



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par **l'Institut Supérieur de l'Aéronautique et de l'Espace**
Spécialité : STIC Réseaux, télécom, système et architecture

préparée en cotutelle internationale de thèse avec
l'Université Autonome de Barcelone (UAB)

Présentée et soutenue par **David PRADAS FERNÁNDEZ**
le **23 septembre 2011**

**Conception et méthodologie cross-layer
pour les réseaux satellite à haut débit**

**Cross-layer design and methodology
for satellite broadband networking**

JURY

Mme Fotini-Niovi Pavlidou, présidente, rapporteur
M. Toufik Ahmed, rapporteur
M. Michel Bousquet
M. Jérôme Lacan, directeur de thèse
M. Guy Lesthievant
M. Harald Skinnemoen
Mme Maria Ángeles Vázquez-Castro, directrice de thèse

École doctorale : **Mathématiques, informatique et télécommunications de Toulouse**

Unités de recherche : **Équipe d'accueil ISAE-ONERA MOIS
UAB/TES**

Directeurs de thèse : **M. Jérôme Lacan – Mme Maria Ángeles Vázquez-Castro**

Cross-Layer Design and Methodology for Satellite Broadband Networking

Ph. D. Dissertation

DAVID PRADAS FERNÁNDEZ

I would like to dedicate this thesis to my loving parents (Pere and Mercè).

Acknowledgements

It is not a fair task to acknowledge all the people who made this Ph.D. thesis possible with a few words. However, I will try to do my best to extend my great appreciation to everyone who helped me scientifically and emotionally through out this study.

First of all, I would like to thank my research advisor, M. Ángeles Vázquez Castro, for her support and motivation through this Ph.D. dissertation, and for guiding me on this exciting project. I also want to thank my co-advisor, Jérôme Lacan for his help during my stay in Toulouse, as well as Michel Bousquet for making it possible. Moreover, I would like to thank Niovi Pavlidou and Toufik Ahmed for reviewing my thesis and for their nice comments, and also to Harald Skinnemoen for his smart feedback of my thesis dissertation and to Guy Lesthievant for attending the defense. Representing the UAB, CNES and ISAE I would like to thank Josep Parrón, Alessandro Cardoso, Laurence Clarac and Emmanuel Lochin.

With great appreciation I would like to thank my research colleagues at UAB, Jiang Lei and Joan Enric, and my office colleagues, Albert Bel, Mercè, not only for many useful discussions but also for your friendship. I would like also to thank Dani, Orlando, Monica, Roman, Laia and Carles. I can not forget my office colleagues during my stay in Toulouse (at TéSA), thank you to Julien, Marianna, Patrick and Fares.

A very special thank you to my family and friends, they know who they are (the list is too big). And in particular to my parents (Pere and Mercè) and to Laura for all the support/affection/love... they have given me, I would never have done it if it was not for them.

Finally, I would also like to thank UAB, CNES and TéSA for the joint financing of my research fellowship. And to our SatNEx partners for all those fruitful collaborations. I would like to acknowledge to all projects where I have been able to participate.

Abstract

The demand for high-speed networking, driven mainly by the rapid expansion of the Internet, has been growing at an exponential rate. A wide range of technologically innovative wireline and wireless solutions already offer competitive broadband connectivity. Communication satellites emerge as an attractive solution in providing broadband connectivity to a variety of users thanks to its inherent global coverage. The broadcast nature of satellites makes them the natural choice for multicasting services, for interconnecting geographically distributed high-speed networks, and for providing multimedia services to both home and business users.

In order to accommodate the heterogeneity of users and multimedia traffic, broadband satellite networks are moving from the crowded K_u band to the more suitable K_a band. However, this higher frequency band imposes challenging channel conditions (attenuations up to dozens of dBs), for which novel technologies have already been adopted for upcoming broadband satellite networks. The main innovation has been the adoption of adaptive coding and modulation at the physical layer. This is the main driver of this thesis due to the crucial fact that such adaptivity makes traditional satellite system design totally inefficient. In addition, and as a natural way to increase the overall satellite capacity, multibeam technology is also assumed in such broadband satellite networks.

In this thesis, we focus on a different paradigm to address such new challenges due to adaptivity and multibeam technology. Such paradigm of system design is based on a joint optimization across layers of the protocol stack and has been the subject of a profusion of research work in the last years both for terrestrial and satellite networks. The fundamental idea behind this concept is the fact that adaptivity at the physical layer should be followed at upper layers in order to achieve efficient management of the system resources. Furthermore, it is the only way to comply with the stringent Quality of Service (QoS) of new applications services.

Instead of focusing on one aspect only of the cross-layer design for broadband satellite systems, we cover several aspects related to the networking optimization, including interoperability with terrestrial networks (thus also covering hybrid networking). Networking optimization deals with allocating resources efficiently, maximizing the throughput and assuring fairness among all the users, according to channel conditions measured at the physical layer. Delay, jitter and packet loss requirements (straightforward related to video and speech quality) for constrained services are specially challenging when they

need to cope with the channel dynamics of the wireless satellite link. In order to meet these requirements, we take profit of adaptive technologies at higher layers (transport and application) such as layered coding and unequal error protection for videostreaming and adaptive codes for Voice over IP (VoIP).

Our efforts have been also focused on choosing the best methodology in terms of:

1. Selection of mathematical tools that best fit with each cross-layer approach. In particular, we have used convex optimization, Network Utility Maximization (NUM), game theory and linear programming.
2. Selection of the most suitable architecture. In particular, hybrid architectures including satellites are of special interest for wireless networks deployment in rural and remote areas. Moreover, conventional centralized vs novel decentralized architectures have also been traded-off.
3. Selection of the best approach to the cross-layer design, implementation, and performance evaluation. In particular, we have first set a taxonomy of approaches so as it is possible to choose the most suitable one for each scenario.
4. Selection, if needed, of truly available standardized tools, hence, ready-to-use for the realistic implementation of our developed schemes.

The contributions of this Ph.D. dissertation can be summarized as follows:

- a) We study the resource allocation in multibeam broadband satellite systems. First, we propose to optimize the multicast transmission of the air interface for satellite and hybrid (satellite+terrestrial) networks. We focus on hierarchical allocation algorithms based on Superposition Coding (SC) and Adaptive Coding and Modulation (ACM). We prove that these algorithms can be adapted according to cross-layer information, depending on the channel conditions and users' mobility.
- b) We address the challenging problem of capacity drop at K_a Band due to weather conditions, for both the forward and the return link of multibeam broadband satellite systems. In order to do so, we propose a cross-layer DiffServ architecture to optimize the interplay of lower layers, i.e. Physical (PHY), Medium Access Control (MAC) and Network (or Internet Protocol (IP)). We formulate the optimization problem according to a NUM paradigm and we take advantage of Game Theory to investigate the impact of different performances of centralized and distributed approaches. We prove that our solutions maximizes capacity, while at the same time meets QoS requirements (in terms of delay/jitter).
- c) The analytical methodology performed and the results obtained allow us to formulate the problem according to fair and efficient policies. We contribute with different tailored trade-offs, such as rate vs. delay or efficiency vs. fairness for different satellite weather conditions.

- d) The problem of capacity drop and resources allocation in satellite multi-cast systems can be also faced up from the higher layers point of view, i.e. Transport and Application (APP) layers, which are able to interact with PHY-layer technologies (adaptive physical layer and Hierarchical Modulation) thanks to our cross-layer approach. Herein, we propose adaptive and scalable solutions in order to fulfil QoS requirements of real-time applications (such as VoIP/Videoconferencing) or strong loss-constrained applications such as videostreaming. In particular, we take advantage of Adaptive Multi Rates coding for VoIP, and Dependency Aware Unequal Error Protection (DA-UEP) and Scalable Video Coding (SVC) for videostreaming. We show that our cross-layer designs outperform non-cross-layer designs in terms of video quality (PSNR) and speech quality (MOS).
- e) We have also addressed and solved scenarios of particular interest, such as the hybrid transmission to rural areas, or the difficulty to give wireless services in tropical areas (due to strong weather conditions and terrain morphology).

Contents

Contents	xi
List of Figures	xvii
List of Tables	xxi
List of Acronyms and Abbreviations	xxiii
1 Introduction	1
1.1 Motivation	1
1.1.1 Wireless/Satellite Communications	1
1.1.2 Cross-layer	2
1.2 Outline of Dissertation	3
1.3 Research Contributions	4
1.3.1 Chapter 3	4
1.3.2 Chapter 4	5
1.3.3 Chapter 5	5
1.3.4 Chapter 6	5
1.4 Thesis Context	7
2 Cross-layer Design Paradigm	9
2.1 Introduction	9
2.2 Taxonomy of Cross-Layer Methodologies	9
2.2.1 Implicit and explicit cross-layer design	10
2.2.2 Based on optimization criteria	10
2.2.2.1 Global-performance-optimization-oriented	10
2.2.2.2 Protocol-optimization-oriented	11
2.2.3 Based on direction of information flow	11
2.2.3.1 Bottom-up approach	12
2.2.3.2 Top-down approach	12
2.2.3.3 Hybrid approach	13
2.2.4 Based on layers involved	13
2.3 Taxonomy of cross-layer signaling methods	13
2.3.1 Based on Packet headers	14
2.3.2 Based on Internet Control Message Protocol (ICMP) messages	14
2.3.3 Based on a network service	14

2.3.4	Based on local profiles	14
2.3.5	Concept mode: CLASS	14
2.3.6	BSM particular case: SI-SAP Primitives	15
2.4	Optimal Solution Challenges	16
3	Satellite and Hybrid Systems and Channel Models	17
3.1	Channel Models	17
3.1.1	Satellite Channel Model	17
3.1.1.1	Rain Correlated Models	17
3.1.1.1.1	Tropical and Temperate	18
3.1.1.2	Land Mobile Satellite Models	19
3.1.2	Terrestrial Propagation Channel Model	19
3.2	System and QoS Models	19
3.2.1	Practical Standards Overview	19
3.2.1.1	DVB-S2	19
3.2.1.1.1	Adaptive physical layer	20
3.2.1.1.2	System architecture	21
3.2.1.1.3	Hierarchical Modulation	21
3.2.1.1.4	IP data encapsulation	21
3.2.1.2	DVB-RCS	22
3.2.1.2.1	PHY-layer: Adaptive Coding	22
3.2.1.2.2	MAC-layer: MF-TDMA	22
3.2.1.2.3	Topology	23
3.2.1.2.4	Mobility	23
3.2.1.3	DVB-SH	23
3.2.1.3.1	Architectures	23
3.2.1.3.2	PHY-layer outline	23
3.2.1.3.3	Upper layer outline	24
3.2.1.4	WiMAX	24
3.2.1.4.1	PHY layer key features	24
3.2.1.4.2	MAC layer key features	25
3.2.2	QoS Model	26
3.2.2.1	DVB-S2/RCS	26
3.2.2.2	WiMAX	27
3.3	Study case: An Hybrid Architecture	27
3.3.1	Demographic aspects	27
3.3.2	Hybrid Reference Architecture	28
3.3.2.1	The DVB-S2/RCS Standard and WiMAX	28
3.3.2.2	Architecture & Protocol Stack	29
3.3.3	QoS Model	31
4	Cross-Layer Design at System Level	33
4.1	Objective: Capacity Maximization	33
4.2	Methodology: Hierarchical Multicast Transmission	34
4.3	Optimization Problem and Solution: HA-SC	35

