and erasures.

Moreover we introduce novel perceptual semantics to enhance video adaptation capabilities in an information-centric compliant setting. Within the QoE/QoS framework we enable a perceptual semantic adaptation loop, which complements the cross-layer optimization and will target the specific user's demand for improved perception.

With the interest of exploring point-to-point video services beyond entertainment purposes, we were inspired by the use of live video streaming for the provision of situation awareness in disaster and emergency response events [3].

While this chapter shows the model applied to this illustrative scenario, it can be used in a more general context, in order to optimize the user's satisfaction of live video streaming in heterogeneous networks.

5.1.1 Link to thesis objectives

This chapter deals with the third objective of this dissertation, as traced in Section 1.2:

Develop a complete model and solution for video adaptation over heterogeneous networks.

In order to achieve this objective, the work presented in this chapter uses the modeling frameworks developed in Chapters 3 and 4 to build a complete and novel model for user-centric adaptive video in heterogeneous networks, coherent with information-centric networking.

This chapter will focus on the QoS/QoE interplay and propose alternatives to enhance user's satisfaction for video services beyond recreational purposes. As a particular study case, we address the problem of provision of situation awareness through video services.

5.1.2 Link to state-of-the-art

PROVISION OVER SATELLITE In this work, we propose congestion avoidance over RTP, specifically for QoE in satellite networks, using a utility-based optimization approach, as formulated in Chapter 4.

With regards to the network constraints due to the wireless nature of the satellite scenario, we propose the use of systematic random linear network coding (SRNC) for erasure protection and its optimal integration into adaptive video to improve QoE. With the use of SRNC, the erasure recovery performance is similar to Maximum Distance Separable (MDS) codes like Reed Solomon (RS) codes [52], allows for feasible coding at intermediate nodes, and it's progressive decoding capability decreases decoding delay.

Concerning QoE in video in particular, several features have been studied to improve user's experience in streaming, e.g. video coding parameters [71], or temporal impairments [72, 74]. Such solutions focus independently on erasure protection solution in lossy networks or dynamic rate adaptation for best effort cases [75] [76].

In contrast to the afore-mentioned approaches, not only do we provide a full solution for both erasures and congestion, but we base upon the notions of QoE from [64], focusing on mission-specific: 1) system influential factors on QoE that relate to the networking scenario at hand and 2) perceptual features in QoE for video that will guarantee delivery of valuable information to the user (e.g in emergency situations). We follow the QoE vs. QoS correlation modeling [65], and propose a time/space graphical analysis of QoE when both congestion and erasures are affecting video delivery. The aim is high temporal (related to congestion avoidance) and spatial (related to reliable transmission) QoE procurement.

ADDITION OF PERCEPTUAL SEMANTICS In multimedia, "classic" semantics deals with heterogeneous metadata that sensors observe and/or tag when capturing video. As such, it has applications in information retrieval, integration and aggregation of varied data types as in semantic-aware delivery of multimedia [106]. Further, semantic tagging describing pure observations is used in computer-based systems with artificial intelligence to perceive and abstract situations [107]. Rather than doing perception through classic semantics, we propose a novel human-analysis-driven perceptual semantics to tag the videos, based on the temporal/spatial characteristics of the video a user is perceiving, as means to more specifically target and improve his perceptual needs.

The term perceptual semantics has been used in [108] for automatic feature extraction of video based on image segmentation and target internal changes within the video source coding mechanisms. Our approach differs, as it does not limit to particular feature extraction and it focuses on offering a solution for video communications in a non-intrusive manner towards video codecs. Further, we involve the user in tagging first-level perceptual features, rather than only observations as in [109], for perceptual-based networking. To the best of our knowledge, such diversion from classic semantics has not been explored before.

We frame the novelty of perceptual semantics such that it complements cross-layer optimization schemes that help cope with network constraints. Additionally, the overall framework can be mapped to current content-centric approaches of information-centric-networks.

5.1.3 Contributions

The main contributions of our work can be summarized as follows.

5.1.3.1 *Contribution for provision over satellite*

- We decouple the problem of combined congestion and erasures in best effort satellite networks affecting QoE of video, identifying congestion with freezes in the video playback and packet erasures with artifacts degrading the video, both perceptual features that contribute to the user's experience. Hence, we are able to formulate separate QoE driven cross-layer optimizations for the time (freezes) and space (artifacts) domain.

- The jointly operative QoE-driven cross-layer optimizations in time and space domains are designed for feasible online video adaptation and optimal adaptive erasure coding suitable for constraint networks.

- Our novel time/space graphical analysis of QoE proves to be useful as it allows time/space graphical representation of 1) network performance through congestion and erasures and 2) Service performance from the user's perspective, i.e., QoE.

- We use SRNC for erasure protection which provides three fold benefits: similar erasure correction performance to state-of-the-art MDS like RS codes, improvement in packet recovery time due to progressive decoding and flexibility in choosing coding parameters with variable code rate for the adaptive video streaming.

- Our experimental results show that the joint QoE optimizations in time and spatial domain achieve in overall a planar, homogeneous, performance with high values of QoE metrics, regardless of both erasures and congestion degrading the network.

5.1.3.2 *Addition of perceptual semantics*

- We introduce novel perceptual semantics for video adaptation to enhance situation awareness in non-computer aided processes as in emergency operations.

- Our proposed perceptual semantics relate to end user requested resolution in the temporal domain for a better assessment of event's evolutions seen from streaming video. Adaptation is enabled at transmission via a perceptual semantics feedback loop to adapt source coding on-the fly in terms of frame rate.

- The overall framework contemplates the use of an underlying cross-layer optimization that copes with network congestion and erasures in best effort scenarios. We show through simulations that within the proposed framework, the perceptual semantics are preserved.

- Moreover, we show it complies with information-centric-networking philosophy and architecture, such that it is in line with content-aware trends in networking.

- In addition, we contribute to the study of situation awareness from the QoE perspective, by identifying the factors that contribute to user's satisfaction with this type of service, and the potential enhancements that would benefit his/her experience.

5.1.4 Structure of the chapter

This chapter is structured as follows. Section 5.2 presents the scenario of interest. Section 5.3 presents the QoE/QoS time/space decoupling model the cross-layer framework that will enable its implementation. Section 5.4 presents the cross-layer optimizations in both time and spatial domains that derive from the QoE/QoS model. Section 5.5 presents the experimental results with the three-dimensional analysis of performance. Section 5.6 presents the perceptual semantics model, how it diverges from classic semantics and how the semantig tagging is performed. Section 5.7 presents how the perceptual semantics model can be integrated into an adaptive video framework. Section 5.8 shows the performance analysis with simulation results. Section 5.9 draws the final conclusions of the chapter, and Section 5.10 lists the publications related to the work presented.

5.2 SITUATION AWARENESS SCENARIO

We consider point-to-point live streaming of user generated content for purposes beyond recreational.

As a study case, we take the emergency scenario, where traditional terrestrial communications infrastructure is unavailable. The video streamed by a user from the affected area will e.g. help trained emergency responders in attaining situation awareness. Such assessment will help to e.g. coordinate rescue and relief operations or aid for telemedicine purposes.

The users involved in this point-to-point setting are of two kinds: the sender acts as a people-as-sensor providing live video content, and the receiver (e.g the decision-maker) has the job of extracting valuable information of the current situation “in the field” from the streamed videos. We assume such job does not use computer-aid techniques (i.e. artificial intelligence), as in [85][86].

The holistic framework to model the heterogeneous networks from Chapter 3 is applicable. We assume in this work that the sender has access to a band-limited communications network (i.e. the satellite network instance from Section 3.3.1) . Long round trip times, best effort service and erasures due to the wireless medium are the network constraints considered.

While the models we propose are inspired by this scenario, we would like to highlight that they can be applied to a wider range of video services being streamed through heterogeneous networks suffering of the network constraints we point out.

5.2.1 Video use-case for situation awareness

In an emergency scenario, good decision-making and rapid response are based on the awareness of the situation, the people, the resources and how the event is evolving over time.

We recall from Section 2.4.3 that video as a specific source of information for situation awareness can both provide 1) precise time-space accounts of the scenario in the critical situation 2) insights to the temporal aspects of dynamically changing situations. Further, the receiver wishes to assess the situation with as close to real-time as possible [85].

5.3 QoS/QoE MODELING BY TIME/SPACE DECOUPLING

5.3.1 System influential factors indicators

Definition 5. Quality of Service QoS is the ability of the network or service to provide or guarantee a certain level of performance to a data flow.

We consider the following QoS metrics to quantify the influence of effects of congestion and erasures in the best effort satellite scenario.

Definition 6. Erasure rate ϵ is modeled as i.i.d random variable. It represents packet erasure rate due to channel fading in wireless links.

Definition 7. Congestion-induced erasure rate ϵ_c is modeled as i.i.d random variable. It represents packet erasure rate due to the congestion in best effort wireless networks.

Definition 8. Delay τ is the cue of congestion in the network.

Definition 9. Degree of congestion η represents how congested the network is with respect to the maximum available rate r_{av}^{max} [1] where $\eta = \frac{r_{av}^{max}}{r_{av}}$, $0 < \eta \leqslant 1$ with r_{av} as the current available network rate.

5.3.2 QoE framework

We first present the framework used to decompose the system and perceptual aspects of the scenario in Section 5.2, according to standard QoE definitions. We identify the problems degrading QoE from each aspect, mapping them to its Quality of Service (QoS) counterpart, to either a space or time domain. Finally, in order to tackle these QoE degradations, we propose to decouple tempoal and spatial domains which entails a decoupling of the system effects of congestion and erasures on QoE.

We follow the definitions of QoE described in Section 2.4.1 and the influential factors affecting QoE from Def. 1, focusing on the system influential factors. Human factors are out of scope, while context factors can be a natural extension of the work we present here.

QoE system influential factors relate to the technical aspects affecting quality of the application or service, including media capture, coding, transmission, playback among others. Such factors can affect user experience of video, with noticeable degradations such as artifacts, blockiness, freezes. In the scenario considered in this chapter, the QoE system influential factors are linked to the underlying network performance, e.g. the best effort wireless satellite network and its QoS.

Perceptual features of QoE are the perceivable characteristics of the user's experience contributing to the overall quality [64]. In our framework, we related the QoE

---

[1] r_{av}^{max} depends on the underlying network (i.e. for the BGAN network in the best effort mode $r_{av}^{max} \approx 500$ kbps)

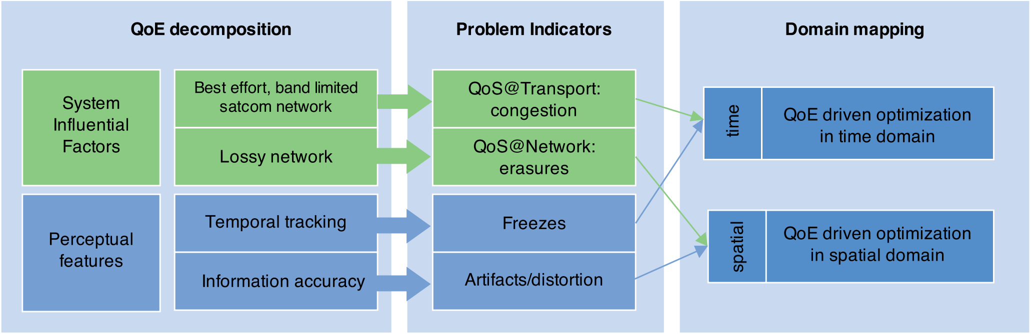


Figure 5.1: Scenario-specific QoE framework with decoupling in time and spatial domains of QoE.

perceptual features to the perceptual characteristics the user needs to e.g. gain situation awareness through the use of video as expressed in Section 5.2. From such relationship, we distinguish between between temporal and spatial domains of QoE.

Definition 10. QoE in the space domain refers to user's dissatisfaction due to lack in accuracy: artifacts in the video that may be caused by packet erasures, source coding effects like compression, and other factors affecting video recovery and degrading video quality.