4.3.1	Time Scheduling (TS) reference	38
4.4	Mobile Hybrid Scenario	40
4.5	Results	41
4.6	Conclusions	44
4.6.1	Conclusions on the methodology	44
4.6.2	Conclusions on the performance gain	45
5	Cross-Layer Design at PHY/MAC/IP Level	47
5.1	Objective: Capacity Maximization with QoS provision	47
5.2	Forward Link	48
5.2.1	Methodology: Adaptive Fairness	48
5.2.1.1	Cross-layer related work	48
5.2.1.2	Network Utility Maximization related work	48
5.2.1.3	Contributions	49
5.2.2	Satellite/wireless network model	50
5.2.3	Architecture, physical and link layers	50
5.2.3.1	IP Traffic model	51
5.2.4	QoS PHY/MAC/IP-layer scheduler: DiffServ	51
5.2.4.1	α -fairness WRR: channel conditions	51
5.2.4.2	Adaptive WRR: delay aware	53
5.2.5	Optimization: Rate-Delay balancing across MAC /IP queues	54
5.2.5.1	Formulation Framework	54
5.2.5.2	NUM Solution: Water Filling Rate/Delay balancing	55
5.2.5.3	Interpretation of Water Filling Solution	56
5.2.6	Standard, Protocol Design and Implementation: FairDB	57
5.2.6.1	Transmission standard: DVB-S2	57
5.2.6.2	Objectives of FairDB	58
5.2.6.3	SI-SAP Architecture and Flows	58
5.2.7	Performance Results	59
5.2.7.1	Per-layer parameters	59
5.2.7.2	Simulation Results	60
5.3	Return Link	61
5.3.1	Methodology: Adaptive Fairness	61
5.3.1.1	Centralized vs. Decentralized Satellite Networking	61
5.3.1.2	Game Theory related work	62
5.3.1.3	Contributions	63
5.3.2	Interactive Satellite Network Model	64
5.3.2.1	Architecture, physical and link layers	64
5.3.2.2	Interactive traffic model	66
5.3.3	Justification of the Game Formulation	67
5.3.4	Optimization: Distributed Bandwidth Request Game	67
5.3.4.1	Modeling the satellite selection game	67
5.3.4.2	N-user non-cooperative game 1: Capacity unaware	68
5.3.4.3	N-user non-cooperative game 2: capacity aware	69
5.3.4.4	N-user non-cooperative game 3: Capacity requests aware	70

5.3.5	Cross-Layer Architecture Design: DistFair	72
5.3.5.1	DVB-RCS framing description	73
5.3.5.2	DVB-RCS capacity request mapping	73
5.3.5.3	Intra Network Cross-layer: <i>XL-net</i>	74
5.3.5.4	MAC to IP Explicit Cross-layer: <i>XL-MAC-IP</i>	74
5.3.6	Results	76
5.3.6.1	Bandwidth requests: channel correlation	76
5.3.6.2	Efficiency and Fairness	79
5.3.6.3	DVB-RCS signaling savings	81
5.4	Conclusion	83
5.4.1	Conclusions on the methodology	83
5.4.1.1	Forward link	84
5.4.1.2	Return link	85
5.4.2	Conclusions on the performance gain	85
6	Cross-Layer Design at PHY/APP Level	89
6.1	Objective: adaptive scalable QoS and error protection (VoIP and Video applications)	89
6.2	Methodology: Bit rate adaptation (Scalable QoS)	90
6.2.1	XL Design 1: Layered Video Multicasting	90
6.2.1.1	Modeling the Scalable Video Coding	91
6.2.1.2	Optimization: Rate-Delay Balancing across Layered Video Multicasting	91
6.2.1.2.1	Video layers and ACM: multicast mapping	91
6.2.1.2.2	Formulation Framework	93
6.2.1.2.3	NUM Solution: Water Filling Layered Multicasting	93
6.2.1.3	Standard, Protocol Design and Implementation: MulFairDB	94
6.2.1.4	Performance Results	94
6.2.2	XL Design 2: Adaptive VoIP Codecs	96
6.2.2.1	Capacity drop due to DVB-S2/RCS adaptation to channel conditions	98
6.2.2.2	Centralized vs. Distributed Approaches	99
6.2.2.3	Distributed design: Rate Control	99
6.2.2.3.1	RTCP-based rate adaptation	99
6.2.2.3.2	AMR-based rate adaptation	101
6.2.2.3.3	Delay Budget Model and Performance Model	103
6.2.2.3.4	Cross-layer and Signaling	105
6.2.2.4	Simulation Results	106
6.2.2.5	Centralized approach: Load Control	109
6.2.2.5.1	Dimensioning of Satellite subnetwork	110
6.2.2.5.2	Dimensioning of Terrestrial subnetwork	110
6.2.2.5.3	Optimization: Load Control	111
6.2.2.5.4	Cross-layer and Primitives	112
6.2.2.5.5	Simulation Results	113
6.3	Methodology: Adaptive Error Protection	116

6.3.1	XL Design 3: High layer-FEC protection and PHY Layer optimization	116
6.3.1.1	Contributions	117
6.3.1.2	Fading Mitigation Techniques	117
6.3.1.2.1	LL-FEC	118
6.3.1.2.2	AL-FEC Framework	119
6.3.1.2.3	Shifted Threshold	120
6.3.1.3	Results	121
6.3.1.3.1	Simulation Platform	121
6.3.1.3.2	Parameters and Results	121
6.3.1.3.3	Video Quality Evaluation	124
6.3.1.3.4	Cross-layer design	127
6.3.2	XL Design 4: Unequal Error Protection (UEP) optimization	127
6.3.2.1	Multicasting and Protection strategies	128
6.3.2.1.1	HM, UEP and SVC Introduction	129
6.3.2.2	Hierarchical Modulation	130
6.3.2.3	DA-UEP for Transport Layer	132
6.3.2.4	Cross-layer Architecture Design	133
6.3.2.5	Optimization: Maximize Distortion	135
6.3.2.5.1	Formulation Framework	135
6.3.2.5.2	NUM: non-convex problem	135
6.3.2.5.3	Parameters Definition	135
6.3.2.6	Simulation Results	136
6.3.2.6.1	LMS Analysis	137
6.4	Conclusions	137
6.4.1	Conclusions on the methodology	137
6.4.1.1	Adaptive bit rate	138
6.4.1.1.1	Layered Video Multicasting	138
6.4.1.1.2	Adaptive VoIP Codecs-Centralized	139
6.4.1.1.3	Adaptive VoIP Codecs-Distributed	141
6.4.1.2	Adaptive error protection	141
6.4.1.2.1	High layer-FEC protection and PHY layer optimization	142
6.4.1.2.2	UEP optimization	142
6.4.2	Conclusions on the performance gain	143
7	Overall Conclusions and Future Work	145
7.1	Conclusions	145
7.2	Conclusions on the methodology	147
7.2.1	Bit rate adaptation	147
7.2.2	Adaptive error protection	148
7.2.3	Final comments	149
7.3	Future works	150
	References	151

List of Figures

2.1	Cross-layer signaling methods: (a) Based on packet headers; (b) Based on ICMP; (c) Based on a network service; (d) Based on local profiles.	15
3.1	Examples of channel conditions caused by a rain event in Temperate and Tropical areas.	18
3.2	Channel attenuation time-series for different LMS Environments at S-band: (a) Open area; (b) Intermediate tree shadowing (ITS); (c) Heavy tree shadowing (HTS); (d) Suburban area.	20
3.3	Block Diagram of a DVB-S2 ACM link.	21
3.4	DVB-S2/RCS star topology with WiMAX networks associated to Satellite terminals.	29
3.5	Reference architecture and protocol stack for the hybrid WiMAX and DVB-S2/RCS network.	30
4.1	Hybrid DVB-SH Architecture with the proposed RRM for Multicast Transmission	41
4.2	Capacity gain of multicast groups when increasing the minimum rate allocated the worst group (group 5), using HA-SC algorithm	42
4.3	Normalized throughput achieved with HA-SC algorithm depending on the number of users served and for different channel condition	43
4.4	Normalized throughput achieved with HA-SC algorithm depending on the number of users served and for different number of SC layers.	43
4.5	Normalized throughput achieved with TS algorithm depending on the number of users served (L) and for different channel condition	44
4.6	Cross-layer interaction and methodology conclusions: HA-SC	45
5.1	Video transmission scenario over a wireless/satellite link, multicast groups organized per video application requested and per geographical area. . . .	50
5.2	Cross-layer scheme of the adaptive PHY/MAC/IP schedulers	52
5.3	Queue balancing interpretation of the water-filling solution method.	57
5.4	Architecture of the cross-layer design, including exchanged information and per-layer block description.	59
5.5	Comparison between AWRR and WRR in terms of EF delay and total delay for different maximum rain attenuations.	60
5.6	Star-like topology of and interactive multi-beam satellite network. Users are organized according to weather conditions	65

5.7	Set of feasible solutions (in terms of Mbps) for different Optimization problems	72
5.8	BSM Architecture of MAC to IP Explicit Cross-layer scheme	75
5.9	SI-SAP primitives signaling flows performing in different scenarios	76
5.10	Centralized (selfish) traffic requests per user during a rain event affecting 60 % of users	77
5.11	Traffic request per user during a rain event affecting 60 % of users, by using DistFair with $\alpha = 1$	78
5.12	Traffic request per user during a rain event affecting 60 % of users, by using DistFair with $\alpha = 0$	78
5.13	Aggregated bandwidth demands and DistFair capacity requests during a rain event	79
5.14	Ratio requested traffic vs. aggregated traffic demands in different rain scenarios	80
5.15	Bandwidth requests efficiency depending on pricing α for different distributed games, when satellite link is affected by rain	80
5.16	Jain's fairness index depending on pricing α for different distributed games, when rain is dropping the satellite link capacity	81
5.17	Bandwidth requested by users under clear sky conditions depending on pricing α and for different distributed games, when satellite link is affected by rain	82
5.18	Comparison between Selfish requests in a centralized network and the DistFair requests	83
5.19	Bandwidth gain (%) depending on the SYNC rate in terms of SFs	84
5.20	Cross-layer interaction and methodology conclusions: Forward link	86
5.21	Cross-layer interaction and methodology conclusions: Return link	86
6.1	Cross-layer scheme of the adaptive PHY/MAC/IP schedulers and Layered Video	92
6.2	Architecture of the cross-layer design, including exchanged information and per-layer block description	95
6.3	Comparison between MultFairDB (for $\alpha = 0$) and a non cross-layer approach in terms of video layers distribution among the users.	96
6.4	Comparison between MultFairDB and a non cross-layer approaches in terms of average rate received per group depending on channel conditions.	97
6.5	Comparison between MultFairDB and non cross-layer approaches in terms of end-to-end delay depending on channel conditions.	97
6.6	RTCP-driven and AMR-WB architectures	100
6.7	RTCP-driven and AMR-WB flows	104
6.8	EF traffic delay and aggregated Throughput for a channel attenuation of 3 dB, VoIP using the G. 711 codec.	107
6.9	EF traffic delay for a channel attenuation of 12 dB, VoIP using the G. 711 codec.	107
6.10	EF aggregated throughput for a channel attenuation of 12 dB for the two cross-layer approaches. AMR-WB (left) and RTCP-driven rate control (right)	108

LIST OF FIGURES

6.11 Probability of exceeding delay requirements (270 ms) for EF depending on rain attenuation. Comparison between reference scenario and cross-layer approaches. 109

6.12 R-factor for G. 729 for different delay guarantees and number of satellite hops. 110

6.13 Percentage of VoIP Connections that can be maintained using cross-layer VoIP load control for different peak attenuations and percentage of users experiencing the attenuation. 114

6.14 Time evolution of percentage of VoIP connections that can be maintained using cross-layer VoIP load control for an attenuation of 15 dB in a temperate area. Comparison to the non-cross-layer scenario. 115

6.15 Time evolution of percentage of VoIP connections that can be maintained using cross-layer VoIP load control for an attenuation of 22 dB in a tropical area. Comparison to the non-cross-layer scenario. 115

6.16 Hybrid Architecture 117

6.17 FEC location in the protocol stack 118

6.18 MPE-FEC Frame and the MPE encapsulation process 119

6.19 Simulation Flow Diagram 122

6.20 PER as a function of thr_s (dB) for several Amazon areas (rain and clear sky conditions), and different super-frame length durations 123

6.21 PER as a function of thr_s (dB) and for different FEC code rates using RS ($SF_{dur} = 200$ ms) 124

6.22 BW efficiency as a function of thr_s (dB) for several Amazon areas (rain and clear sky conditions) 125

6.23 Cumulative distribution of PSNR. 126

6.24 Average time length of PSNR persistence under a given threshold. 126

6.25 Mean PSNR versus threshold offset for different Amazon areas and rain types. 127

6.26 Mean duration of PSNR corruption depending on the thr_s and for different SF_{dur} 128

6.27 QPSK/16QAM Hierarchical Modulation Scheme, where noise emulates the symbols received by QPSK users 131

6.28 Cross-layer Unequal Protection Scheme 133

6.29 Comparison in terms of Video Distortion between the cross-layer optimization and the independent non-cross-layer optimizations. 137

6.30 Evolution of DA-UEP (redundancy allocated to each video layer, left axis) and HM (λ , right axis) parameters for the cross-layer optimization. 138

6.31 Performance of the UEP Cross-Layer for different LMS scenarios: (a) Open area; (b) Intermediate tree shadowing (ITS); (c) Heavy tree shadowing (HTS). 139