Definition 11. QoE in the time domain refers to user's dissatisfaction due to persistent freezes in the video playback that do not allow to track dynamically changing situations.

5.3.3 QoE/QoS decomposition

Fig. 5.1 summarizes QoE framework adopted. Our analysis to perform the mapping into time and space domains, and how we propose to solve it is as follows.

Congestion affects QoE mainly in the time domain, inducing freezes in video playback. If congestion can be tracked at transport layer, rate adaptation to the network's available rate can performed and QoE in the time domain will be improved. Congestion can be detected online at the transport layer through end-to-end feedback using delay as congestion cue.

Erasures affect QoE in the space domain, inducing artifacts in video displayed. If at network layer channel coding is performed to recover from erasures, QoE in the space domain is improved.

By mapping congestions to QoE in the time domain, and erasures to QoE in the space domain, we are able to propose a decoupled solution for the joint problem affecting our scenario. Hence, we propose a two QoE driven optimizations, jointly operative but working separately, one for the time domain, the other for the space domain, to work at transport and network layers respectively.

A decoupled solution provides advantages in terms of flexibility of the design, as the formulation and performance evaluation of the two optimizations can be treated separately.

The possible price to pay in our decoupled approach is whether, when working jointly, the QoE optimizations are affecting one another. In Section 5.5 we will show that under reasonable assumptions, this cross-influence is minimal.

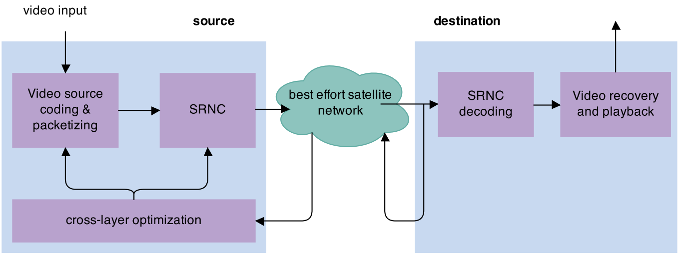


Figure 5.2: Scenario.

5.3.4 Selected metrics

With respect to QoS metrics, η and ϵ represent congestion and erasures respectively.

With respect to QoE, we have chosen low complexity metrics with a reasonable correlation to subjective QoE studies that can help us evaluate the performance of our solution in an experimental platform. From the metrics defined in Section 2.4.2, we will use QoE_T from (2–6) to measure QoE in the temporal domain, and QoE_{ST} from (2–5) for QoE in spatial domain.

Note that our framework and later solution is not dependent on the QoE metrics used. More sophisticated metrics as cited in Section (2.4.2) could be used for performance assessment.

5.3.5 Topology

We consider a point-to-point scenario where the underlying heterogeneous network topology can have several intermediate nodes. However, in this paper, for reliability, we consider channel coding only at the source node, but our system model can be extended for the case where joint channel-network coding can be used in the intermediate nodes of the network (which is out of the scope of this paper and is left out for future work).

Fig. 5.2 shows the overall block diagram of the source-destination topology and our proposed solution.

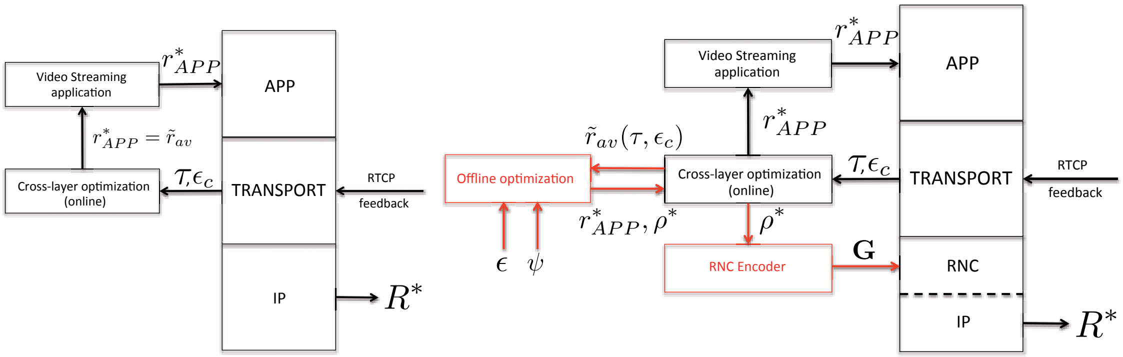
5.3.6 Cross-layer optimization

As Fig. 5.1 shows, we propose a decoupled cross-layer QoE-driven optimization framework consisting of two optimizations, one in the time domain and one in the space domain of QoE.

In the time domain, as shown in Fig. 5.3a, we use an online adaptation strategy that uses end-to-end feedback at transport layer to cope with congestion. Network delay τ and congestion induced erasures ϵ_c can be inferred from the feedback, and used to estimate $\tilde{r}_{av}$. The application layer rate r^*_{APP} to be used by the video streaming application, will be $r^*_{APP} = \tilde{r}_{av}$. As a result, the transmission rate at network layer R, is after optimization $R^* \approx r^*_{APP}$[2]

---

[2] we consider overhead due to layer encapsulation to be negligible when calculating the rates



(a) Cross-layer diagram for QoE-driven optimization in the time domain (online).

(b) Cross-layer optimized video streaming in the time domain (online) and space domain (offline).

Figure 5.3: Proposed cross-layer optimization framework.

Fig. 5.3b shows the integration of QoE optimizations in time and space domains.The available rate $\tilde{r}_{av}$ is estimated online by the optimization in the time domain. The optimization in the space domain determines the optimal code rate ρ^* for an erasure protection code using RNC encoding. As a result, the application layer rate is adapted to $r^*_{APP} = \rho^* \tilde{r}_{av}$, the transport layer packets are encoded using RNC at a sublayer of the network layer, and the IP sublayer transmits at a rate $R^* \approx \tilde{r}_{av}$. The optimization in the space domain can be performed offline, and look-up tables can be available online with the optimal values for a certain set of input values.

Note that the erasure correction coding is chosen to be at the network layer in order to enable the possibility of coding at intermediate nodes in future work. Network layer packets are accessible at intermediate nodes and hence coding at this layer would be more efficient in our model.

5.4 CROSS-LAYER OPTIMIZATIONS IN SPACE AND TIME

5.4.1 Optimization in time domain

The cross layer optimization in time is as proposed in Section 4.2. Integration within the cross-layer framework follows Figs. (5.3b) and (5.3a), such that $R^* \approx \tilde{r}_{av}$.

In order to maintain coherence of our model in time and avoid synchronization issues at the different layers, we assume that $T_{GoP} < T_{samp}$, such that the obtained application layer rate r^*_{APP} for the n^{th} GoP is invariant for the duration of the whole GoP.

5.4.2 Optimization in space domain

In our decoupling approach, we have identified erasures with artifacts in the video to be solved with an optimization in the space domain. In this Section, we present the formulation and implementation of such QoE driven optimization, as part of the cross-layer model in Fig. 5.3.

The objective of the optimization in the space domain is to optimize application rate r^*_{APP} and code rate ρ^*, in order to use SRNC to cope with erasures with maximized QoE of video.

Let us consider $\tilde{r}_{av}$ to be the available rate estimated using the algorithm in (4–5). In order to protect the video stream from network erasures, SNRC coding will be used with a certain allocated code rate ρ. A low value of ρ implies more erasure protection, at the expense of a lower rate for the application layer $(r_{APP} = \rho\tilde{r}_{av})$. Given that lower r_{APP} is a result of higher compression rates, QoE in the space domain is damaged with low values of ρ.

Hence, we propose in (5–1) to maximize QoE by maximizing r_{APP}, such that SRNC is used with an optimal code rate ρ^* that guarantees a residual erasure rate ψ.

$$\begin{aligned} r^*_{APP} &= \max r_{APP} \\ &\quad \text{s.t.} \quad r_{APP} \leqslant \tilde{r}_{av} \ \text{ and } \ \epsilon^{res}(\epsilon, q, r_{APP}, \tilde{r}_{av}) \leqslant \psi \end{aligned} \tag{5–1}$$

where $\epsilon^{res}(\epsilon, q, r_{APP}, \tilde{r}_{av})$ is the residual erasure rate of the SRNC code with field size q.

We target an offline solution to (5–1), in order to obtain the optimal values (r^*_{APP}, ρ^*) corresponding to all the possible estimated available rates $\tilde{r}_{av}$. A look-up table with these values is generated. As $\tilde{r}_{av}$ is time varying and estimated from the feedback, the look-up table is accessed online and optimal r^*_{APP} and ρ^* are obtained corresponding to $\tilde{r}_{av}$.

More details on the SRNC specifics can be followed in [12]. In systematic codes input data is embedded in the encoded output, reducing the decoding overhead at the receiver side.

Further, due to the inherent random structure of SRNC, progressive decoding is possible, improving packet recovery time as compared to RS codes, an advantage in long-delay scenarios.

5.5 PERFORMANCE EVALUATION I

5.5.1 Experimental setup

The setup consists of a point-to-point streaming connection. The receiver and sender applications are connected through an emulated network using the NetEM emulator.

5.5.1.1 *Setup*

Following Fig. 5.3b, we describe each block.

At application layer we use the state-of-the-art video codec VP8 [103]. At transport layer, we use the RTP/UDP protocol and a standard implementation of RTCP (RFC 3550) protocol for the feedback. At network layer, each transport layer packets is encapsulated into an IP packet.

The cross-layer modules work as follows. The online optimization has been implemented to output a rate control update of R^* with every new RTCP report, according to (4–7). The offline optimization uses a look-up table to output the optimal r^*_{APP} and code rate ρ^* values from the budget rate $\tilde{r}_{av}$.

Table 6: Parameters in experimental setup.

Experiments		1. QoE (time)	2. QoE (space)	3. Joint QoE
Video sequences		*pedestrian, foreman, coastguard*		
T*(streaming time)*		3min		
APP	N_{frame}	15		
	r_{frame}	15fps		
	$r_{APP}(t_1)$	500kbps	[100- 500kbps]	500kbps
Transport	pkt size l	1400B		
	T_{samp}	2s		
Network	q	-	256	256
	ψ	-	10^{-3} [110]	10^{-3} [110]
Network emulation	τ_D	250ms		
	ϵ	no	[0-15]%	[0-15]%
	r_{av}^{max}	500kbps	[100- 500kbps]	500kbps
	η	[100-50]%	0%	[100-50]%

We simulate SRNC coding by adding, for each GoP coming from the transport layer at rate r_{APP}, redundant (dummy) packets such that $R^* = r_{APP}/\rho^*$.

5.5.1.2 *Network emulation*

With respect to erasures, packets are erased at the random rate ϵ when no erasure protection is performed. When SRNC is used, packets are erased corresponding to the residual erasure probability of SRNC $\epsilon = \epsilon_{res}(\epsilon, q, r^*_{App}, \tilde{r}_{av})$.

Congestion events are emulated as a drop (step-like) in maximum available rate r_{av}^{max} that occurs halfway through one streaming session, at $T/2$. In practice, we use the traffic shaping in NetEm emulator to create the drops in r_{av}^{max}, such that $r_{av} = \eta \cdot r_{av}^{max}$.

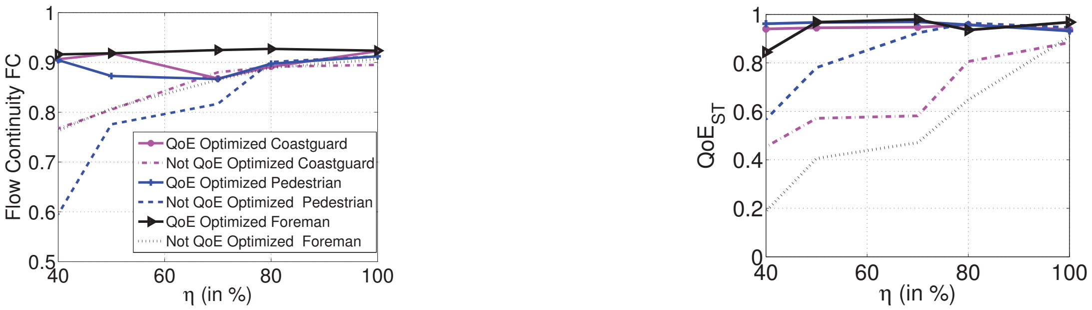
5.5.1.3 *Experiments*

Table 6 summarizes the values of the parameters used for the experiments.

Experiments with and without the space and time domain QoE optimizations are considered. Each experiment consists of one streaming session lasting T seconds. A large value of T (3 minutes) is such that it guarantees statistical significance with respect to erasure rates as well as spatio-temporal variations in the video.

For each experiment, a looped standard video sequence served as input source. Further, to each experiment correspond one value of ϵ and η. The ranges of values for ϵ are 0-15%, while for η, the range is from 100% to 50%.

The range of values considered for r_{av}, and r_{APP} correspond to realistic values for an application using a Mobile Satellite Service like the BGAN network. Such network offers roughly a maximum $r_{av}^{max} = 500kbps$ in a best effort configuration. The propagation delay τ_D corresponding to a GEO-stationary satellite network is also configured in NetEm. The value of ψ was chosen according to 3GPP specifications for real-time scenarios.



(a) Flow Continuity FC. **(b)** SSIM.

Figure 5.4: Evaluation of only QoE optimization in the time domain.

5.5.2 QoE Performance assessment

Application-layer information, both at media and bitstream level, are collected at receiver and sender in order to perform offline QoE assessment. The common problem of sent/received video misalignment affecting full reference spatio-temporal QoE assessment, is solved by using the frame concealment strategy. This means that a lost frame at the receiver is replaced in the sequence by the last frame received, making use of bitstream level data from actual frames sent and received. In addition, bitstream level data also provides frame play-out timestamps.

The two QoE metrics, QoE_{ST} and QoE_{T}, will be used to validate our results. QoE_{ST} and QoE_{T} correspond to the spatio-temporal metric and the temporal metric of flow continuity respectively, both of them described in Section 4.2.1.

Along with QoE_{ST},other spatio-temporal metrics [111] were tested with similar behavior.

5.5.3 Optimization in time domain

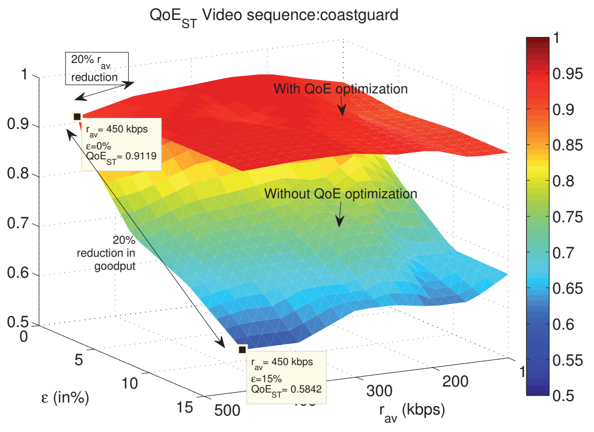
The purpose of this experiment is to evaluate the isolated performance of our optimization in the time domain. In order to assess this, we consider only degradations due to congestion. Further, we compare the performance to a solution that is unaware of such degradations. The results are summarized in Fig. 5.4 for videos *coastguard*, *pedestrian*, and *foreman*.

5.5.3.1 *Description of experiments:*

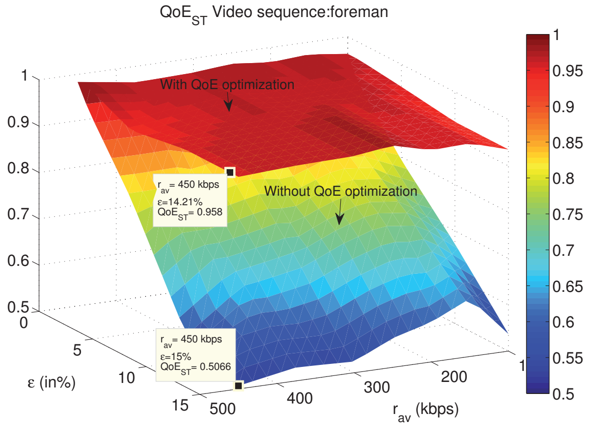
The parameters for this experiment correspond to Experiment (1) from the Table 6.

5.5.3.2 *Effect on* QoE_{T}

As can be observed in Fig. 5.4a, the flow continuity measured with QoE_{T}, is improved up to 50%. The highest advantage is achieved for lower values of degree of congestion η.



(a) Video Coastguard



(b) Video Foreman

Figure 5.5: QoE Optimization in the space domain. QoE_{ST}metric.

5.5.3.3 *Effect on* QoE_{ST}

The online optimization is able to avoid congestion events, hence packet losses due to congestion are minimized. As a consequence, the improvements in QoE are not only in time domain metrics but also in the space domain.

It can be observed in Fig. 5.4b that most significant improvements occur for congestion events with $\eta < 70\%$ in QoE_{ST}, with improvement of over 100%. For higher values of η, the metric also shows a gain from 4.5% to 50% using the QoE optimization in the time domain.

5.5.4 Optimization in space domain

In this experiment we compare the optimization in the space domain (where for each r_{av} an optimal ρ^* is obtained to use SRNC) to a non-optimized strategy with no erasure protection ($\rho = 1$).

5.5.4.1 *Description of experiments*

This case considers only degradations in the network due to erasures. We assume r_{av} is constant throughout the entire streaming session ($\eta = 100\%$, there is no congestion). We assume the transmission rate $R = r_{av}$, and $\tilde{r}_{av} = r_{av}$. For each experiment, there is a corresponding pair of values (ϵ, r_{av}). The parameters are set as in Table 6 for Experiment (2).

5.5.4.2 *Effect on* QoE_T

By optimizing the rate budget $\tilde{r}_{av}$ when using SRNC, we ensure that the redundancy added will not congest the network. QoE in the time domain is therefore not affected by the use of SRNC, as intended in our decoupling approach.

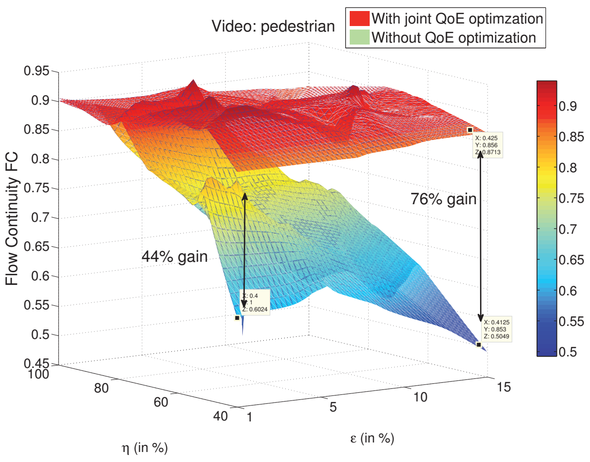
Notwithstanding, SRNC as any block erasure code adds delay at encoding/decoding. The systematic characteristic of SRNC as well as the possibility of performing progressive RNC decoding, significantly reduce the delays imposed by erasure protection. Therefore, we can assume, as a small price to pay, a reduced start-up delay in the video playback, such that SRNC does not affect QoE in the time domain, on no longer than the duration of a GoP, T_{GoP}.

5.5.4.3 *Effect on* QoE_{ST}

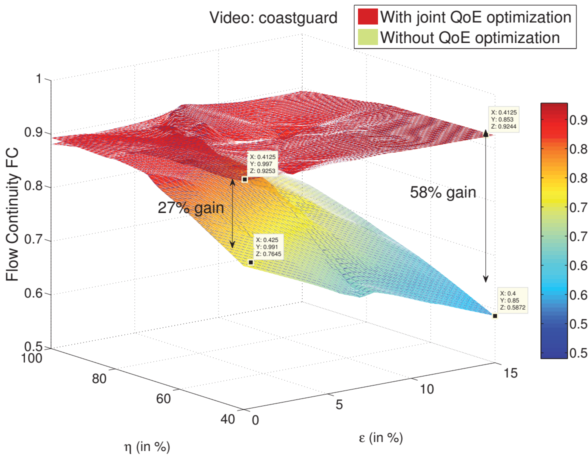
Fig. 5.5 shows the results for videos *foreman* and *coastguard*, in our three-dimensional analysis of QoE, where we have plotted QoE metrics vs. ϵ vs. η. Using the optimization in the space domain, the main advantage is achieved in scenarios with high available rate and high erasure rates, with up to a 38% improvement in QoE_{ST} metric compared to a non-optimized strategy. The higher advantage occurs for higher values of r_{av}.

Further, we show minimal effects of the decoupling approach on QoE_{ST} in the following example. Fig. 5.5 proves that space QoE is hardly sacrificed when lower application layer rates are used to either cope with congestion (QoE in the time domain) or with erasures (QoE in the space). In an erasure-less scenario ($r_{av} = r_{APP}$), a reduction of 20% in r_{APP}, has a degradation of under 1% in QoE, while a reduction in goodput due to erasures of $\epsilon = 20\%$ represent a 40% degradation in QoE.

Regarding video complexity, when comparing Fig. 5.5a for *coastguard* and Fig. 5.5b for *foreman*, the analysis is as follows. Using the QoE optimization, higher QoE values can be achieved for less complex types of videos, since these videos can obtain higher QoE with less amount of application layer rate r_{APP}. Higher complexity videos are more susceptible to the amount of r_{APP}, compared to low complexity videos. Hence, in higher complexity videos the difference in QoE at low r_{APP} and QoE at high r_{APP} is more noticeable than for lower complexity videos. Additionally, the gains of using the QoE optimization in the space domain for lower complexity videos are higher than for high complexity videos at lower values of r_{APP}.

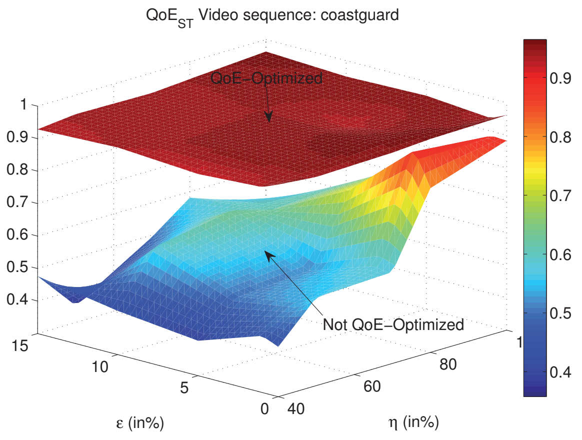


(a) Video: Pedestrian

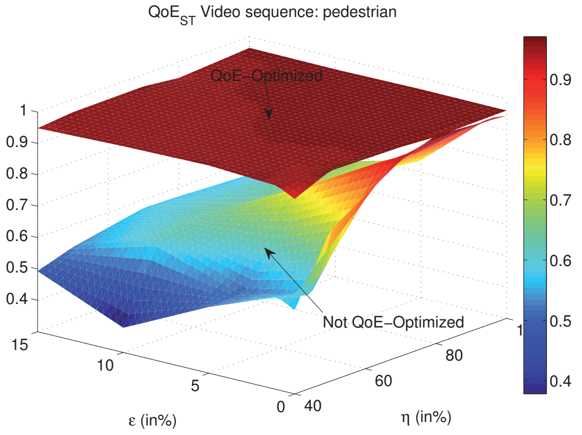


(b) Video: Coastguard

Figure 5.6: Flow Continuity, QoE_T for joint QoE in Space and Time domain.



(a) Video coastguard.



(b) Video Pedestrian.

Figure 5.7: QoE$_{ST}$ metrics for joint QoE in space and Time domain.

5.5.5 Joint optimization in time and space domains

The purpose of this experiment is to evaluate the performance of the joint optimization in the time and space domain according to the proposed model in Fig. 5.3b. Hence, we consider degradations both due to congestions as well as erasures.