6.32 Cross-layer interaction and methodology conclusions: Layered Video Multicasting 140

6.33 Cross-layer interaction and methodology conclusions: Adaptive VoIP Codecs-Centralized 140

6.34 Cross-layer interaction and methodology conclusions: Adaptive VoIP Codecs-Distributed 141

6.35	Cross-layer interaction and methodology conclusions: High layer-FEC protection and PHY layer optimization	142
6.36	Cross-layer interaction and methodology conclusions: UEP optimization . .	143

List of Tables

3.1	Situation of temperate and tropical areas and satellite features.	18
3.2	Comparison of the percentage of time the shown attenuation is exceeded at K_a frequency band	19
3.3	Mobile WiMAX Applications and QoS	27
3.4	QoS mapping between DVB-S2 and WiMAX Classes of Service	31
6.1	Delay and MOS for different ways to distribute video layers, for fairness parameter $\alpha = 0$	98
6.2	Codecs considered for RTCP-driven scenario	101
6.3	Codecs considered for the AMR-WB based scenario	102
6.4	Relevant Codec Delays	104
6.5	Relationship between R-value and MOS	105

List of Acronyms and Abbreviations

AAA	Authentication, Authorization and Accounting	CBR	Constant Bit Rate
AAS	Adaptive Antenna System	CCM	Continuous Coding and Modulation
AC	Adaptive Coding	cdf	cumulative distribution function
ACM	Adaptive Coding and Modulation	CGC	Complementary Ground Component
ACQ	ACQuisition	CLASS	Cross-Layer Signaling Shortcuts
AF	Assured Forward	CoS	Class of Service
API	Application Program Interfaces	CRA	Continuous Rate Assignment
APP	Application	CSC	Common Signaling Channel
APSK	Asymmetric PSK	CSI	Channel State Information
ARQ	Automatic Repeat Request	DAMA	Demand Assigned Multiple Access
ASN	Access Service Network	DA-UEP	Dependency Aware Unequal Error Protection
ATM	Asynchronous Transfer Mode	DBA	Dynamic Bandwidth Allocation
AWRR	Adaptive-WRR	DiffServ	Differentiated Services
BE	Best Effort	DistFair	Distributed Fairness Capacity Request
BRASIL	Broadband to Rural America over Satellite Integrated Links	DPC	Dirty Paper Coding
BS	Base Station	DVB	Digital Video Broadcasting
BSM	Broadband Satellite Multimedia	DVB-RCS	DVB-Return Channel via Satellite
CA	Correlated Area		

DVB-RCS+M	DVB-RCS Mobile	ITU	International Telecommunication Union
DVB-S	DVB-Satellite	ITU-R	ITU Radiocommunication sector
DVB-S2	DVB-Second Satellite generation	KKT	KarushKuhnTucker
DVB-SH	DVB-Satellite to Handheld	LDPC	Low Density Parity Check Code
EF	Expedite Forward	LL	Link-layer
ertPS	extended-rtPS	MAC	Medium Access Control
ETSI	European Telecommunications Standards Institute	MAC-CPS	MAC-Common Part Sublayer
FA	Foreign Agent	MAC-CS	MAC-Convergence Sublayer
FairDB	Fair Delay Balance	MBS	Multibeam Broadband Satellite
FCA	Free Capacity Assignment	MF-TDMA	Multiple Frequency Time Division Multiple Access
FDD	Frequency Division Duplex	MIMO	Multiple-Input Multiple-Output
FEC	Forward Error Correction	MulFairDB	Multicasting Fair Delay Balance
FRR	Fast Round Robin	MPDU	MAC packet data unit
FTP	File Transfer Protocol	MPE	Multi-Protocol Encapsulation
GEO	Geostationary Earth Orbit	MPEG	Moving Picture Experts Group
GS	Generic Stream	MOS	Mean Opinion Score
GSE	Generic Stream Encapsulation	NBS	Nash Bargaining Solutions
GW	Gateway	NCC	Network Control Center
HA	Hierarchical Allocation	NE	Nash Equilibrium
HA	Home Agent	nrtPS	non-real-time Polling Service
HARQ	Hybrid ARQ	NSP	Network Service Provider
ICMP	Internet Control Message Protocol	NUM	Network Utility Maximization
IntServ	Integrated Services		
IP	Internet Protocol		

0. List of Acronyms and Abbreviations

OBP	On-Board Processors	RU	Remote User
OFDM	Orthogonal Frequency Division Multiplexing	SAC	Satellite Access Control
OFDMA	OFDM Access	SC	Superposition Coding
OSI	Open Systems Interconnection	SF	Super-Frames
PHB	Per-Hop Behavior	SI-SAP	Satellite Independent-Service Access Point
QAM	Quadrature Amplitude Modulation	SNIR	Signal to Noise plus Interference Ratio
QID	Queue IDentifier	ST	Satellite Terminal
QoS	Quality of Service	STQRM	ST QID Resource Manager
QPSK	Quadrature PSK	SVC	Scalable Video Coding
PER	Packet Error Rate	SYNC	SYNChronization
pdf	probability distribution function	TBTP	Terminal Burst Time Plan
PHY	Physical	TCP	Transmission Control Protocol
PO	Pareto Optimal	TCT	Timeslot Composition Table
PSK	Phase-Shift Keying	TDD	Time Division Duplex
PSNR	Peak Signal to Noise Ratio	TDM	Time Division Multiplex
RBDC	Rate Based Dynamic Capacity	TDMA	Time Division Multiple Access
RCS	Return Channel via Satellite	ToS	Type of Service
RSVP	Resource ReserVation Protocol	TRF	TRaFfic burst
rtPS	real-time Polling Service	TS	Transport Stream
RTT	Round Trip Time	UE	User Equipment
RRA	Radio Resource Allocation	UGC	Unsolicited Grant Service
RRC	Radio Resource Control	VBDC	Volume Based Dynamic Capacity
RRM	Radio Resource Management	VCM	Variable Coding and Modulation

VoIP	Voice over IP
WiMAX	Worldwide Interoperability for Microwave Access
WRR	Weighted Round Robin

Chapter 1

Introduction

1.1 Motivation

The demand for high-speed networking, driven mainly by the rapid expansion of the Internet, has been growing at an exponential rate. A wide range of technologically innovative wireline and wireless solutions already offer competitive broadband connectivity. Communication satellites emerge as an attractive solution in providing broadband connectivity to a variety of users thanks to its inherent global coverage.

The broadcast nature of satellites makes them the natural choice for multicasting services, for interconnecting geographically distributed high-speed networks, and for providing multimedia services to both home and business users. In order to accommodate the heterogeneity of users and multimedia traffic, broadband satellite networks are moving from the crowded K_u band to the more suitable K_a band. However, this higher frequency band imposes challenging channel conditions, particularly the capacity drop. Such capacity drop leads to bigger delay, jitter and packet loss.

1.1.1 Wireless/Satellite Communications

In wired scenarios, the capacity drop is due to the congestion collapses suffered by the network, which has led to the development of congestion and admission control mechanisms at transport layer. Congestion control based on Transmission Control Protocol (TCP), is known to provide reliability to elastic application over Internet Protocol (IP) based networks, i.e. traditional data services (such as file transfer, web browsing and email services). On the other hand, admission control aims at fulfilling the strict QoS requirements of inelastic services (delay/jitter-sensitive).

In wireless terrestrial networks, main efforts have been focused on smart tailored resource management schemes, depending on the size and coverage of the network, power limitations of terminals, etc. In these types of networks, the effect of shadowing in typical Rice/Rayleigh channels cause a capacity drop when transmission schemes adds redundancy in order to protect data [1]. Capacity or quality might also decrease when the number of users increases, and the resource manager needs to reallocate power in order to avoid the interferences among users [2].

In multi-beam broadband satellite communication networks, it is vital to design network architectures to be cost-competitive with respect to the terrestrial media of wireless and fiber. Since on-board resources such as power, bandwidth, receivers and spotbeams are scarce and expensive, it is critical to share them efficiently among as many users as possible. The issue of providing fair and efficient resource allocation and channel access should be considered together with system performance guarantees on throughput and delay. In particular, it is foreseen the use of K_a (20-30 GHz) to compete with terrestrial networks. However, it is also well known that the problem of capacity drop becomes more challenging at these bands, where the satellite might suffer from strong bandwidth decrease. Specially if satellite bandwidth is not enough (under bad weather conditions) to satisfy all requests, and each Satellite Terminal (ST) needs to compete for an adequate amount of bandwidth in order to transmit data and fulfill an expected Quality of Service (QoS) level. For example, in the case of an adaptive physical layer, bad weather conditions can result in significant capacity drops [3]. While existing typical algorithms such as TCP to counteract congestion effects in wired networks, the problem of substantial capacity drop due to fading events has not been so far properly addressed. In this case, control over fairness is a key issue for controlling the overall network operation while maximizing user satisfaction.

Such higher frequency bands are seen as a solution to cope not only with the saturation at lower bands but mostly with the explosive growth of inelastic traffic applications in these networks, such as real time and video streaming. In order to accommodate the heterogeneity of users, multimedia traffic constraints and channel conditions, adoption is underway of essential technologies at Physical (PHY), Medium Access Control (MAC), Network (or IP) Transport and Application (APP) layer such as Adaptive Coding and Modulation (ACM), time and frequency multiplexing, data scheduling, higher layer error protection, layered video multicasting or adaptive Voice over IP (VoIP) codecs. However, the interplay of these technologies for an optimal service provision is not solved yet.

1.1.2 Cross-layer

Open Systems Interconnection (OSI) layered protocols and algorithms are sometimes not able to adapt to new multimedia contents requirements, which are transmitted through the often scarce resources of wireless networks. This is because, at upper layers, multimedia applications do not consider the mechanisms provided by the lower layers for error protection, scheduling, resource management, etc, and vice versa. Such design leads to a suboptimal performance.

Cross-layer designs have been demonstrated to jointly optimize the overall network performance by taking advantage of the available mechanisms at different layers. In this dissertation, we choose the cross-layer methodology in order to bridge the gap between different layers and allow the system adapt and cope with physical layer variability. Although cross-layer design has been around for some time, formal studies and the actual term for these techniques are quite recent. Basically, cross-layer design provides a mechanism for increased awareness between layers, without destroying the layered stack paradigm. It rather redefines a stack of independent layers into a stack of inter-dependent layers.

1.2 Outline of Dissertation

In general terms, the focus of this dissertation is on the joint optimization design across layers of a multibeam broadband satellite air interface that implements ACM at the PHY layer. We cover several aspects related to the networking optimization, including interoperability with terrestrial networks (thus also covering hybrid networking). The outline of each of the chapters is as follows:

- Chapter 1: The present chapter, gives the motivation, outline and context of this dissertation.
- Chapter 2: The chapter overviews the cross-layer design paradigm which arise in many different wireless and satellite network scenarios. We open the question of cross-layer optimization and its effectiveness in providing an improved solution with respect to traditional layered protocols. This chapter includes also a taxonomy for cross-layer according to type of interaction, layers involved, performance criteria and signaling methodology used to evaluate and optimize the network communication.
- Chapter 3: The chapter introduces the satellite and hybrid (terrestrial and satellite) systems that are considered in this dissertation, and in particular, we define the channel and the system models. From the channel point of view, terrestrial and satellite models are considered. The latter case is studied according to the weather conditions in tropical and temperate areas. Regarding the system model, a brief technical overview of the main standards on which most results are based is also included: DVB-Second Satellite generation (DVB-S2), DVB-Return Channel via Satellite (DVB-RCS), DVB-Satellite to Handheld (DVB-SH) and Worldwide Interoperability for Microwave Access (WiMAX). The QoS concept, the requirements and the traffic models evaluated in this dissertation are discussed for the satellite and terrestrial models. Moreover, the chapter also includes our design of a detailed hybrid architecture, which is able to provide service to rural and tropical areas, and in particular to the Amazon Rainforest.
- Chapter 4: In this chapter, we address the multicast resource allocation of the air interface in multibeam mobile broadband satellite/hybrid systems. The chapter deals with the two main transmission schemes considered in this dissertation: adaptive physical layer (or ACM) and Superposition Coding (SC). Moreover, two mobility scenarios are considered depending on channel variations: a pedestrian/fixed scenario and mobile scenario with slow and fast channel variations respectively. From a PHY and MAC layer perspectives, we aim at optimizing the multicast transmission by means of an allocation algorithm based on SC, which is able to deal with fast mobility. We maximize the capacity while fulfilling a minimum rate per multicast group, and we perform a comparison with a time-slot based algorithm. We finally justify that a combination of SC and ACM, based on the standard recommendations of DVB-SH is suitable for resource allocation over mobile scenarios.
- Chapter 5: The chapter addresses the challenging problem of capacity drop of satellite capacity at K_a Band when bad weather conditions are affecting the satellite

link (i.e. from satellite to satellite terminal). In particular, we focus on how the resource allocation at MAC/IP scheduler follows the adaptivity of the physical layer (i.e. the ACM technology). In order to do so, we propose a cross-layer DiffServ architecture to optimize the interplay of PHY, MAC and IP layers. Two different designs are implemented, one for the forward link (DVB-S2) and a second one for the return link (DVB-RCS), which take advantage of advanced mathematical tools; Network Utility Maximization (NUM) and Game Theory, respectively. We use an adaptive fairness methodology to distribute the available resources. In order to do so on the return link, we propose to partially decentralize the bandwidth allocation of the multiple access air interface, achieving high bandwidth allocation efficiency and reducing the signaling overhead.