5.5.5.1 *Description of experiments*

Congestion events and erasures are emulated as in the previous sections with the parameters of Table 6 for Experiment (3). For each experiment a different degree of drop in available rate is considered, together with an erasure rate ϵ.

We compare the joint optimization to a solution unaware of network dynamics, where application layer is blind to the network dynamics, the transport layer is not performing any congestion control, and there is no protection against erasures.

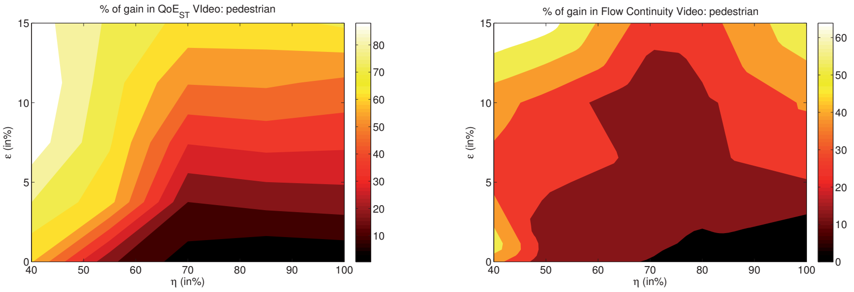
5.5.5.2 *Effect on* QoE$_T$

The three-dimensional QoE plots in Fig. 5.6 show that, for all cases of congestion and erasures tested, the values of flow continuity metric are all above 0.9 when using both optimizations . This represents that more than 90% of the time, the user is not experiencing freezes in the video playback.

5.5.5.3 *Effect on* QoE_{ST}

The complete solution achieves in overall a planar surface in QoE_{ST}, as shown in Fig. 5.7. This means that, regardless of both erasures and congestion affecting QoE, the combination of the online and offline strategies is able to deliver a smooth performance.

Moreover, for all cases of congestion and erasures tested, the values of QoE_{ST} metric are all above 0.9 when using the complete QoE framework, guaranteeing very small variations of quality over time.



(a) gains in QoE_{ST}.

(b) gain in Flow Continuity.

Figure 5.8: Contour plots of gains compared to no QoE solution. VIdeo: Pedestrian..

Fig. 5.8 shows that the gains obtained from using the complete framework compared to a non-QoE optimized approach. With respect to Flow Continuity, the gains are of up to 60% in Flow Continuity.

With respect to the QoE_{ST} metric, gains are of up to 80%. The cases with higher improvement correspond to higher erasure rates and higher degree of congestion η. The gradient of the gains in QoE_{ST} for higher values of η is only dependent on the increase in erasure rates, while for lower values of η, the gain increases jointly as η and ϵ increase.

These results represent a high QoE in space domain with additional smooth QoE performance throughout an entire streaming session, a characteristic highly valued by end-users, This behavior was observed with all video sequences tested.

5.6 PERCEPTUAL SEMANTICS MODEL

The motivation of the work presented here is the possibility of enhancing user's satisfaction by targetting the specific time/space perceptual features he identifies as relevant in his application scenario. This is particularly relevant in video services that go beyond enternainment fullfillment.

We consider band-limited, wireless best effort networks, as possible means of communications for point-to-point live video streaming . Such networks pose a number of constraints affecting Quality of Service (QoS), namely, congestion and erasures. The topology envisioned is that of live user-generated content being upstreamed e.g. a

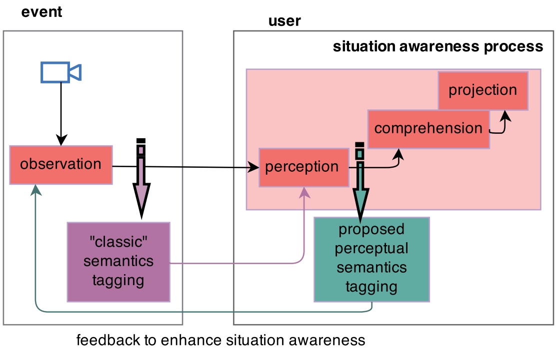


Figure 5.9: Perceptual semantics vs "classic semantics".

user in an emergency location upstreaming to proper decision-makers in a disaster management operation.

While the transmission framework proposed in Section 5.3.6 mitigates the effects of the network constraints, we propose an additional dimension to target specific user demands, e.g. in scenarios using video for other purposes. We propose to improve the user's satisfaction beyond standard improvements by means of perceptual semantics.

In this section we derive the model for perceptual semantics and how we propose to perform semantic tagging.

5.6.1 Spatial/temporal decoupling for situation awareness

We present an approach inspired by situation awareness and its three level model description as defined in Section 2.4.3, which highlights perception as a first level process in attaining a particular target of comprehension and projection of situations.

We focus on the spatial and temporal advantages of video as a source of information that can help in e.g. emergency operations. Firstly, the possibility of capturing dynamic scenes improves the assessment of temporary evolving events. Secondly, video can provide visual spatial accurate accounts of an ongoing situation [86].

If the temporal and spatial perceptual characteristics of the video satisfy the situation-specific user resolution, then the user satisfaction will be fulfilled, and further higher level cognitive processes will be benefited.

5.6.2 Semantic tagging

Based on the spatio/temporal identification of perceptual features, our proposal is to utilize the end-user's (analyst) perception, to do semantic tagging that enables an enhancement of the received video stream signal tailored to the user's demand.

Semantic tagging is hence performed to describe perceptual features in the video and as such represents more complex abstractions of a viewed scene. In comparison, classic semantics tagging would focus on unprocessed sensorial observations [107]. The difference between both approaches in semantics is shown in Fig. 5.9.

In scenarios where perception is not achieved by artificial intelligence, it is the human analysis that will interpret the sensory information (i.e. perceiving). Hence, the

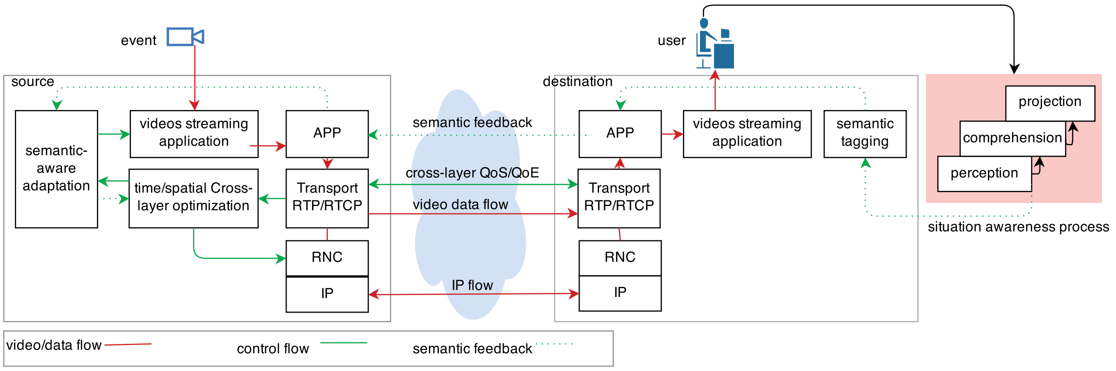


Figure 5.10: Block diagram proposed cross-layer framework and APP-to-APP perceptual semantics loop.

semantic tagging is performed by the user, as he is ultimately the one perceiving and foreseeing what might be of interest in the video.

We propose a tagging that would indicate the temporal/spatial predominance according to the level of perception of the user. A tag indicating predominance of temporal features, means the user is perceiving a situation that demands more attention to the dynamics of the scene (e.g. rapid movements). On the other hand, a predominance of spatial features indicate moments of less movement but densely overloaded frames that requires more detail to identify features.

5.7 INTEGRATION OF PERCEPTUAL SEMANTICS TO VIDEO ADAPTATION

We propose to integrate the perceptual semantics with video adaptation in order to provide to the user a required perceptual level. Further, we propose to map the tags to actions such that the specific perceptual features are enhanced.

Following, we describe how our perceptual semantics model can be mapped to video coding characteristics and propose an algorithm to meet the end-user's demands. Further, we comment on protocol aspects in the implementation and propose an integrated framework with an adaptive video solution.

We highlight that this is a first approximation to integrate perceptual semantics to video adaptation in this promising research direction. Future lines of work on this matter are discussed in Chapter (6).

5.7.1 Mapping

We focus on using our proposed perceptual semantics for enhancement at source coding level. At codec level, in single layer or scalable layer video encoding of state-of-the-art codecs, three types of resolution are defined, namely temporal (frame rate), amplitude (quantization step), and spatial (frame size).

We map enhancement of temporal features to higher frame rates, and predominance of spatial features to higher spatial and amplitude frame resolution. In this way, dynamics of the scene can be more closely followed (temporal preference) and details of a scene can be better identified (spatial preference). The mapping is intuitive and relies on the intrinsic architecture of video codecs currently in use in a non-intrusive manner, to facilitate the video communications. Finally, we show an example architecture within ICN networking of a typical emergency scenario.

5.7.2 Algorithm

We propose to map the perceptual semantics to a system quantified with the variable $\alpha \in [0, 1]$. $\alpha = 0$ and $\alpha = 1$ express full preference of the spatial and temporal perceptual features, respectively. Intermediate values of α represent weighed combinations of spatial and temporal preferences.

We denote the feasible set of finite values of frame rate, as $F_T(r_{APP})$, while $F_S(r_{APP})$ is the feasible set for the spatial factors, both a function of video coding rate r_{APP}. Note that higher frame rates and frame sizes are possible to attain with higher r_{APP} [112], hence the feasible sets $F_S(r_{APP})$ and $F_T(r_{APP})$ corresponding to higher values of r_{APP} will contain more number of possible values that can be chosen from. For example, in the case of scalable video coding, if temporal dyadic scalability is performed, the available values of frame rate contained in $F_T(r_{APP})$ would be the base layer frame rate and the frame rates from enhancement layers that would add up to i.e. a full 30Hz frame rate if r_{APP} is sufficient: $F_T(r_{APP}) = \{3.75\text{Hz}, 7.5\text{Hz}, 15\text{Hz}, 30\text{Hz}\}$.

In order to choose the appropriate value of frame rate and resolution according to our mapping of perceptual semantics, we formulate the following optimization function:

$$\begin{aligned} (r^*_{fr}, s^*_{fr}) &= \max \alpha\bar{r}_{fr} + (1-\alpha)\bar{s}_{fr} \\ \text{s.t.} &\quad r_{fr} \in F_T(r_{APP}) \text{ and } s_{fr} \in F_S(r_{APP}) \end{aligned} \tag{5–2}$$

where $\bar{r}_{fr} = {}^{r_{fr}}/_{r^{max}_{fr}}$, and $\bar{s}_{fr} = {}^{s_{fr}}/_{s^{max}_{fr}}$ are the normalized values of frame rate r_{fr} and spatial/amplitude resolution s_{fr} with respect to maximum available values set for the application.

Note that the optimization in (5–2) can be applied to single layer video coding or scalable video coding.

5.7.3 Implementation and compliance with standards

We assume an underlying standard cross-layer framework, which provides the application layer rate r_{APP} that can be used by the codec (such that the video coding rate equals the application layer rate), for an on-the-fly adaptive video subject to network constraints.

The cross-layer optimization has been designed such that it copes with the network impairments that directly affect negatively the spatial/temporal aspects of video and is therefore QoE-driven. This cross-layer optimization associates congestion with temporal impairments in video playback such as freezes. In addition, it associates erasures with artifacts degrading video quality. Further, we keep in mind that higher video quality is achieved with higher video codec rate, r_{APP}.

The above assumption is relevant in the design as it is a guarantee that the network degradations are not an issue when enhancing the temporal and spatial features necessary for enhanced perception. We will further discuss this with the numerical results in Section 5.8.

Fig. 5.10 shows the block diagram of the proposed framework, where the cross layer optimization is integrated with the perceptual semantics loop. The video streaming application uses a state-of-the-art codec such that the frame rate, frame size and codec rate can be configured on-the-fly, either as a single layer or a scalable layers.

In order to facilitate the perceptual semantics role, we use a return path to send via feedback the tags chosen by the user according to the perceptual semantics.

Following the trends in current network architectures, we propose to use semantic web protocols to enable the APP-to-APP cross talk of the semantic tagging [109]. At the transport layer, the application-specific information can be encapsulated into RTCP feedback packets compliant with the extended reports defined in RFC4585. The cross layer optimization is handling feedback with the standard RTP/RTCP protocol. Note we have assumed Forward erasure protection being performed at network layer, in particular using random linear network coding.

Considering future network architecture, our framework complies with a semantic information-based network [58]. Moreover, our scenario would be suitable to information-centric networks' (ICN) architectures such as the publish/subscribe for live streaming [59]. ICN could allow for future enhancements to the perceptual semantics as proposed in this paper. In particular, our approach is coherent to the receiver-driven nature of ICN. Further, caching, one of the appealing attributes of ICN in data delivery, could enable more actions at intermediate nodes concerning the incoming video stream. The philosophy of ICN by which content information is available to network/forwarding layers will allow the semantic loop we have created to trigger further actions at these intermediate nodes, such as adaptive network coding to enhance last-mile network reliability.

Fig. 5.11 shows the topology of our framework mapped to publish/subscribe architecture within ICN. This example shows a typical emergency application over a satellite access network, whose gateway can be mapped to the *rendevouz* (RN) and *topology manager* (TM) nodes. The publisher (operator in the ground during an emergency situation) announces that it has a publication available to the RN node. The subscriber (end-user/decision maker) issues a subscription, as he is interested in obtaining live feed of the current on-going events of the emergency. The RN and TM nodes find the publisher and resolve the the publisher/subscriber path. The subscriber can issue petitions or unsubscribe, and in our framework, issue perceptual semantics tagging, which the publisher will receive through the RN nodes.

5.8 PERFORMANCE EVALUATION II

We simulate a realistic scenario typical of emergency operations, where mobile satellite services are used to upstream live video from field, to proper decision makers remotely located.

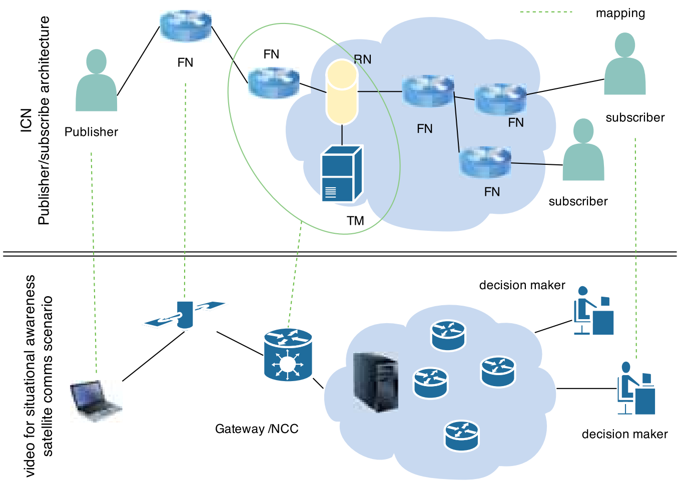


Figure 5.11: Publish/Subscribe architecture suitable for our proposed perceptual semantics framework.

r_{APP} (in kbps)	Feasible Set F_T	Feasible set F_S
$r_{APP} \leqslant 64$	{3.75, 7.5,10,15}	{QCIF}
$64 < r_{APP} \leqslant 192$	{3.75, 7.5,10,15}	{QCIF,CIF}
$192 < r_{APP} \leqslant 384$	{3.75, 7.5,10,15}	{CIF,QCIF}
$384 < r_{APP} \leqslant 500$	{3.75, 7.5,10,15,30}	{QCIF,CIF,640x360}

Table 7: Feasible sets considered for simulation.

5.8.1 Setup

We use a simulation system that allows to test the proposed framework shown in Fig. 5.10. The video streaming application is simulated by generating packets of size l encoded at a rate r_{APP} and frame rate r_{frame}.

5.8.1.1 *Cross-layer optimization setup for congestion and erasures*

This block receives as inputs the feedback from the receiver on current network conditions, and outputs the rate r_{APP} that the video streaming application is allowed to use, and the code rate ρ to be used for erasure correction, such that the transmission rate is $R = {}^{r_{APP}}/_{\rho}$.

The transmission rate is online optimized through a QoE delay-driven optimization at the sender side that uses receiver feedback, as the one proposed in [9] and described in Chapter 4. The resulting discrete rate control update is given by (5–3)

$$R(t_{k+1}) = R(t_k) + f(\tau(t_k - \tau_D)) \tag{5–3}$$

where $f(\cdot)$ is a function of the delay τ measured at time $t_k - \tau_D$, a delayed value due to the propagation delay τ_D in the feedback loop. . Updates on network measurements are received every $T_{samp} = t_{k+1} - t_k$ seconds.

Further, our additional novelty to cope with erasures, is the use of adaptive network coding with systematic random linear network coding, (SRNC), in coherence with Section 5.4.2. We use SRNC due to similar performance to optimal forward erasure correction codes like Reed-Solomon [52], but higher flexibility and compliance with future network-coded networks. For a rate budget given by R in (5–3), the code rate $\rho = r_{APP}/R$, chosen for SRNC is maximized such that the performance meets a target residual erasure rate given the current erasure rate ϵ of the network. Hence the application layer rate r_{APP} is maximized.

5.8.1.2 *Network simulation*

We simulate a network as a FIFO finite queue of available rate r_{av} with erasure rate ϵ. Simulated packets are transmitted at the obtained rate R.

SRNC uses the allocated code rate ρ to meet the complete budget rate R, such that its performance meets the residual erasure rate ϵ^{res}.

Congestion events are simulated as a drop (step-like) in maximum available rate r_{av}^{max} to r_{av}^{min} that occurs halfway through one streaming session, at $T/2$ such that $\eta = \frac{r_{av}^{max} - r_{av}^{min}}{r_{av}^{max}}$, with $\eta \in (0, 1]$. (Higher η means higher congestion). Each simulation, corresponding to one streaming session, lasts 300s, one corresponding value of η and ϵ.

The values used correspond to a realistic satellite network commonly used in emergency operation, operating in the L-band offering up to 500kbps uplink in best effort mode.

5.8.1.3 *Perceptual semantics*

We model the user's semantic tagging from temporal/spatial features with the parameter α. α may vary over time throughout one single streaming session, such that the sender is receiving feedback of this changes and will adapt to them using i.e. (5–2). We assume these tags are changed by the user with a period of at least 10s. Three cases are considered for variation of semantic tagging, namely, TAG_T: only temporal tagging for the entire session, TAG_S: only spatial tagging, TAG_{TS}: alternating tags, each of 10 seconds.

Table 7 summarizes the feasible sets for values of frame rate dependent on r_{APP}, in order to solve the algorithm in (5–2). The values chosen correspond to typical feasible combinations in current state-of-the art codecs.

5.8.2 Metrics

The following metrics relate to the effects of the network constraints on the performance in terms of Quality of Experience.

1. QoE_A. This metric is related to degradations due to erasures in the network, that cause artifacts in the image: $QoE_A = 1 - \bar{p}$, where $\bar{p}$ is the average packet loss rate at the receiver. $QoE_A \in [0, 1]$.

2. QoE_F. This metric is related to degradation due to congestion, that cause freezes in video playback. $QoE_F = 1 - \bar{f}$ where $\bar{f}$ is the probability of freezes occurring in the playback. A freeze is the event where the time elapsing between two consecutive frames displayed exceeds a tolerated threshold. $QoE_F \in [0, 1]$.

3. $\hat{\alpha}$ and Δ_α: These metrics are related to the performance of the adaptation through perceptual semantics. We measure the value achieved by the algorithm

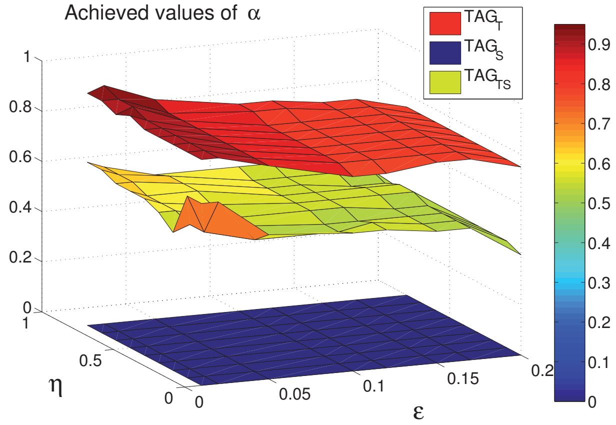


Figure 5.12: Achieved values of α vs. η vs. ϵ, using cross-layer optimization.

as $\hat{\alpha}$, and the mean absolute error with respect to the user's demanded α, as $\Delta_\alpha = |\hat{\alpha} - \alpha|$.

4. Ω: Combined metric to measure tradeoffs of using perceptual semantics with and without cross-layer optimization. It is defined as:

$$\Omega = w_1 \cdot \mathrm{QoE}_A + w_2 \cdot \mathrm{QoE}_F + w_3 \cdot (1 - \Delta_\alpha)$$

with $w_1 + w_2 + w_3 = 1$. $\Omega \in [0, 1]$. The best performance, i.e. $\Omega = 1$, occurs when no losses degrade the video ($\mathrm{QoE}_A \to 1$), freezes in playback are minimal ($\mathrm{QoE}_F \to 1$) and the perceptual semantic adaptation matches the one requested by the user (Δ_α).

5.8.3 Results

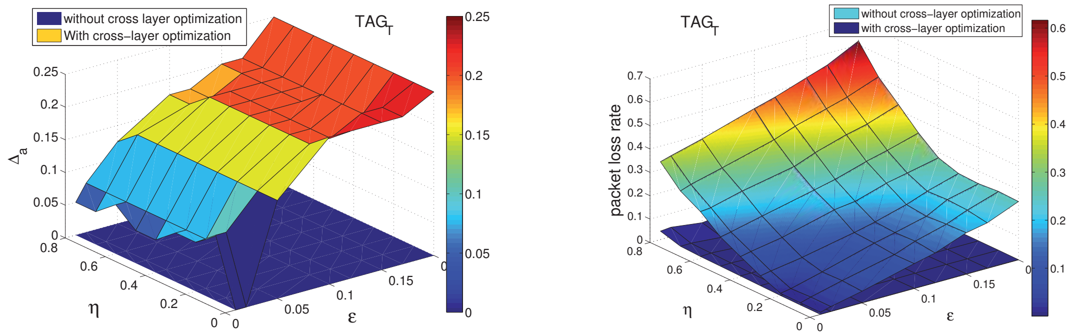
5.8.3.1 *Perceptual semantics with and without cross layer optimization*

Fig. 5.12 shows the performance of the perceptual semantics together with the cross layer optimization, as a function of congestion drops, η, and erasures ϵ. Each surface corresponds to one of the three cases of time varying semantic tagging. TAG_T achieves high values of α close to the tagged from the user, representative of preference on temporal features, while TAG_{TS} offers an intermediate values, corresponding to the alternating tags.

In order to observe the combined effects of the adaptation through perceptual semantics with an underlying cross-layer optimization, we observe the individual metrics. The comparison is made between using cross-layer optimization to cope with the network constraints, or no use of it.

Fig. 5.13a shows the value of Δ_α as a function of η and ϵ. In order to achieve high QoS and QoE with the cross layer optimization, the application layer rate r_{APP} is sacrificed, as more rate is needed to protect from network erasures. Hence, the feasible set of frame rates is reduced, and the obtained α can not achieve the highest expected value. This can be observed with higher values of Δ_α as ϵ increases.

Nevertheless, the cross layer optimization is guaranteeing very low packet losses, as Fig. 5.13b shows, which translates into minimal artifacts in the video. Hence,



(a) Δ_α for case 1.

(b) $\bar{p}$

Figure 5.13: Δ_α and $\bar{p}$ for TAG_T.