- Chapter 6: The chapter extends the channel conditions adaptivity to higher layers, i.e. Transport and APP layers. Applications such as VoIP, Videoconferencing or Videostreaming, known to require stringent QoS in terms of delay and packet loss, are transmitted among different mobility scenario and channel conditions. In this chapter, the satellite dependant layer (MAC and PHY) are successfully jointly optimized with the efficient features at Transport and APP layers: adaptive VoIP rates and layered video. Some of the designs presented in this chapter aim at solving the challenges of dealing with heavy weather conditions, typical of tropical areas, and therefore, they are part of the Brasil Project.
- Chapter 7: The last chapter concludes the dissertation. We not only deal with the goals of our research contribution and summarize the main results obtained, but we also define a cross-layer methodology framework for adaptive satellite systems. We finally propose the future lines of work.

1.3 Research Contributions

The main goals of this dissertation are the design and study of cross-layer solutions for adaptive broadband satellite systems in mobile scenarios. Our work is backed by many published contributions, which are detailed and organized per chapter as follows.

1.3.1 Chapter 3

The detailed cross-layer hybrid architecture has been published in the ITU-R standard:

- David Pradas and M.A. Vázquez Castro, "Cross-layer Design over Hybrid Satellite-WiMAX Networks", *ITU-R WP4B*, Spain, May 2008.

Note also that our hybrid architecture has been used for different cross-layer designs of this dissertation and published in several publications: those specified in Chapter 6.

1.3.2 Chapter 4

Two conference papers:

- David Pradas and M.A. Vázquez Castro. "Cross-layer Rate Allocation of Multicast Transmission over Hybrid DVB-SH", in *Proc. of International Workshop on Cross-layer Design (IWCLD)*, Mallorca, Spain, Jun. 2009.
- David Pradas and M.A. Vázquez Castro, "Multicast Transmission Optimization over Hybrid DVB-SH Systems", in *Proc. of the IEEE Vehicular Technology Conference (VTC)- Spring*, Barcelona, April 2009.

1.3.3 Chapter 5

Three journal papers:

- David Pradas, M. A. Vázquez Castro and Min-Su Shin, "Distributed Fair Bandwidth Request Solution and Corresponding Cross-layer Architectural Design", submitted to *ETRI Journal*, 2011.
- David Pradas and M. A. Vázquez Castro, "Study and Implementation of Bandwidth Request Games for Interactive Satellite Networks", submitted to *IEEE Transaction on Broadcasting*, 2011
- David Pradas and M. A. Vázquez Castro, "NUM-based Fair Rate-Delay Balancing for Layered Video Multicasting over Adaptive Satellite Networks", *IEEE Journal on Selected Areas in Communications (JSAC)*, no. 5, vol. 29, p. 969, May 2011.

Moreover, a patent is currently being processed:

- David Pradas, M. A. Vázquez Castro and J. E. Barceló, "Distributed MAC for Next Generation Satellite Networks", by the Research Council for Industrial Science & Technology, Korea.

1.3.4 Chapter 6

One journal paper, one chapter of a book and five conference papers. Note that relevant content of the journals of the previous chapter is detailed in this chapter.

- M.A. Vázquez Castro and David Pradas, "VoIP Cross-Layer Load Control for Hybrid Satellite-WiMAX Networks", in *IEEE Wireless Communications*, vol. 15, pp. 32-39, Jun. 2008.
- M. A. Vázquez Castro, David Pradas and Catherine Morlet, "Distributed Cross-Layer Approaches for VoIP Rate Control over DVB-S2/RCS", *Advances in Mobile and Wireless Communications: Views of the 16th IST Mobile and Wireless Communication Summit*, pp. 315-330, Ed: Springer, 2008.

- David Pradas, Amine Bouabdallah, Jérôme Lacan, M.A. Vázquez Castro and Michel Bousquet. "Cross-layer Optimization of Unequal Protected Layered Video over Hierarchical Modulation", in Proc. IEEE GLOBECOM, Dec. 2009.
- David Pradas, Paolo Barsocchi, Lei Jiang, Francesco Potortì, and M.A. Vázquez Castro. "Satellite PHY-layer Selector Design for Video Applications in Tropical Areas", in Proc. of International Workshop on Satellite and Space Communications (IWSSC), Siena, Italy, Sep. 2009.
- David Pradas, Paolo Barsocchi, Lei Jiang, Francesco Potortì and M.A. Vázquez Castro. "Cost-efficient design of hybrid network for video transmission in tropical areas", in Proc. of the IEEE Vehicular Technology Conference (VTC) - Spring, Barcelona, April 2009.
- M. A. Vázquez Castro, David Pradas and Sastri Kota, "VoIP Transmission Cross-Layer Design over Satellite-WiMAX Hybrid Network", in *Proc. Military Communications Conference (MILCOM 2007)*, Orlando, Florida, USA, Oct. 2007.
- M. A. Vázquez Castro, David Pradas and Catherine Morlet, "Cross-Layer Rate Control of VoIP over DVB-S2/RCS", in *Proc. 16th IST Mobile & Wireless Communications Summit*, Budapest, Hungary, Jul. 2007.

Moreover, our research contribution has been published in two standards. A cross-layer architecture and a cross-layer design for ITU-R and ETSI standards respectively:

- David Pradas and M.A. Vázquez Castro, "Cross-layer Design over Hybrid Satellite-WiMAX Networks", *ITU-R WP4B*, Spain, May 2008.
- SI-SAP primitives defined in ETSI TS 102 463 for VoIP Cross-Layer Design: "Survey on the envisaged explicit cross-layer approaches for BSM", 2009.

Additional result have been obtained thanks to the contribution included in Chapter 6 and the related publications. Although these results are not included in this dissertation, they have been published in two conference papers:

- José Radzik, Oriol Vidal, David Pradas, Michel Bousquet and M.A. Vázquez Castro, "Cross-layer optimization of hybrid satellite network over the Amazonian region", in *Proc. of International Workshop on Satellite and Space Communications (IWSSC)*, Siena, Italy, Sep. 2009.
- José Radzik, Oriol Vidal, David Pradas, Michel Bousquet and M.A. Vázquez Castro. "Performance of Hybrid Satellite-WiMAX Networks based on DVB-RCS satellite systems for rural and remote areas", in *Proc. of International Communications Satellite Systems Conference (AIAA-ICSSC)*, Edinburgh, UK, Jun. 2009.

1.4 Thesis Context

This document reflects the path taken in the studies towards obtaining the doctorate degree. Let us introduce that this thesis has been jointly supervised by the Universitat Autònoma de Barcelona (UAB) and the Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), and thus it has included stays between Barcelona (Spain) and Toulouse (France), location of both universities.

In the context of their research programme, within the field of satellite communications, the UAB and the French Space Agency (Centre National d'Études Spatiales, CNES) decided to finance the study by means of a fellowship collaboration. The agreement included also an additional work supervision performed by the TeSA (Telecommunications for Space and Aeronautics) laboratory.

The idea of a collaborative thesis between all the parts was originally agreed within a SatNEx context, i.e. the European Satellite Communications Network of Excellence that was launched within the 6th Research Framework Programme of the European Commission. This project was designed to foster cooperative work between universities and research institutes across Europe so that they can play a crucial role in supporting European industry. Precisely, previous research studies between the parts involved in this thesis triggered off the idea of designing an optimization framework for adaptive satellite systems, which was the original aim of the thesis. Further discussion led to choosing the versatility of cross-layer designs as a way to improve the satellite performance.

Within SatNEx, a wide range of joint activities were performed, where the collaborative tools like the personal exchange missions, work meetings as well as the satellite-based video-conferencing platform provided the framework for fruitful collaborative research with other partners, such as the Italian National Council of Research (ISTI-CNR), the Office National d'Études et de Recherches Aérospatiales (ONERA) and other entities contributing to the Broadband to Rural America over Satellite Integrated Links (BRASIL) project. These parallel collaborations worked as a complementary platform to develop some of the cross-layer designs included in this dissertation. Additionally, some designs are the result of European projects, such as European Space Agency (ESA) projects, and international projects, e.g. the Electronics and Telecommunications Research Institute (ETRI).

This document is not the result of a single project but rather a collection of knowledge obtained from several projects, some with more practical-oriented studies while others rather more theoretical in nature. Furthermore, the topics covered in this document have been studied from both a theoretical and a practical point-of-view due to this broad range of activities.

Chapter 2

Cross-layer Design Paradigm

2.1 Introduction

In recent years, the research focus has been to adapt existing algorithms and protocols to the rapidly varying and often scarce resources of wireless networks. However, these solutions often do not provide adequate support for new traffic applications in crowded wireless/satellite networks, when interference is high, stations are mobile or under bad weather conditions. This is because the resource management, adaptation, and protection strategies available in the lower layers of the stack—the PHY, MAC, and Network/Transport layers—are optimized without explicitly considering the specific characteristics of new multimedia applications, and conversely, applications such as adaptive VoIP or layered video do not consider the mechanisms provided by the lower layers for error protection, scheduling, resource management, and so on. This "layered" optimization leads to a simple independent implementation, but results in suboptimal (objective and/or perceptual quality) performance. Alternatively, under adverse conditions, transmitters need to optimally adapt their transmission strategies jointly across the protocol stack in order to guarantee a predetermined quality at the receiver.

The cross-layer framework is of fundamental importance in order to adapt the different strategies available at the various OSI layers. Moreover, it not only leads to improved multimedia performance over existing satellite networks, but also provides valuable insights into the design of next-generation algorithms and protocols. The proposed cross-layer approach does not necessarily require a redesign of existing protocols [4], and can be performed by selecting and jointly optimizing the application layer and the strategies available at the lower layers, such as admission control, resource management, scheduling, error protection, and power control.

2.2 Taxonomy of Cross-Layer Methodologies

In this dissertation, we deal with many cross-layer designs of different types. Not only the layers involved or the mechanisms optimized can change between designs, but also the optimization criteria, the nature of the interaction (either explicit or implicit) or the flows direction. In order to highlight the cross-layer versatility, next, we provide a classification of

cross-layer methodologies, based on a review of current literature and [5, 6], which offers a useful classification of cross-layer mechanisms based on the direction of information flow within the protocol stack.

2.2.1 Implicit and explicit cross-layer design

An important aspect for differentiating cross-layer methods is the presence or absence of signaling of the internal protocol state between protocol layers. This may be used as a basis to differentiate between implicit and explicit cross-layer design/techniques, as summarized below.

In case of implicit cross-layer design, there is no exchange of information between the layers during run-time, but in the design criteria cross-layer interactions are taken into account. Layers are designed to complement each other and unnecessary duplicate implementations of functionalities are eliminated, preventing for example MAC level collisions in case of network flooding, and applying only one retransmission policy and not one at the Transport layer and one at the MAC layer.

In the explicit case, feedback information is used to communicate and optimize the parameters used at different layers. To exchange this signaling information, a cross-layer signaling scheme is introduced.

As an example of implicit and explicit optimization, we refer to two different approaches to transmit video services. In the explicit case the application generates single layer video traffic with a given rate and passes it to the MAC layer. Depending on the channel state the MAC layer may ask the application explicitly to adjust the source rate. In the implicit case, the application generates multiple descriptors of the video stream and passes them to the MAC. The MAC can decide how many descriptors are transmitted depending on the available resources. If the resources are limited, some randomly chosen descriptors are not transmitted. The receiver resolves the video image from the received descriptions by using advanced decoding techniques. This way the MAC layer does not have to inform the source of the available resources, but still the transmitted flow can be adapted to the available resources.

The following criteria can be used in principle for explicit and for implicit cross-layer design.

2.2.2 Based on optimization criteria

According to the optimization criteria, cross-layer optimization can be essentially tackled in two different ways.

2.2.2.1 Global-performance-optimization-oriented

The approach of a global-performance-oriented cross-layer methodology consists of identifying global performance parameters to be optimized through cross-layer signaling flow (e.g throughput). In this approach, a global performance parameter is selected and intervening protocols in the communication are reviewed in order to identify how their behavior can be improved so as the selected parameter is maximized (minimized).

In this approach, it is easy to see that the criteria of evaluation would simply consist of comparing the value of the global parameter(s) selected before using any cross-layer techniques and after using them. The degree of relevance of the improvement cannot be generalized at this point. For example, it is extremely different to focus on improving throughput to improving complexity. Moreover an eventual improvement should not be at the expenses of affecting negatively on other parameters

In this approach, the impact on system design of a cross-layer optimization is hardly parameterized from a general viewpoint since it will depend on the global parameter to be optimized. However, it can be expected that this technique would have a higher impact on system design than other techniques described hereafter. The reason is that a global performance parameter depends normally on a number of communication subsystems. However, it would be feasible to constrain this optimization to a number of reduced communication subsystems susceptible of being modified.

2.2.2.2 Protocol-optimization-oriented

The approach of a protocol-optimization-oriented methodology consists of optimizing per-layer protocol and related mechanisms performance. In this approach, current widely used protocols are chosen and the performance of its different functionalities analyzed. Performance improvement is then pursued by tuning involved procedures that could benefit from a cross-layer optimization (either implicit or explicit but mostly explicit).

In this approach, the criteria of evaluation not only would consist of characterizing a protocol performance before and after applying cross-layer techniques. In fact, a second criterion to assess is an overall performance evaluation since it might happen that an individual protocol is significantly optimized but the overall system performance is not worth the cost of modifying the protocols to include the cross-layer optimizations.

In this approach, the impact on system design of a cross-layer optimization is maybe less than that of a global-performance optimization oriented approach. In this case, a possible solution would be that current layered protocols would carry the signaling across layer with a minimum impact in system design.

2.2.3 Based on direction of information flow

Another cross-layer classification method considers the direction of the cross-layer information flow. This approach is appropriate to an explicit cross-layer design and primarily focuses on optimizing the information flow. Such an approach should allow efficient ad hoc optimizations for each layer and/or protocol, compatible with future versions of current protocols. Moreover, it could provide an optimized cross-layer mechanism that could be used for different kinds of optimization, rather than defining isolated cases that are optimized for a particular communication system.

Developing an integrated cross-layer framework may be important to the satellite community, since it not only leads to improved multimedia performance over existing networks, but could also provide valuable insights into the design of next-generation algorithms and protocols for satellite multimedia systems.

This approach does not follow a traditional layered design. History has shown that

devoting time to build a solid framework (like the OSI reference model) failed, when the more integrated TCP/IP protocol stack has succeeded. However, if the Internet continues its current gradual evolution, this may be too slow to be able to satisfy the immediate needs for cross-layer satellite optimizations.

One criterion for the evaluation of cross-layer methods is the efficiency, i.e., a flow of information would be considered more efficient if a maximum of information is available to other layers when passing a minimum set of parameters or signaling. Another criterion is the evaluation of the chosen ad hoc performance parameter that benefits from the information flow.

The impact on system design is a key constraint when designing to achieve efficient information flow. A cross-layer approach does not necessarily require a re-design of existing protocols, and can be performed by selecting and jointly optimizing the upper layers and the strategies available at the lower layers, such as admission control, resource management, scheduling, error protection, power control, etc.

2.2.3.1 Bottom-up approach

This approach seeks to design an efficient information flow among layers from the lowest layers up to the application one. In a satellite system implementing ACM, this approach seems certainly appropriate since the idea would be to inform upper layers with the dynamics and adaptability that is taking place at the physical layer.

It is interesting to note that in terrestrial wireless, this cross-layer solution may not be optimal for multimedia transmission, due the incurred delays and unnecessary throughput reductions. However, in broadband satellite communications, slow channel variations would allow for a timely bottom-up flow in such a way that upper layers would have enough time to react.

This approach requires defining general per-layer parameters that could be useful to the upper layers. Moreover protocols operating at each layer should be reviewed assuming that all cross-layer parameters flowing up from lower layers are instantaneously available. It would be up to the designer the selection of the parameters that optimize particular procedures or overall protocol performance.

2.2.3.2 Top-down approach

A top-down approach designs an efficient information flow among layers from the application layer down to the physical layer. This can be seen as an application-centric approach: applications indicate their expectations of required network behavior, and lower-layers can then use this information to optimize lower layer parameters.

There are drawbacks with this approach. One problem is that applications are frequently unaware of the network paths over which they operate. They are therefore unable to express their requirements in a way that maps easily to the capabilities of specific lower-layers. Moreover, applications typically operate over longer time-scales with coarser data granularities (multimedia flows or blocks of data) than those used at lower layers (operating on bits or frames). It is therefore non-trivial to perform adaptive source-channel coding trade-offs, given the time-varying channel conditions and the fact that multimedia

applications cannot be expected to adapt instantaneously their behavior to achieve an optimal performance.

While lower layers can benefit from notifications of requirements (capacity estimates, delay bounds, FEC/ARQ needs, priority, etc.) this does not provide a complete solution. For example, it has limited benefit for a satellite system implementing ACM, since the upper layers may not be able to influence usefully the behavior of lower layers, rather, the channel dynamics require upper layers to adapt themselves.

2.2.3.3 Hybrid approach

There are cases in the related literature in which system level constraints are refined in a top-down fashion, while target architecture performance is abstracted in a bottom-up fashion and a "meet in the middle" approach decides the final optimization.

This approach may not be attractive if a general methodology is desired. The fact of exhaustively trying all the possible strategies (and their parameters) in order to choose the composite strategy that would lead to the best quality performance is impractical due to the associated complexity. Even when designing a cross-layer methodology, general software architecture principles such as information hiding, modularization, and separation of concerns should be considered.

2.2.4 Based on layers involved

This approach would focus on individual and well characterized problems that could be solved by joint optimization of a certain number of layers. As an extension, this approach would model the protocol stack as a whole, rather than a collection of individually modeled layers. Such model would have tunable parameters that affect the operation of each layer to achieve globally specified performance and efficiency.

It can be noticed that this approach can be actually equivalent to the approach based on protocol optimization since the optimization of a specific protocol involves a given number of layers.

2.3 Taxonomy of cross-layer signaling methods

This dissertation includes detailed signaling between layers in most of the designs (the two cross-layer designs of Chapter 5, the design of Section 6.2.1 and the centralized design of Section 6.2.2) and more simple signaling in the others. As it can be seen through the document, we have specified the signaling between MAC and Network (or IP) layers by using the Broadband Satellite Multimedia (BSM) reference architecture, its methodology and the interaction between Satellite Independent and Satellite Dependent layers, which are introduced at the end of this section.

Although we have also proposed cross-layer signaling between the rest of the layers, we do not specify the methodology and thus it is still open. A number of signaling methods to carry cross-layer information across layers have been proposed to date [7]. A non-exhaustive list is given in the following.

2.3.1 Based on Packet headers

This method makes use of IP data packets as in-band message carriers with no need to use a dedicated internal message protocol. However, an IP packet normally can only be processed layer by layer, and it is not easy for higher layers to access to the IP-level header. Furthermore, the conceptual bottom-up "pipe" seems excessive in most cases. This method can be visualized like a *signaling pipe* (see Figure 2.1(a), extracted from [7]).

2.3.2 Based on Internet Control Message Protocol (ICMP) messages

Internet Control Message Protocol (ICMP) is a widely deployed signaling protocol in IP-based networks. Compared to the "pipe" described above, this method tries to "punch holes in the protocol stack" and propagate information across layers by using ICMP messages as shown in Figure 2.1(b). In this scheme, desired information is abstracted to parameters, measured by corresponding layers wherever convenient. A new ICMP message is generated only when a parameter changed beyond the thresholds. Since cross-layer communications are carried out through selected "holes" not a general "pipe", this method seems more flexible and efficient. Furthermore, this method is more mature since it has been implemented on Linux Operating System with Application Program Interfaces (API) developed. However, an ICMP message is always encapsulated in an IP packet, and this indicates that the message has to pass by Network Layer even if the signaling is only desired between MAC layer and Application layer.

2.3.3 Based on a network service

In this scheme, channel and link states from PHY Layer and MAC layer are gathered, abstracted and managed by third parties; distributed servers. Interested applications then access to the servers for their required parameters from the lowest two layers. Although it is not a cross-layer signaling scheme within a terminal, we can deem it complementary to the two above schemes, as further implementation problems are considered in parameter definition, abstraction, coding, and decoding. However, any intensive use of this method would introduce considerable signaling overhead and delay over a radio access network.

2.3.4 Based on local profiles

In this approach, local profiles are used to store periodically updating information from the nodes. Cross-layer information is abstracted from each necessary layer respectively and stored in separate profiles within the nodes. Other interested layer(s) can then select the profile(s) to fetch the desired information.

2.3.5 Concept mode: CLASS

From the above discussion, a couple of major drawbacks of the existing methods can be identified:

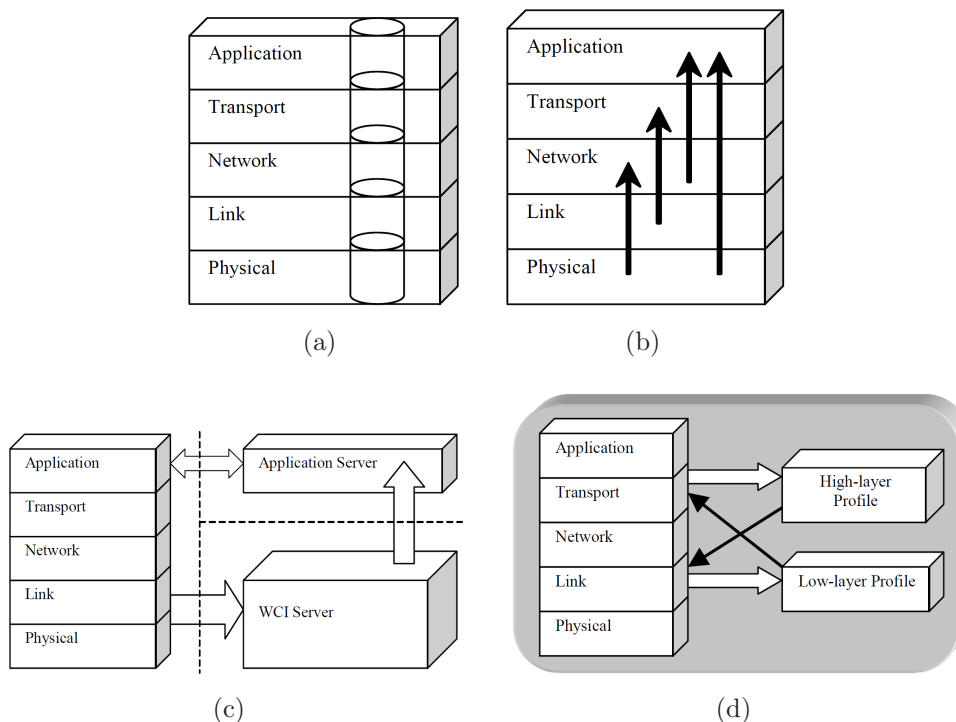


Figure 2.1: Cross-layer signaling methods: (a) Based on packet headers; (b) Based on ICMP; (c) Based on a network service; (d) Based on local profiles.

- The signaling propagation paths across the protocol stack are not efficient. The layer-by-layer propagation approach just follows the data propagation mode. Consequently, the intermediate layers have to be involved even if only the source layer and the destination layer are actually targeted. This will cause unnecessary processing overhead and propagation latency.
- The signaling message formats are either not flexible enough for active signaling in both upward and downward directions, or not optimized for different signaling inside and outside the Protocol Stack respectively. Furthermore, the desired message formats should be scalable enough for rich signaling more than cross-layer hints and notifications.

Therefore, it could be considered a new method proposed in [7], named Cross-Layer Signaling Shortcuts (CLASS). It is an improvement of the ICMP to "punch holes in the protocol stack" so that local out-of-band shortcuts are created for the exchange of signaling among non-adjacent layers. Practically, a lightweight ICMP version is used for internal cross-layer signaling, while ICMP is used for external cross-layer signaling.

2.3.6 BSM particular case: SI-SAP Primitives

The European Telecommunications Standards Institute (ETSI) working group BSM has defined a reference architecture [8] where the protocol stack is divided into two main blocks

connected by the Satellite Independent-Service Access Point (SI-SAP) [9]: the upper part of protocols is characteristic of the Internet and independent of the satellite technology (i.e., IP-based protocol suite), while the lower part depends on the satellite system implementation. Defined primitives [10] are used to exchange signaling on the control plane across SI-SAP between these two blocks of protocol layers. The BSM standard envisages a framework for QoS support in the mapping of layer 3 (Network/IP layer) traffic classes and layer 2 (MAC layer) technology-dependent allocation methods.