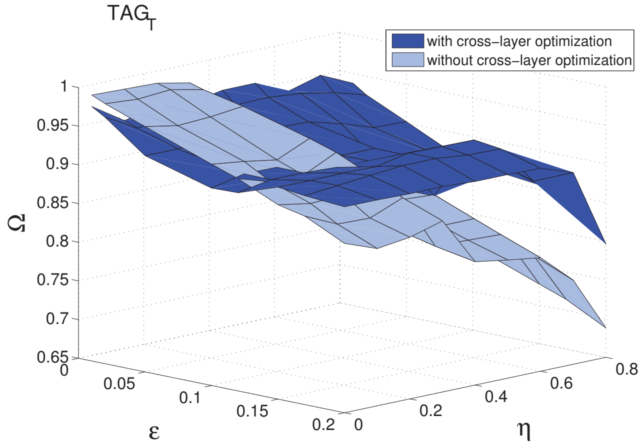


Figure 5.14: Ω metric for TAG_T.

while seemingly $\Delta_\alpha$is not as low as expected, the user is guaranteed a seamless video playback.

Fig. 5.14 shows the combined metric Ω, where the above trade-off result into higher performance when using cross-layer optimization in combination with the perceptual semantics loop, especially for highly degraded networks.

5.8.3.2 *Time varying perceptual semantics tagging*

We analyze the effects of time-varying perceptual tagging, representing a realistic case where the user identifies different situations that demand attention towards temporal or spatial features. These variations are represented as alternations of temporal and spatial tagging. Fig. 5.15 shows the performance in terms of the combined metric Ω.

In addition to achieving the expected α demanded through the semantic tagging, the performance is above 80% regardless of the degradations of the network, thanks to the cross-layer optimization. The performance is highly degraded due to congestion as well as erasures when no cross-layer optimization is used, with performance dropping to 40%.

in conclusion, Fig. 5.15 shows that the cross layer optimization preserves the perceptual semantics.

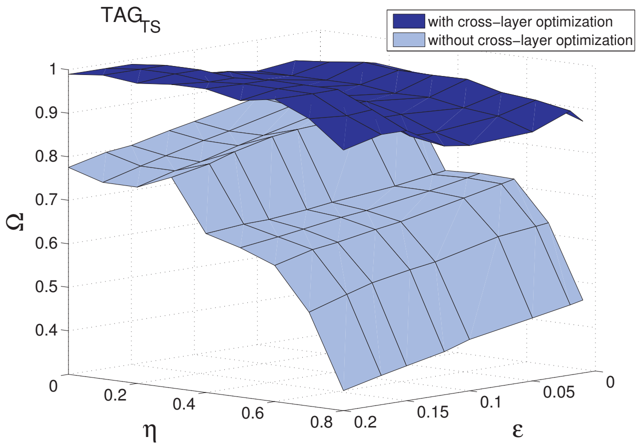


Figure 5.15: Ω for time-varying semantic tagging TAG_{TS}.

5.9 CONCLUSIONS

In this chapter we have presented the culmination of the work from the preceding chapters by proposing a complete model and solution to deliver point-to-point video services in an information-centric network. This model meets the main objective of this thesis, Objective (3) by proposing a novel QoS/QoE interplay framework that allows for an optimal adaptation to erasures and congestions in heterogeneous networks and presenting and alternative to enhance user's satisfaction by means of perceptual semantics in an information-centric setting. The framework is inspired by situation awareness applications, but it is not limited to it and opens a promising door towards the analysis of QoE for video services beyond recreational purposes.

First, we have focused on the provision of such services in constraint networks scenarios. We used a QoE framework to decouple the problems inherent to the scenario, identifying congestion with freezes in the time domain and packet erasures with artifacts in the space domain, both degrading the QoE of video. Our decoupled approach facilitates the design to optimize QoE both in the time and in the space domain, to obtain a feasible solution for dynamic adaptive streaming tailored to the scenario's needs As a consequence of decoupling and tackling these two problems separately, we have a time/space graphical analysis with varying network conditions in form of congestion and erasures. The tradeoffs in this decoupling paradigm are minimal, given the design choices for minimal delay effects by the erasure correction mechanism, and the delay-driven nature of the optimization in the time domain that prevents erasures due to congestions to happen.

Second, we have presented in this paper a framework where we introduce perceptual semantics for video adaptation. Perceptual semantics are used to acknowledge the user's particular demand in different video services applications, e.g. where special attention is required when using video as means to perceive, comprehend and project ongoing situations. We have presented a novel model for perceptual semantics, based upon these demands, and propose a framework to be integrated into a video adaptive solution, where there is no artificial intelligence behind interpreting sensory information from the video. We discussed how to practically implement perceptual semantics into an adaptive loop that works with an underlying cross-layer optimization in charge of coping with network constraints typical of best effort wireless scenarios. Further, we have shown an adaptive algorithm that translates the perceptual semantics into temporal and spatial resolutions ad video codec level.

Finally, our framework is contextualized for the upcoming trends in information-centric-networking. The perceptual semantics proposal could be further enhanced within the information-centric networking architectures, thus offering future opportunities such us active roles of intermediate nodes that are aware of content information.

Our experimental results show the benefits of this decoupled approach in terms of objective QoE metrics. The gains obtained from using the complete solution proposed in the thesis are of up to 80% in QoE_ST and 60% in QoE_T, occurring when congestion and erasure rate are the highest. In cases with no erasures and highest congestion, the gains can be of up to o 40%. This gains are obtained when comparing the solution with not performing any no cross-layer signaling between the APP layer and the lower layers, and where there is no active transport layer mechanism for congestion or reliability. Moreover, we are able to achieve overall homogenous high QoE performance , regardless of both erasures and congestion degrading the network.

Moreover, simulation results show how the perceptual semantic tagging achieves the expected user demands while the underlying cross-layer optimization preserves such performance.

5.10 PUBLICATIONS

1. M. A. Pimentel-Niño, P. Saxena and M. A. Vázquez-Castro. “Multimedia Delivery for Situation Awareness Provision over Satellite”. Submitted to IEEE Transactions on Wireless Communications, April 2014.
2. M. A Pimentel-Niño, M. A. Vázquez-Castro, and I. Hernáez-Corres, "Perceptual Semantics for Video in Situation Awareness", accepted to ICSNC 2014, Nice France. June 2014.

6 CONCLUSIONS AND FUTURE DIRECTIONS

This thesis was motivated by three main challenges: 1) the heterogeneity of converged networks leading time varying constraint networks with limited QoS guarantees and uneven performance, 2) how this network configurations affect video content delivery and the user's satisfaction, and 3) the increasing user demands for video services and its use expanding to applications beyond recreational purposes user expects different requirements to be met. A particular application that included all of this challenges was the provision of situation awareness by means of video services.

Hence the objectives of this thesis were set accordingly, and have been successfully accomplished as each corresponding chapter of this thesis details. Moreover, a number of scientific publications and contributions to technical reports support our work.

The active involvement in user driven practical projects shaped this dissertation. Thus, this thesis was granted a well-founded background concerning the requirements of each scenario, user's needs and constraints. These foundations paved the way towards finding not only theoretically sound and novel models but also feasible solutions suitable for implementation.

6.1 CONCLUSIONS

To conclude we summarize the contributions with respect to the three objectives originally set and discuss future lines of work.

6.1.1 Objective 1: Propose a framework of heteregeneous networking

In Chapter 3 we presented our contribution where we propose to model heterogenous networks with a holistic approach. The methodology of this holistic system design is based on two novel concepts. First, to provide a framework by which heterogeneous network instances can be modeled to guarantee generality and robustness. Second, to uniquely characterize the network instances via their min-cut. The framework meets the coherence with information-centric networking as expected in Objective (1)

The strength of this framework is its usage as an underlying system model that can guarantee seamless content delivery regardless of the network instance. The latter is possible by formulating a general cross-layer optimization for content delivery, coherent to information-centric networking philosophy. To show the extent of the holistic approach we show examples of holistically-modeled realistic network instances common to emergency operations in disaster management.

6.1.2 Objective 2: Propose a user-centric performance framework for adaptive video transmission

We propose a QoE-driven adaptive video framework, based on a cross-layer optimization formulation in Chapter 4. Such formulation uses QoE correlation models with QoS, and addresses heterogeneous networks in best effort mode, in particular in wireless mediums and with long round trip times, as required in Objective (2). The derived adaptive video algorithm for time-variant networks is delay-driven, hence not affected by network erasures that are not congestion-induced. The design contemplates the constraints of long-delayed networks and the challenges of establishing a feedback loop to enable the network adaptability.

The frameworks is evaluated systematically, first in an emulation environment, second with a full system implementation. QoE metrics show the advantage of using the user centric framework in time-varying network scenarios.

Moreover, we study the feasibility of jointly tackling network erasures due to the wireless medium with the proposed QoE-driven adaptive video framework for congestions. A content-aware erasure protection mechanism based on network coding is tested. The experimental results show the potential in terms of improvements in the QoE metrics.

6.1.3 Objective 3: Develop a complete model and solution for video adaptation over heterogeneous networks

Chapter 5 presents the main contribution of this thesis with a complete model for video services in an information-centric networking setting complaint with Objective (3). It builds up from the contributions Objective (1) and (2) to propose a model that includes a novel QoE/QoS framework for video with a mission specific angle inspired by situation awareness scenarios. The problem of combined erasures and congestion in best effort network setting is decoupled to match the specific degrading effects on the video. This allows for two separate QoE driven optimization approaches in time (freezes) and space (artifacts) domain. The complete solution offers a feasible dynamic streaming adaptation that suits constraint heterogenous networks such as satellite. The performance is evaluated with QoE metrics using a novel three-dimensional analysis with respect to both erasures and congestion.

Moreover, we propose a novel alternative to enhance user satisfaction by adding perceptual semantics. This approach is compatible with the time/spatial decoupling to tackle network constraints, and further proposes, within the context of information-centric networking, an un explored take on semantics that intends to reflect on the perceptual needs of the end user.

6.2 FUTURE LINES OF WORK

Following, we discuss future directions to be considered in relation to the contributions of this thesis. We list them in relation to each of the main objectives.

6.2.1 Framework for heterogeneous networking

The holistic approach to heterogenous networking we have proposed could include challenging aspects in heterogeneous networks such as geographical location, time of day, as the min-cut characterization could be affected by this aspects. This aspect could be useful especially for network instances in emergency scenarios, which as noted in Chapter 3.2 offer diverse performances.

Further, in coherence to the IoT trend, our modular approach could be further refined with respect to the different devices and requirements that could affect the network.

Finally, while this dissertation considers a more evolutionary approach towards information-centric networking, it could be worth studying the implications of using clean-slate ICN architectures, and the role of of content-aware nodes in the holistic design approach.

6.2.2 Framework for user-centric for adaptive video transmission

With regards to adaptive video, our framework could consider evolutions that exploit different network topologies, to increase robustness towards congestion events or erasures. Some of the possible directions could be multipath streaming and hop-by-hop congestion control for adaptive video that could be of interest within the clean-slate ICN approaches.

Moreover, in the ICN paradigm, exploitation of in-network caching, would push the application of network coding at intermediate nodes for reliability.

6.2.3 Complete model and solution for video adaptation over heterogeneous networks

6.2.3.1 *Extensions to the QoS/QoE framework*

As future line of work, we could extend our time/spatial QoS/QoE decoupling model to include other QoE influential factors such as context. With the enhancement of context-awareness to this model we could map it to the context-aware networking, and identify influential factor at QoE level that are affected by e.g. geographical location, end device, time.

Moreover, our QoE three-dimensional analysis would then be extended to the additional dimension given by the context.

Considerations to be further studied are for example how system influential factors like congestion could be decoupled from context factors like time of day or location.

6.2.3.2 *Extensions to perceptual semantics*

Extension to perceptual semantics can be classified according to interactions with different layers of the protocol stack.

At application layer, closer interaction to video coding techniques and image processing could be worth exploiting for feature extraction. Perceptual semantics would aid in semantically describing the perceptual aspects the end user wishes to be enhanced from ongoing situations. Such approach would be of use in augmented reality or 3D videos.