Note that if SI-SAP primitives are used, a complex scheme of signaling should be considered involving primitives from other layers and coordinated SI-SAP primitives from layer 3 to layer 2. It seems that ICMP and CLASS potentialities could be exploited at the SI-SAP level by creating suitable short-cuts; this could make easier the implementation of the proposed cross-layer schemes.

The performance and the architecture of SI-SAP interface architecture and primitives will be fully explained within the cross-layer designs.

2.4 Optimal Solution Challenges

Finding the optimal solution of a cross-layer optimization problem is difficult because:

- Deriving analytical expressions for the design objectives (e.g., maximize video quality or minimize PER), the resources to be distributed (bandwidth, power, etc.) or performance criteria (delay, rate, fairness, etc.) as functions of channel conditions is very challenging, since these functions are nondeterministic (only worst case or average values can be determined) and nonlinear, and there are dependencies between some of the strategies of the different layers involved in the design. Since we need to map all the expressions into the corresponding utility or price values and to optimize the established utility pricing function, the task becomes even harder.
- The algorithms and protocols at the various layers are often designed to optimize each layer independently and often have different objectives. For instance, the PHY layer is concerned with symbols and depends heavily on the channel characteristics, while the application layer is concerned with semantics and dependencies between flows, and depends heavily on the multimedia content.
- The wireless channel conditions and multimedia content characteristics may change continuously, requiring constant updating of parameters.
- Formal procedures are required to establish optimal initialization, grouping of strategies at different stages (i.e., which strategies should be optimized jointly), and ordering (i.e., which strategies should be optimized first) for performing the cross-layer adaptation and optimization.

Finally, different practical considerations (e.g., buffer sizes, set of modulations or protection properties at the packet level) for the deployed satellite standard must be taken into account to perform the cross-layer optimization.

Chapter 3

Satellite and Hybrid Systems and Channel Models

3.1 Channel Models

3.1.1 Satellite Channel Model

Adaptive Multibeam Broadband Satellite (MBS) systems operating at K_a/K_u bands such as the one assumed in this dissertation are known to provide wide bandwidth and thus small size antennas can be used. However, as have we already highlighted that, at these bands, weather conditions are the main cause of transmission impairments, which can cause big capacity drops.

We consider the transmission and reception models of [11], including received Signal to Noise plus Interference Ratio (SNIR), which depends on the time and space variability, and takes into account inter-beam interferences, free space losses, antenna gain, and sky and rain attenuations among others. The behavior, characterization and the effects of the rain attenuation in a MBS system are some of the main challenges analyzed in this dissertation.

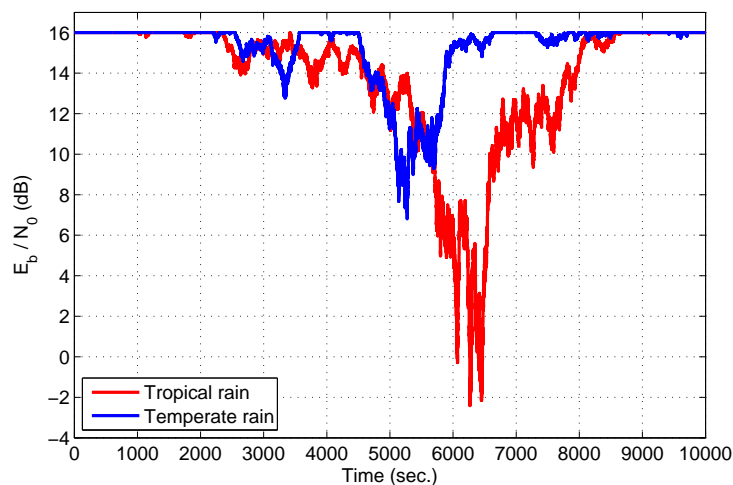
3.1.1.1 Rain Correlated Models

In this section, we present the rain channel model proposed in [11], which gathers both the time and space variability of the K_a band real channel. In particular, it is able to correlate the rain cell size with the geographical areas affected. Note that here we just give a short explanation.

Real channel conditions show a marked space correlation; i.e. the probability that very close locations undergo similar channel conditions is very high, and the probability decreases with the distance between the locations. Resulting on the definition of Correlated Area (CA): the geographical zone within which channel conditions are highly correlated at a given time. Which means that users located within a CA undergo similar channel attenuation, and users in different CAs experience channel attenuations that are virtually uncorrelated. The size of CAs (around a few dozens of Km), is relevant in order to assess the space variability of the channel within the coverage of one beam (up to 200 Km).

Table 3.1: Situation of temperate and tropical areas and satellite features.

	Temperate	Tropical
Locations	North of Italy	Amazon (Tabatinga and Caxias)
Satellite	EUTELSAT II F4	HISPASAT AMAZONAS-1
Type of orbit	Geostationary	Geostationary
Orbital position	7°E	61°W
Frequency band	K _u	K _u and C

**Figure 3.1:** Examples of channel conditions caused by a rain event in Temperate and Tropical areas.

From the propagation point of view, the size of the correlated areas is related to the spatial correlation of rain cells. Following an exponential model of the rain cell [12], cell radius ranges up to 30-50 Km for K_a band propagation. From the point of view of the practical design of the communication network, the location of CAs should be associated with centers of population, i.e. areas whose size is on the order of a typical rain cell and that are the main sources of traffic. As a result, each area would typically be associated to one of the main cities in the beam coverage and its surroundings.

3.1.1.1.1 Tropical and Temperate

In order to accurately design our cross-layer models and avoid the quality of service decrease under heavy weather conditions, we need for a system dimensioning and link budget assessment. Therefore, precise propagation models of both temperate and tropical areas need to be analyzed.

Among the different rain fading statistics analyzed in this dissertation, two are the most used to represent tropical and temperate areas as channel attenuations. The situation and the features of the satellites feeding the beam area are briefly detailed in Table 3.1.

Table 3.2: Comparison of the percentage of time the shown attenuation is exceeded at K_a frequency band

Rain area	Attenuation			
	5 dB	10 dB	15 dB	20 dB
Temperate	0.15	0.02	0.008	0.0001
Tropical	1	0.3	0.15	0.05

An example of temperate and tropical rain conditions are depicted in Figure 3.1. And the comparison in terms percentage of time a given attenuation is exceeded is shown in Table 3.2. The tropical data has been taken from the NASA K_a band Radio Frequency Atmospheric Propagation Experiment [13], from ONERA (Office National d'Études et de Recherches Aérospatiales), also known as the French Aerospace Lab, and related literature [14].

3.1.1.2 Land Mobile Satellite Models

For the LMS scenario, we assume different channel environments at S band [15], which are shown in Figure 3.2. The depicted scenarios are open area, intermediate and heavy tree shadowing and suburban area, for a mobile user that covers a distance of 10 km at a speed of 100 km/h , with an antenna elevation of 40° .

3.1.2 Terrestrial Propagation Channel Model

A terrestrial propagation model is also needed to evaluate some of the cross-layer design in hybrid networks. In particular, we assume a WiMAX terrestrial subnetwork.

The WiMAX standard [16] proposes COST 231 suburban model for WiMAX channels, but this empirical model is not developed for the 3.5 GHz band assumed here. In [17], it is concluded that COST 231 overestimates the path loss for this band, and a mean prediction error is proposed as a rectification. We consider the Modified COST 231 model since it is generally accepted as a suitable propagation model for mobile applications in the 1900 MHz band and is assumed, in this case, to be acceptable for 2500 and 3500 MHz.

3.2 System and QoS Models

3.2.1 Practical Standards Overview

3.2.1.1 DVB-S2

DVB-Second Satellite generation (DVB-S2) [18, 19] is the newly ratified air interface protocol specification designed to promote the development of interoperable technology and services in the satellite broadcast and Internet access communication market. It

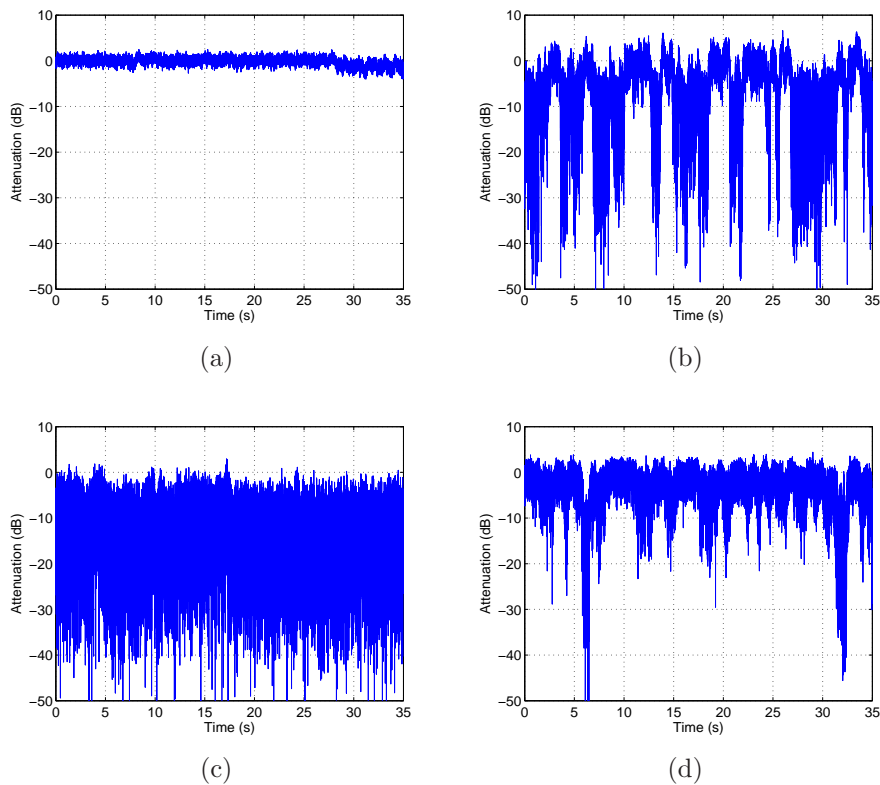


Figure 3.2: Channel attenuation time-series for different LMS Environments at S-band: (a) Open area; (b) Intermediate tree shadowing (ITS); (c) Heavy tree shadowing (HTS); (d) Suburban area.

satisfies the evolutionary needs of the consumer direct-to-home services market and sets the stage for a paradigm shift in the delivery of broadband interactive services via satellite.

3.2.1.1.1 Adaptive physical layer

The standard includes Continuous Coding and Modulation (CCM), Variable Coding and Modulation (VCM) and ACM operative modes to satisfy different needs from broadcast to unicast applications, with constant protection (CCM), service-based differentiated protection (VCM) and adaptive protection (ACM) to counteract fading dynamics. The ACM functionality allows different modulation formats and error protection levels (i.e. coding rates) to be used and changed on a frame-by-frame basis within the transmitted data stream. The principle is that each receiving station has the possibility to control the protection mode of the transmission by sending reports on traffic quality (see Figure 3.3). The technical characteristics of the adaptive physical layer includes:

1. An advanced forward error correction coding scheme with near-Shannon performance, using Low Density Parity Check Code (LDPC) with long code block sizes (64800 bits for the long block size or 16200 bits for the short block size).

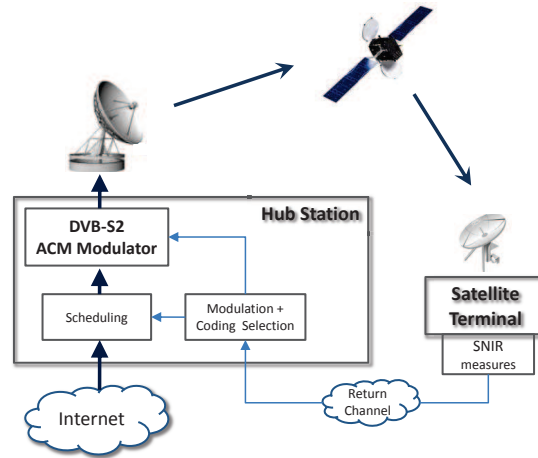


Figure 3.3: Block Diagram of a DVB-S2 ACM link.

2. A variety of modulation schemes from QPSK, to 8PSK, 16APSK and 32APSK which combined with a large set of supported code rates provides a large range of spectral efficiencies spanning from about 0.5 b/s/Hz to 4.5 b/s/Hz.

3.2.1.1.2 System architecture

DVB-S2 accommodates any input stream format. In addition to the widely used Moving Picture Experts Group (MPEG) Transport Stream (TS), Generic Stream (GS) (of constant or variable length packets) are encompassed by the standard, allowing IP, Asynchronous Transfer Mode (ATM) packets or any future data format to be used without the need for a new specification. Basically, input stream is transformed into a sequence of frames. Within these frames, the modulation and coding scheme is homogeneous. They also carry signally data to configure the receiver according to the application scenarios: single or multiple input streams, generic or transport stream, CCM or ACM/VCM. Moreover, PHY layer inserts signaling pilot symbols where required for robust carrier recovery.

3.2.1.1.3 Hierarchical Modulation

A hierarchical mapper has been specified to combine two TS on a single satellite channel, each TS having a high or low priority in reception at terminal satellite receiver. This is a means to provide a differentiated level of protection for each TS. This can be implemented on backwards compatible DVB-S2 systems by synchronously combining High Priority and Low Priority streams at modulation symbol level on an asymmetric 8PSK constellation.

3.2.1.1.4 IP data encapsulation

Another real benefit of DVB-S2 is its flexible architecture regarding input streams. Currently, IP data are usually carried over Transport Streams following the Multi-Protocol

Encapsulation of [20]. With DVB-S2, it is possible to directly access physical media with IP data and avoid lengths of headers inserted in the MPEG stream.

This smart encapsulation of IP data allows for a better efficiency, thus for achieving required QoS with less efforts. Indeed, longer frames (64800 or 16800 bits) are used for encapsulation while MPEG packet size (fixed to 188 bytes) statistically results in recurrent payload slicing (implying all the required MPEG+IP headers addition).

3.2.1.2 DVB-RCS

DVB-Return Channel via Satellite (DVB-RCS) [21, 22] is the well known standard concerning the air interface of the return link of interactive satellite terminals. It defines the end-to-end connectivity link between the satellite operator hub and the Satellite Terminal (ST). To this end, it embraces the DVB-S and the DVB-S2 standards as the forward link specifications, while it describes in more details the physical and MAC layer aspects of the return channel.

3.2.1.2.1 *PHY-layer: Adaptive Coding*

Instead of using ACM, as in the case of DVB-S2, Adaptive Coding (AC) is performed in DVB-RCS, with a fixed QPSK modulation. Two coding approaches are possible: an outer Reed Solomon outer coding followed by a convolutional inner coding, and turbo coding.

3.2.1.2.2 *MAC-layer: MF-TDMA*

DVB-RCS is designed for accommodating a large number of users in a system, sharing the available bandwidth using the Multiple Frequency Time Division Multiple Access (MF-TDMA) method, where STs have allocated capacity in slots within a certain time-frequency structure. Both static and dynamic slot assignments are possible together with the use of frames and superframes. The possibility to operate with fixed time slots lengths and frequency slots bandwidth simplifies the operation on the forward link. Both ATM and MPEG profiles are defined together with a variety of capacity assignment techniques: Continuous Rate Assignment (CRA), Rate Based Dynamic Capacity (RBDC), Volume Based Dynamic Capacity (VBDC) and Free Capacity Assignment (FCA).

The entire system is controlled by a Network Control Center (NCC) (e.g., at the Gateway side of the satellite) controlling the ST behavior. The NCC is responsible for synchronization of the system, and timeslot allocation within the superframe shared by competing STs. It periodically broadcasts a signaling frame, the Terminal Burst Time Plan (TBTP) that contains the information on which STs relies to know when to transmit their bursts. Each one is defined by a frequency, a bandwidth, a start time and duration. The NCC also broadcasts additional tables informing on frame composition, capacity allocation, regulation of ST timing and frequency offsets, etc.

3.2.1.2.3 *Topology*

Beyond the basic hub-and-spoke architecture, the DVB-RCS air interface has also been deployed in systems that provide direct terminal-to-terminal mesh connectivity, either through satellite On-Board Processors (OBP) that mirror the functions of a ground-based hub, or through transparent satellites, using terminals equipped with an additional demodulator.

3.2.1.2.4 *Mobility*

DVB approved the DVB-RCS Mobile (or DVB-RCS+M) specification in 2008, providing support for mobile and nomadic terminals as well as enhanced support for direct terminal-to-terminal (mesh) connectivity. DVB-RCS+M includes features such as live handovers between satellite spot-beams, spread-spectrum features to meet regulatory constraints for mobile terminals, and continuous-carrier transmission for terminals with high traffic aggregation. It also includes Link-layer (LL) and Forward Error Correction (FEC), used as a countermeasure against shadowing and blocking of the satellite link.

Note that DVB-RCS+M is the version 1.5.1 of ETSI 301 790 standard [21].

3.2.1.3 **DVB-SH**

The DVB-Satellite to Handheld (DVB-SH) standard provides an efficient way of carrying multimedia services over hybrid satellite and terrestrial networks, at frequencies below 3 GHz, to a variety of mobile and fixed terminals having compact antennas with very limited directivity. Target terminals include handheld (PDAs, mobile phones), vehicle-mounted, nomadic (laptops, palmtops) and stationary terminals.

The DVB-SH standard provides a universal coverage by combining a Satellite Component and a Complementary Ground Component (CGC): in a cooperative mode, the Satellite Component ensures geographical global coverage while the CGC provides cellular-type coverage.

3.2.1.3.1 *Architectures*

Orthogonal Frequency Division Multiplexing (OFDM) is the natural choice for terrestrial modulation and is the basis of both the DVB-H and DVB-T systems. DVB-SH introduces a second scheme, a Time Division Multiplex (TDM), leading to two reference architectures termed SH-A and SH-B:

- SH-A uses OFDM both on the satellite and the terrestrial link;
- SH-B uses TDM on the satellite link and OFDM for the terrestrial link.

3.2.1.3.2 *PHY-layer outline*

The transmission scheme of DVB-SH in the case of SH-B architecture is composed of a common part including forward error correction and interleaving, and two different modulators:

- Physical Layer with multi-carrier modulation (OFDM). The multi-carrier modulation concept is derived from DVB-T;
- Physical Layer with single-carrier modulation (TDM). The single-carrier modulation concept is adapted from DVB-S2 technology.

For the OFDM part, the possible choices are QPSK, 16QAM and non-uniform 16QAM with support for hierarchical modulation. A 1K mode is proposed in addition to the usual 2K, 4K and 8K modes which does not exist in either DVB-T or DVB-H. For the TDM part, the choices are QPSK, 8PSK, 16APSK for power and spectral-efficient modulation formats, with a variety of roll-off factors (0.15, 0.25, 0.35).

DVB-SH uses a highly flexible channel interleaver that offers time diversity from about one hundred milliseconds to several seconds, depending on the targeted service level and corresponding capabilities (essentially memory size) of the terminal class.

3.2.1.3.3 Upper layer outline

The main link layer features of DVB-SH are inherited from DVB-H: support of MPEG2-TS packets at the input, although the specification allows for the introduction of a Generic Stream at a later date; Multi-Protocol Encapsulation (MPE) encapsulation and support of MPE Time Slicing (powersaving) and handover between frequencies/coverage beams; compatibility with MPE-FEC; Generic Stream Encapsulation (GSE) compatible.

However, the longer fading experienced in the satellite context tends to require longer protection than provided by MPE-FEC. Therefore it is currently foreseen that an MPE-FEC extension (interburst FEC) should be supported. This MPE-FEC extension is expected to combat the deep and long shadowing encountered in some satellite channels by providing additional time diversity.

Furthermore, DVB-SH benefits from the set of IP Datacast specifications, which were originally defined to turn the DVB-H transmission system into an end-to-end solution. IP Datacast will now be amended, if necessary, in order to act as the "higher layer" of DVB-SH.

3.2.1.4 WiMAX

Mobile WiMAX, which follows the IEEE 802.16 air interface specifications [23], is a broadband wireless solution that enables convergence of mobile and fixed broadband networks through a common wide area broadband radio access technology and flexible network architecture. In the following, some detailed descriptions of key PHY and MAC features in the mobile WiMAX system profile are offered.

3.2.1.4.1 PHY layer key features

The Mobile WiMAX Air Interface adopts OFDM Access (OFDMA) for improved multi-path performance in non-line-of-sight environments. Scalable OFDMA (SOFDMA) is introduced in the IEEE 802.16e Amendment [23] to support scalable channel bandwidths from 1.25 to 20 MHz. WiMAX operates at the S-Band (principally around 3.5 GHz).

The mobile WiMAX profile has only Time Division Duplex (TDD) as the duplexing mode even though the baseline IEEE Standards contain both TDD and Frequency Division Duplex (FDD). TDD is in many ways better positioned for mobile Internet services than FDD. First, because Internet traffic is asymmetric typically with the amount of downlink traffic exceeding the amount of uplink traffic, and thus, conventional FDD with the same downlink and uplink channel bandwidth does not provide the optimum use of resources. Moreover, TDD is inherently better suited to more advanced antenna techniques such as Adaptive Antenna System (AAS), such as Multiple-Input Multiple-Output (MIMO) and Beamforming (BF), which are implemented in the mobile WiMAX profile.

Full mobility support is yet another strength of the mobile WiMAX products. The baseline standard of mobile WiMAX was designed to support vehicles at highway speed with appropriate pilot design and Hybrid ARQ (HARQ), which helps to mitigate the effect of fast channel and interference fluctuation. The systems can detect the mobile speed and automatically switch between different types of resource blocks, called subchannels, to optimally support the mobile user.

3.2.1.4.2 MAC layer key features

The flexibility of the WiMAX MAC layer is defined by the following features.

The WiMAX technology provides an environment for connection-oriented services. For each service, certain classification rules are specified to define the category of traffic associated with the connection. Certain QoS parameters can be defined, such as minimum reserved rate, delay and jitter requirements, etc. Types of scheduling for real-time services, streaming or web browsing can be applied based on the application requirements. Special scheduling types are defined for the VoIP service with silence suppression and adaptive codecs, see the end of the section for detailed QoS types.

Bandwidth allocation mechanism is based on real time bandwidth requests transmitted by the terminals, per connection. Bandwidth requests may be transmitted using a contention based mechanism or they can be piggybacked with the data messages. The Base Station executes resources allocation based on the requests and QoS parameters of the connection.

Multicast and Broadcast Services (MBSs) allow WiMAX mobile terminals to receive multicast data even when they are in idle mode. The most popular application of this feature is TV broadcasting to mobile terminals.

A summary of other MAC mechanisms and properties are listed next:

- Mobility Support: Handover procedures include numerous means of optimization.
- The security sublayer provides Extensible Authentication Protocol (EAP)-based mutual authentication between the mobile and the network.
- WiMAX technology includes support of the general Purpose Header Suppression (PHS) and IP Header Compression (ROHC).
- Power saving modes: idle mode when the mobile is away from the base station and sleep mode when the mobile has no traffic for a long time.

A brief explanation of the MAC sub-layers is included in the hybrid design at the end of the section.

3.2.2 QoS Model

As we will see, each type of traffic has different requirements on average delay, delay jitter, average rate, peak rate, burst size, possibility of retransmissions, adaptability of the application layer, etc.

Two different models have been defined to facilitate end-to-end QoS over IP-networks: Integrated Services (IntServ) and Differentiated Services (DiffServ). IntServ follows the signaled-QoS model, where the end-hosts signal their QoS needs to the network, while DiffServ works on the provisioned-QoS model, where network elements are set up to service multiple classes of traffic with varying QoS requirements. IntServ provides for a rich end-to-end QoS solution, using end-to-end signaling, state-maintenance and admission control at each network element. DiffServ, on the other hand, addresses the clear need for relatively simple and coarse methods of categorizing traffic into different classes, also called Class of Service CoS (CoS), and applies QoS parameters to those classes. To accomplish this, packets are first divided into classes by marking the Type of Service (ToS) byte in the IP header. A small bit-pattern (6 bits), called the Differentiated Services Code Point (DSCP), is used in the IPv4 or the IPv6 Traffic Class Octet to mark a packet, which in this way receives specific forwarding treatments, formally called Per-Hop Behavior (PHB), at each network element. This allows providing the packet the appropriate delay-bound, jitter-bound, bandwidth, etc.

This combination of packet marking and well-defined PHBs results as a scalable QoS solution for any given packet, and any application. In DiffServ, signaling for QoS is eliminated, and the number of states required to be kept at each network element is drastically reduced, resulting in a coarse-grained, scalable end-to-end QoS solution.