Further, traditional ways of measuring QoE in video have limited correlation to the outcome of adaptive video using perceptual semantics, in particular with close interactions with the application layer. Video displayed is no longer homogeneously important to the user, and depending on the stress of perceptual feature he/she needs, his notion of QoE will be tainted. Hence, alternative views on quantifying QoE for video in different scenarios, not limited to entertainment, are worth pursuing .

At lower layers of the protocol stack future directions of research could consider the more active roles of intermediate nodes in network architectures, as in content-awareness in information-centric networking. Perceptual semantics could exploit these network architectures to disseminate demands from the user that could be solved at nodes closer to him/her (when i.e. caching is used), or to provide specialized robustness at specific points of the network with strategies such as network coding.

REFERENCES

[1] "Real-Time Communication in WEB-browsers (Active WG)." [Online]. Available: http://tools.ietf.org/wg/rtcweb/charters

[2] T. Stockhammer, "Dynamic adaptive streaming over http –: Standards and design principles," in *Proceedings of the Second Annual ACM Conference on Multimedia Systems*, ser. MMSys '11. New York, NY, USA: ACM, 2011, pp. 133–144. [Online]. Available: http://doi.acm.org/10.1145/1943552.1943572

[3] FP7 Project GEO-PICTURES. http://www.geo-pictures.eu. [Online]. Available: http://www.geo-pictures.eu

[4] BEATLES- BGAN Enhanced Alphasat Technology for L-band Extended Spectrum. European Space Agency. [Online]. Available: http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=31865

[5] MASERATTI- Mejora de la atención sanitaria en entornos rurales mediante aplicaciones de telemedicina sobre tecnologías inalámbricas . CYTED, Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo. [Online]. Available: http://www.cyted.org/cyted_investigacion/detalle_accion.php?un=5487315b1286f907165907aa8fc96619

[6] M. A. Vázquez-Castro, M. A. Pimentel-Niño, and R. Alegre-Godoy, "Las redes de comunicación basadas en satélite," in *Tecnologías de la Información y las Comunicaciones para zonas rurales: Aplicación a la atención de salud en países en desarrollo*, A. Rendón Gallón, P. J. Ludeña González, and A. Martínez Fernández, Eds. CYTED. Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo, 2012.

[7] M. A. Pimentel-Niño and M. A. Vázquez-Castro, "Cross layer content delivery optimization for holistic network design in disaster preparedness and recovery scenarios," in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, july 2011.

[8] M. A. Pimentel-Niño, M. A. Vázquez-Castro, and H. Skinnemoen, "Optimized ASMIRA Advanced QoE Video Streaming for Mobile Satellite Communications Systems," in *30th AIAA International Communications Satellite Systems Conference*, 2012.

[9] M. A. Pimentel-Niño, P. Saxena, and M. A. Vázquez-Castro, "QoE Driven Adaptive Video with Overlapping Network Coding for Best Effort Erasure Satellite Links," in *31st AIAA International Communications Satellite Systems Conference*, Florence,Italy, 2013.

[10] S. Gupta, M. A. Pimentel-Niño, and M. A. Vázquez-Castro, "Joint network coded cross-layer optimized video streaming over relay satellte channel," in *3rd*

International Conference on Wireless Communications and Mobile Computing (MIC-WCMC), June 2013.

[11] M. A. Pimentel-Niño, M. A. Vázquez-Castro, and I. Hernáez-Corres, "Perceptual semantics for video adaptation," in *accepted to ICSNC14*, Nice, France, October 2014.

[12] M. A. Pimentel-Niño, P. Saxena, and M. A. Vázquez-Castro, "Multimedia delivery for situationawareness provision over satellie," *Submitted to IEEE Transactions on Wireless Communications*, April 2014.

[13] X. Sanchez-Loro, J. Paradells, J. Casademont, and J. L. Ferrer, "Optimizing the delivery chain in heterogenous networks," in *Wired/Wireless Internet Communications*, ser. Lecture Notes in Computer Science, X. Masip-Bruin, D. Verchere, V. Tsaoussidis, and M. Yannuzzi, Eds. Springer Berlin Heidelberg, 2011, vol. 6649, pp. 207–219. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21560-5_18

[14] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3gpp heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 10–21, June 2011.

[15] L. Franck, M. Berioli, P. Boutry, G. Harles, L. S. Ronga, R. Suffritti, and L. Thomasson, "On the role of satellite communications for emergency situations with a focus on europe," *International Journal of Satellite Communications and Networking*, vol. 29, no. 5, pp. 387–399, 2011. [Online]. Available: http://dx.doi.org/10.1002/sat.979

[16] Inmarsat BGAN. http://www.inmarsat.com/service/bgan/. [Online]. Available: http://www.geo-pictures.eu

[17] C. Dario *et al.*, "Opportunities and challenges of health and telemedicines via satellite," in *Proceedings of ESRIN-Symposium, July 5, Frascati, Italy*, 12 2004.

[18] F. Fontan, M. Vazquez-Castro, C. Cabado, J. Garcia, and E. Kubista, "Statistical modeling of the lms channel," *Vehicular Technology, IEEE Transactions on*, vol. 50, no. 6, pp. 1549–1567, 2001.

[19] Ericsson, "Ericsson mobility report," Tech. Rep., June 2014. [Online]. Available: http://www.ericsson.com/mobility-report

[20] IHP, "Ihp support modules concept paper," International Humanitarian Partnership, Tech. Rep., 2012. [Online]. Available: http://www.ihp.nu/about/background.html

[21] EUTAC, "EU Assessment Mission Kit," European Technical Assistance Cooperation (EUTAC), Tech. Rep., 2009. [Online]. Available: http://www.eutac-project.eu/

[22] A. Rendón Gallón, P. J. Ludeña González, and A. Martínez Fernández, Eds., *Tecnologías de la Información y las Comunicaciones para zonas rurales: Aplicación a la atención de salud en países en desarrollo.* CYTED. Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo, 2011.

[23] G. Pérez Bénitez, "Guía de Tecnologías de Conectividad para Accesso en áreas Rurales." Unión Internacional de Telecomunicaciones UIT Oficina de Desarrollo de las Telecomunicaciones, Informe Final, Noviembre 2007.

[24] J. Radzik, O. Vidal, D. Pradas Fernández, and M. A. Vázquez-Castro, "Cross-layer optimization of hybrid satellite network over the amazonian region," in *Satellite and Space Communications, 2009. IWSSC 2009. International Workshop on*, 2009, pp. 398 –402.

[25] M. A. Vázquez-Castro and D. Pradas, "VoIP cross-layer load control for hybrid satellite-wimax networks," *Wireless Communications, IEEE*, vol. 15, no. 3, pp. 32 –39, june 2008. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4547520&tag=1

[26] D. Vouyioukas, I. Maglogiannis, and V. Pasias, "Pervasive e-health services using the DVB-RCS communication technology," *Journal of Medical Systems*, vol. 31, pp. 237–246, 2007, 10.1007/s10916-007-9061-4. [Online]. Available: http://dx.doi.org/10.1007/s10916-007-9061-4

[27] H. Skinnemoen, M. V. Castro, and R. Aroso, "BRASIL: Broadband to rural america over satellite integrated links," in *Ka Band and Broadband Conference*, 2007.

[28] H. Skinnemoen and A. de Mentzingen, "Latest impact of supporting DVB-RCS for rural America," in *Satellite and Space Communications, 2009. IWSSC 2009. International Workshop on*, Sept 2009, p. 412.

[29] E. T. T. Force, "Cost benefit analysis of satellite-enhanced telemedicine and ehealth services in Sub-Saharan Africa," European Space Agency ESA, Report, November 2008.

[30] K. Sunitha and P. Vidya, "Impact of satellite based telemedicine systems in rural areas of Tamilnadu," *Calicut Medical Journal*, vol. 6, 2008.

[31] P. K. Misra, "A dedicated satellite for meeting health education needs of afro-asian nations: Possibilities, action plan and benefits," *The Electronic Journal of Information Systems in Developing Countries*, vol. 41, no. 6, pp. 1–12, 2010.

[32] "Amerhis system," http://www.alcatel.es/espacio/pdf/amerhis.pdf.

[33] I. Sachpazidis, E. Callejo, and G. Saka, "T@HIS: A satellite based telemedicine network for imaging teleconsultations," *Latin-Am J Telehealth*, vol. 1, no. 1, pp. 105–122, 2009.

[34] D. Panopoulos, I. Sachpazidis, D. Rizou, W. Menary, J. Cardenas, and J. Psarras, "MedNET: Telemedicine via satellite combining improved access to health-care services with enhanced social cohesion in rural Peru," in *Information Systems Development*. Springer US, 2009, pp. 933–940. [Online]. Available: http://dx.doi.org/10.1007/b137171_98

[35] V. Singh, J. Ott, and I. Curcio, "Rate-control for conversational video communication in heterogeneous networks," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2012 IEEE International Symposium on a*, 2012, pp. 1–7.

[36] B. Wang, J. Kurose, P. Shenoy, and D. Towsley, "Multimedia Streaming via TCP: An Analytic Performance Study," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 4, no. 2, pp. 1–22, May 2008. [Online]. Available: http://doi.acm.org/10.1145/1352012.1352020

[37] S. Floyd, M. Handly, J. Padhye, and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification, IETF RFC 5348," United States, 2008.

[38] H. Seferoglu, A. Markopoulou, U. C. Kozat, M. R. Civanlar, and J. Kempf, "Dynamic FEC algorithms for TFRC flows," *Trans. Multi.*, vol. 12, no. 8, pp. 869–885, Dec. 2010. [Online]. Available: http://dx.doi.org/10.1109/TMM.2010.2053840

[39] H.-P. Shiang and M. van der Schaar, "A quality-centric TCP-Friendly congestion control for multimedia transmission," *Multimedia, IEEE Transactions on*, vol. 14, no. 3, pp. 896–909, 2012.

[40] K. Tappayuthpijarn, G. Liebl, T. Stockhammer, and E. Steinbach, "Adaptive video streaming over a mobile network with tcp-friendly rate control," in *IWCMC '09: Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing*. New York, NY, USA: ACM, 2009, pp. 1325–1329.

[41] L. De Cicco, G. Carlucci, and S. Mascolo, "Experimental investigation of the google congestion control for real-time flows," in *Proceedings of the 2013 ACM SIGCOMM workshop on Future human-centric multimedia networking*, ser. FhMN '13. New York, NY, USA: ACM, 2013, pp. 21–26. [Online]. Available: http://doi.acm.org/10.1145/2491172.2491182

[42] "RTP Media Congestion Avoidance Techniques (rmcat)." [Online]. Available: https://datatracker.ietf.org/wg/rmcat/charter/

[43] T. Schierl, M. Hannuksela, Y.-K. Wang, and S. Wenger, "System layer integration of high efficiency video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1871–1884, 2012.

[44] Authors, "Rtp payload for vp8 ietf RFC," United States, 2012. [Online]. Available: http://tools.ietf.org/html/draft-ietf-payload-vp8-08

[45] M. van der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 wlans," *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 10, pp. 1752 – 1763, dec. 2003.

[46] P. Tournoux, E. Lochin, J. Lacan, A. Bouabdallah, and V. Roca, "On-the-fly erasure coding for real-time video applications," *Multimedia, IEEE Transactions on*, vol. 13, no. 4, pp. 797–812, 2011.

[47] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *41st Annu. Allerton Conf. Communication, Control and Computing*, October 2003.

[48] T. Ho, M. Medard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *Information Theory, IEEE Transactions on*, vol. 52, no. 10, pp. 4413–4430, 2006.

[49] D. S. Lun, M. Medard, and M. Effros, “On coding for reliable communication over packet networks,” in *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, September 2004.

[50] H. Balli, X. Yan, and Z. Zhang, “On randomized linear network codes and their error correction capabilities,” *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 3148–3160, 2009.

[51] D. Vukobratovic, C. Khirallah, V. Stankovic, and J. Thompson, “Random network coding for multimedia delivery services in lte/lte-advanced,” *Multimedia, IEEE Transactions on*, vol. 16, no. 1, pp. 277–282, Jan 2014.

[52] P. Saxena and M. A. Vázquez-Castro, “Network Coding advantage over MDS codes for multimedia transmission via erasure satellite channels,” in *Personal Satellite Services (PSATS), 2013 International Conference on*, October 2013.

[53] D. Pradas and M. A. Vázquez-Castro, “NUM-Based Fair Rate-Delay Balancing for Layered Video Multicasting over Adaptive Satellite Networks,” *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 5, pp. 969–978, 2011.

[54] D. Pradas, A. Bouabdallah, J. Lacan, M. Vazquez Castro, and M. Bousquet, “Cross-layer optimization of unequal protected layered video over hierarchical modulation,” in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, 30 2009-dec. 4 2009, pp. 1 –6.

[55] M. A. Vázquez Castro and D. Pradas Fernández, “VoIP cross-layer load control for hybrid satellite-WiMAX networks,” *Wireless Communications, IEEE*, vol. 15, no. 3, pp. 32 –39, june 2008. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4547520&tag=1

[56] G. Xylomenos, C. N. Ververidis, V. A. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. V. Katsaros, and G. C. Polyzos, “A survey of information-centric networking research,” *Communications Surveys Tutorials, IEEE*, vol. 16, no. 2, pp. 1024–1049, Second 2014.

[57] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, “A survey of information-centric networking,” *Communications Magazine, IEEE*, vol. 50, no. 7, pp. 26–36, July 2012.

[58] H. Wirtz and K. Wehrle, “Opening the loops - towards semantic, information-centric networking in the internet of things,” in *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2013 10th Annual IEEE Communications Society Conference on*, June 2013, pp. 18–24.

[59] C. Tsilopoulos, G. Xylomenos, and G. Polyzos, “Are Information-Centric Networks Video-Ready?” in *Packet Video Workshop (PV), 2013 20th International*, Dec 2013, pp. 1–8.

[60] F. Kuipers, R. Kooij, D. De Vleeschauwer, and K. Brunnström, “Techniques for measuring quality of experience,” in *Proceedings of the 8th international conference on Wired/Wireless Internet Communications*, ser. WWIC’10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 216–227. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13315-2_18

[61] M. T. Ivrlac, L. U. Choi, E. Steinbach, and J. A. Nossek, "Models and analysis of streaming video transmission over wireless fading channels," *Signal Processing: Image Communication*, vol. 24, no. 8, pp. 651 – 665, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923596509000411

[62] K. Laghari and K. Connelly, "Toward total Quality of Experience: A QoE model in a communication ecosystem," *Communications Magazine, IEEE*, vol. 50, no. 4, pp. 58–65, April 2012.

[63] S. Barakovic and L. Skorin-Kapov, "Survey and Challenges of QoE Management Issues in Wireless Networks," *Journal Comp. Netw. and Communic.*, vol. 2013, 2013.

[64] P. L. Callet, S. Moeller, and A. Perkis, "Qualinet white paper on definitions of Quality of Experience Version 1.1," European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Tech. Rep., June 2012.

[65] M. Alreshoodi and J. Woods, "Survey on QoE/QoS correlation models for multimedia services," *CoRR abs/1306.0221*, 2013.

[66] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between Quality of Experience and Quality of Service," *Network, IEEE*, vol. 24, no. 2, pp. 36–41, 2010.

[67] T. Hosfeld, M. Fiedler, and T. Zinner, "The QoE provisioning-delivery-hysteresis and its importance for service provisioning in the Future Internet," in *Conference on Next Generation Internet Networks*, 2011.

[68] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo, "The Logarithmic Nature of QoE and the Role of the Weber-Fechner Law in QoE Assessment," in *Communications (ICC), 2010 IEEE International Conference on*, 2010, pp. 1–5.

[69] ITU-T, *G.1070 Opinion model for video-telephony applications*, ITU-T Std., 07 2012.

[70] M. Mu, A. Mauthe, and F. Garcia, "A utility-based qos model for emerging multimedia applications," in *2nd Intl. Conference on Next Generation Mobile Applications, Services, and Technologies*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 521–528. [Online]. Available: http://portal.acm.org/citation.cfm?id=1510534.1511835

[71] S. Hussain, R. Harris, and G. Punchihewa, "Dominant factors in the content domain that influence the QoE of an IPTV service," in *TENCON Spring Conference, 2013 IEEE*, April 2013, pp. 572–577.

[72] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial delay vs. interruptions: Between the devil and the deep blue sea," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, July 2012, pp. 1–6.

[73] A. Vakili and J.-C. Grégoire, "QoE management for video conferencing applications," *Computer Networks*, vol. 57, no. 7, pp. 1726 – 1738,

2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128613000558

[74] A. ParandehGheibi, M. Medard, S. Shakkottai, and A. Ozdaglar, "Avoiding interruptions - QoE trade-offs in block-coded streaming media applications," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, 2010, pp. 1778–1782.

[75] G. Tian and Y. Liu, "Towards agile and smooth video adaptation in dynamic HTTP streaming," in *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, ser. CoNEXT '12. New York, NY, USA: ACM, 2012, pp. 109–120. [Online]. Available: http://doi.acm.org/10.1145/2413176.2413190

[76] O. Habachi, Y. Hu, M. van der Schaar, Y. Hayel, and F. Wu, "MOS-based congestion control for conversational services in wireless environments," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 7, pp. 1225–1236, 2012.

[77] *ITU-T. Recommendation ITU-T P.910 on Subjective video quality assessment methods for multimedia applications. (04/2008).*, ITU-T Std., 04 2008.

[78] ITU, *ITU-T J.247 Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference*, ITU Std., 08 2008.

[79] 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, June 2011.

[80] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600 –612, april 2004.

[81] C. Yim and A. C. Bovik, "Evaluation of temporal variation of video quality in packet loss networks," *Signal Processing: Image Communication*, vol. 26, no. 1, pp. 24 – 38, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923596510001153

[82] J. Strassner, J. Betser, R. Ewart, and F. Belz, "A Semantic Architecture for Enhanced Cyber Situational Awareness," in *Secure and Resilient Cyber Architectures Conference*, 2010.

[83] M. R. Endsley, "Theoretical underpinnings of situation awareness: A critical review," *Situation awareness analysis and measurement*, pp. 3–32, 2000.

[84] M. Ackerman and C. Locatis, "Advanced networks and computing in healthcare," *Journal of the American Medical Informatics Association*, vol. 18, no. 4, pp. 523–528, 2011. [Online]. Available: http://jamia.bmj.com/content/18/4/523.abstract

[85] F. Bergstrand and J. Landgren, "Information sharing using live video in emergency response work," in *Proceedings of 6th International ISCRAM Conference*. ISCRAM, May 2009.

[86] S. S. Krupenia, C. Aguero, and K. C. Nieuwenhuis, "The value of different media types to support command and control situation awareness," in *Proceedings of 9th International ISCRAM Conference*, 2012.

[87] S. Eichler and C. Schroth, "A multi-layer approach for improving scalability of vehicular ad-hoc networks," in *Proceedings of the 4th Workshop ÃŒber Mobile Ad-Hoc Netzwerke (WMAN)*, Mar. 2007.

[88] K. Lee, E. Modiano, and H.-W. Lee, "Cross-layer survivability in wdm-based networks," *Networking, IEEE/ACM Transactions on*, p. (accepted), 2010. [Online]. Available: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5643973

[89] C. hsin Hsu and M. Hefeeda, "A framework for cross-layer optimization of video streaming in wireless networks," *ACM Trans. on Multimedia Computing, Communications, and Applications*, 2009, accepted to Appear.

[90] B. BollobÂ·s, *Modern Graph Theory*. Springer-Verlag, 2002, pp. 70–71.

[91] G. Giambene and S. Kota, "Cross-layer protocol optimization for satellite communications networks: a survey," *International Journal of Satellite Communications and Networking*, vol. 24, no. 5, pp. 323–341, 2006. [Online]. Available: http://www3.interscience.wiley.com/journal/113328334/abstract?CRETRY=1&SRETRY=0

[92] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design of ad hoc networks for real-time video streaming," *Wireless Communications, IEEE*, vol. 12, no. 4, pp. 59 – 65, aug. 2005.

[93] C. An and T. Nguyen, "Cross-layer optimization for video communication over AWGN channel," in *Global Telecommunications Conference, 2007. GLOBECOM '07. IEEE*, 2007, pp. 3348 –3352.

[94] I. Djama, T. Ahmed, A. Nafaa, and R. Boutaba, "Meet in the middle cross-layer adaptation for audiovisual content delivery," *Multimedia, IEEE Trans. on*, vol. 10, no. 1, pp. 105 –120, jan. 2008.

[95] N. Tizon, B. Pesquet-Popescu, and M. Cagnazzo, "Adaptive video streaming with long term feedbacks," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2009, pp. 949 –952.

[96] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications, ietf RFC 3550," United States, 2003.

[97] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, 2007. [Online]. Available: http://dx.doi.org/10.1109/JPROC.2006.887322

[98] R. Vaz, B. W. M. Kuipers, and M. Nunes, "Video quality optimization algorithm for video-telephony over IP networks," in *Personal Indoor and Mobile Radio Communications, IEEE 21st International Symposium on*, 2010, pp. 2545–2550.

[99] D. H. Finstad, H. K. Stensland, H. Espeland, and P. Halvorsen, “Improved multi-rate video encoding,” in *Proceedings of the 2011 IEEE International Symposium on Multimedia*, ser. ISM ’11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 293–300. [Online]. Available: http://dx.doi.org/10.1109/ISM.2011.53

[100] I. E. Richardson, *The H.264 Advanced video Compression Standard: electronic version.* John Wiley & Sons Ltd, 2010.

[101] Z. Rosberg, “Rate control with end-to-end delay and rate constraints,” in *INFOCOM Workshops, IEEE*, 2008, pp. 1–6.

[102] T. Alpcan and T. Basar, “A globally stable adaptive congestion control scheme for internet-style networks with delay,” *Networking, IEEE/ACM Transactions on*, vol. 13, no. 6, pp. 1261–1274, 2005.

[103] J. Bankoski, P. Wilkins, and Y. Xu, “Technical overview of VP8, an open source video codec for the web,” in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, 2011, pp. 1–6.

[104] “Video test media,” http://media.xiph.org/video/derf/, Jun. 2013.

[105] “Netem network emulator.” [Online]. Available: http://www.linuxfoundation.org/collaborate/workgroups/networking/netem

[106] J. Thomas-Kerr, C. Ritz, and I. Burnett, “Semantic-aware delivery of multimedia,” in *Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on*, Sept 2009, pp. 1498–1503.

[107] C. Henson, A. Sheth, and K. Thirunarayan, “Semantic perception: Converting sensory observations to abstractions,” *Internet Computing, IEEE*, vol. 16, no. 2, pp. 26–34, 2012.

[108] A. Cavallaro and S. Winkler, “Perceptual semantics,” in *Multimedia Technologies: Concepts, Methodologies, Tools, and Applications*, S. M. Rahman, Ed. Information Science Reference, Jun. 2008, ch. 7.5, pp. 1441–1455.

[109] S. B. Kodeswaran and A. Joshi, “Content and context aware networking using semantic tagging,” in *Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on*. IEEE, 2006, pp. 77–77.

[110] 3GPP, *Policy and charging control architecture*, 3GPP Std., 03 2010.

[111] V. Joseph and G. de Veciana, “Jointly optimizing multi-user rate adaptation for video transport over wireless systems: Mean-fairness-variability tradeoffs,” in *2012 Proceedings IEEE INFOCOM*. IEEE, Mar. 2012, pp. 567–575. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6195799

[112] Z. Ma, F. Fernandes, and Y. Wang, “Analytical rate model for compressed video considering impacts of spatial, temporal and amplitude resolutions,” in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, July 2013, pp. 1–6.