3.2.2.1 DVB-S2/RCS

DVB-S2/RCS does not standardize QoS provision. Instead a number of tools are given at access layer which can be used to provide the desired QoS guarantees.

We assume a DiffServ model for the satellite subnetwork thus enabling easy networking with IP external networks. We consider that the DVB-S2/RCS deals with three CoS; Expedite Forward (EF) for real-time services, Assured Forward (AF) and Best Effort (BE). EF is implemented to provide premium service, for applications such as VoIP, video, and online trading programs that require a robust network treatment. The main goal of EF is providing low loss rate, guaranteed end-to-end delay, low-latency, low-jitter and an assured bandwidth service. On the other hand, AF is usually defined for providing low loss rates, but not guaranteeing delays and jitters, such in streaming video/audio applications, FTP, etc. The last service in the architecture is BE, which corresponds to all other IP packet flows without priority or flagging, e.g. used in web browsing.

Table 3.3: Mobile WiMAX Applications and QoS

CoS Category	Applications	QoS Specifications
ErtPS Extended Real-Time Polling Service	Adaptive VoIP	Minimum Reserved Rate Maximum Sustained Rate Maximum Latency Tolerance Jitter Tolerance Traffic Priority
UGS Unsolicited Grant Service	VoIP, ATM CBR	Maximum Sustained Rate Maximum Latency Tolerance Jitter Tolerance
rtPS Real-Time Polling Service	Audio/Video Streaming, VoIP, ATM	Minimum Reserved Rate Maximum Sustained Rate Maximum Latency Tolerance Traffic Priority
nrtPS Non-Real-Time Polling Service	File Transfer Protocol (FTP)	Minimum Reserved Rate Maximum Sustained Rate Traffic Priority
BE	Data transfer, web browsing	Maximum Sustained Rate Traffic Priority

3.2.2.2 WiMAX

One of the main contributions of WiMAX respect to other wireless technologies (such as WiFi) is the introduction of QoS. Table 3.3 summarizes CoS applications for WiMAX networks.

3.3 Study case: An Hybrid Architecture

A relevant part of this dissertation has focused on satellite and hybrid solutions for rural and tropical areas, designed for the BRASIL [24] project. The main objectives are focused on important aspects of interactive applications. The target region is Latin America in general, but specifically supporting European interactive satellite broadcast communications technologies such as DVB-RCS in Brazil. Before a detailed description of the architecture, let us introduce the demographic context.

3.3.1 Demographic aspects

Amazon region of interest for this project is an area composed by the following Brazilian states: Amazonas, Par, Roraima and Amap with a total population of 11.188.820 inhabitants in more than 3 millions of square kilometers. Only in their capitals there are around 3.648.148 inhabitants ($\approx 33\%$). Besides the capital municipalities, 35 municipalities have between 50.000 and 500.000 inhabitants totalizing approximately 3.500.000 inhabitants ($\approx 31.2\%$) that represents an average of 100000 inhabitants/municipality. The rest of the population in this area represents the remaining 35.8% with municipalities up to 50.000 inhabitants and currently with no connectivity [25].

Hybrid networks composed of satellite backhauling and terrestrial radio access networks naturally come into place to provide broadband connectivity to sparsely populated areas, such as the villages along the Amazon River, which can be hundreds of kilometers apart from each other with a small but scattered population. The hybrid solution discussed here is much more cost effective than having a circuit-switched telephone network plus a satellite.

Two main hybrid architectures has been selected to provide IP services for this populations. The municipalities can be divided between those having less than 5.000 inhabitants and those between 5000 and 50.000 as with the following characteristics:

- **Community network:** For this first group, we assume a hybrid architecture consisting of DVB-RCS satellite terminals with a WiFi card and up to 200 meters coverage. In this case, the objective is that users will have access to the network from specific access point only, probably located at some public buildings.
- **Consumer network:** For this second group, we assume a hybrid architecture of DVB-RCS terminals with WiMAX cards and up to 10 km coverage. Conversely to the previous page, this consumer-type architecture allows individuals access to the network from user's premises.

In this dissertation, we focus on the second group, where users can be served by WiMAX network(s) attached to a satellite terminal. Moreover, the architecture can be used as a solution not only for the Amazon area, but also would be excellent for rural areas or in scenarios with very remote residential areas where families live some dozens of kilometers apart from each other. Finally, let us note that the following architecture is taken as reference model in several of the cross-layer designs presented in Chapter 6.

3.3.2 Hybrid Reference Architecture

3.3.2.1 The DVB-S2/RCS Standard and WiMAX

We consider a Geostationary Earth Orbit (GEO) MBS link with a DVB-S2/RCS [18, 19] air interface. DVB-S2 allows ACM and can be adapted on a per-time slot basis depending on the SNIR at the destination terminal. DVB-RCS implements MF-TDMA and adaptive coding only is allowed [21, 22]. The satellite transmits several beams to the earth, covering ideally non-overlapping geographical areas. The total available power is distributed among the beams either evenly or not according to traffic and additional requirements. Apart from differences in the link budget, all beams are conceptually identical. We assume each terminal has 2-3 associated WiMAX networks as shown in Figure 3.4. Each WiMAX network consists of mobile users with mobility within the coverage area. In particular, we consider the AMAZONAS-1 satellite operated by Hispasat.

For the WiMAX architecture [16] (IEEE 802.16e-2005) connected to the DVB-S2/RCS terminal, we assume an access point based simple architecture in spite of using the rich feature set and flexibility offered by WiMAX. We have assumed so as it reduces the complexity of service deployment and provisioning for fixed and mobile networks in non-urban environments. Finally, we also assume DVB-S2/RCS terminal is fixed while user device is either fixed or nomadic.

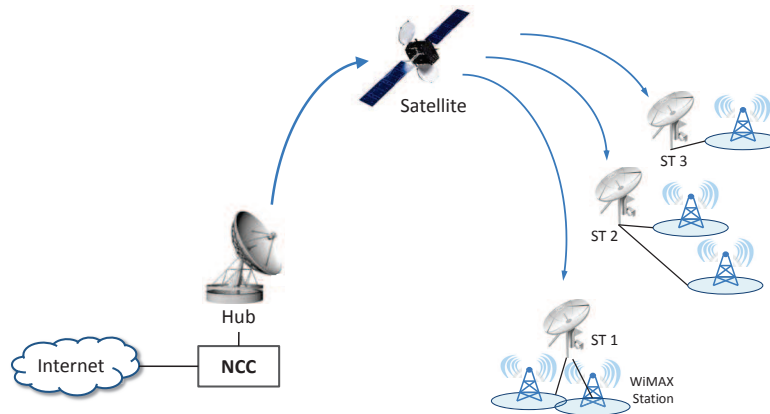


Figure 3.4: DVB-S2/RCS star topology with WiMAX networks associated to Satellite terminals.

3.3.2.2 Architecture & Protocol Stack

Many solutions have been proposed for heterogeneous networks [26, 27], but our efforts have been made in order to design a model for providing service in areas where network deployment is not consistent.

Figure 3.5 shows the hybrid architecture that we are proposing. The terrestrial link is based on IEEE 802.16, but adapted to our scenario, where a user in the core network directs traffic to a mobile user, called User Equipment (UE). The UE is placed in an area serviced by WiMAX network, and the satellite link is used as backhaul. The Base Station (BS) is responsible of the IEEE 802.16 connectivity through the radio link to UE located inside its coverage area. An adaptive physical layer is specified in the standard, to maximize the data rate by adjusting transmission modes to channel variations while maintaining a prescribed PER. The WiMAX Radio Resource Management (RRM) is in charge of utilizing the limited radio spectrum resources and radio network infrastructure of its associated BS as efficiently as possible. In order to achieve it, RRM involves some strategies and algorithms for controlling parameters; transmit power, channel allocation, handover policies, FEC and modulation features for the WiMAX users.

Other components shown in Figure 3.5 include the Access Service Network (ASN) Gateway (GW), and the DVB-RCS RRM. Therefore, ST is in charge of controlling the transmission and reception through the satellite link. It manages the bandwidth allocation of all associated BS. In case a new UE requests service, the WiMAX-RRM sends the request to the RCST-RRM, it checks if enough resources are available. If not, the ST sends a request to the Hub, which can broadcast a new TBTP if necessary. The Hub is associated to a NCC. The NCC controls the interactive network, user service requests via satellite access and manages the satellite spectrum depending on the satellite terminals requests.

The versatility of WiMAX perhaps is due to the way the MAC layer has been defined. The protocol stack of the WiMAX RRM (Figure 3.5) shows three sublayer forming the whole MAC layer. The Convergence Sublayer (MAC-CS) provides the transformation or mapping of external network data (Ethernet, IP, etc.). The Common Part Sublayer

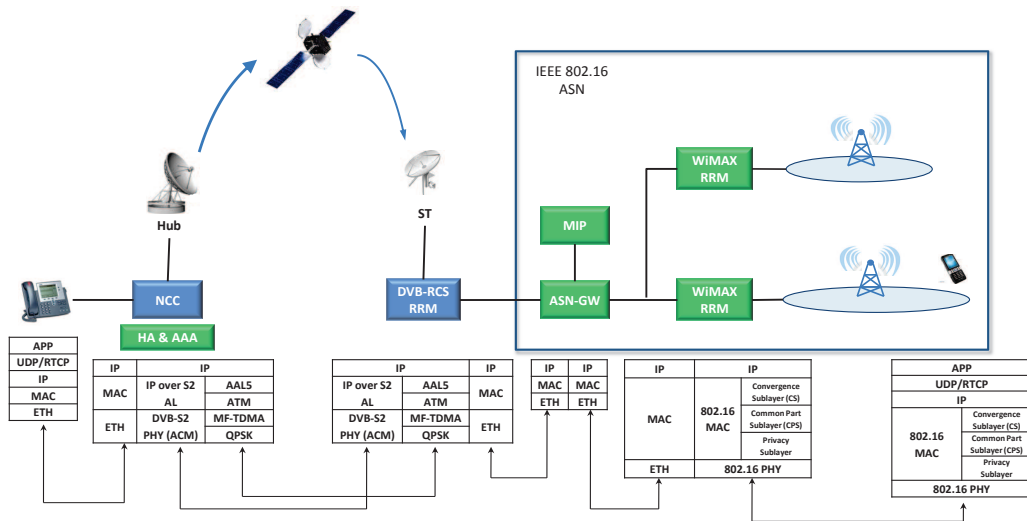


Figure 3.5: Reference architecture and protocol stack for the hybrid WiMAX and DVB-S2/RCS network.

(MAC-CPS) performs packing into MPDU (MAC packet data unit) of information coming from MAC-CS. And the Privacy Sublayer, which provides authentication, key exchange and data encryption. MAC-CPS is the core of MAC layer, since it is responsible of the already commented versatility. It provides QoS, manages bandwidth, multiplex VoIP flows directed to the BS, establishes and maintains the connection, performs FEC, enables ARQ mechanisms, and so on.

Several Access Service Network (ASN) profiles have been specified in WiMAX as a tool to manage diversity node usage and implementation [28]:

- Profile A: Centralized ASN model with ASN-GW and BS in separate platforms with split RRM. Radio Resource Allocation (RRA) in BS and Radio Resource Control (RRC) in the gateway.
- Profile B: BS and ASN-GW functionalities implemented in a single platform.
- Profile C: Separate platforms, with the RRM controlled by the BS.

Profile A is suitable for soft handover, used in high speed mobiles, but that is not our case, where the typical users are mobiles inside a rural area. Moreover, profile A entails difficult interoperability between BS and ASN-GW from different vendors, and heavy workload at ASN-GW. Although profile B is the most simple architecture, operators prefer separate platforms, since with them, it is easier to customize IP and wireless functions. The distributed platform of profile C is the one better matches with our scenario. By placing the ASN-GW between the two RRM; in the Satellite Terminal and the WiMAX network, it allows an interaction between both in order to manage all the resources in a friendly way. This means that each BS is in charge of managing the IEEE 802.16 service within its area, while the ST carries out the resource assignment of all ASNs.

Table 3.4: QoS mapping between DVB-S2 and WiMAX Classes of Service

WiMAX CoS	DVB-S2/RCS CoS	Specifications
ErtPS / UGS	EF	Guaranteed end-to-end delay, jitter and assured bandwidth (VoIP)
rtPS / nrtPS	AF	Low loss rates but not guaranteed delays and jitters (Audio/Video Streaming)
BE	BE	Web browsing

ASN-GW incorporates the Mobile IP [29], used to provide an efficient and scalable mechanism for roaming within the Internet. The gateway works as Foreign Agent (FA) of all associated mobile terminals within this ASN. In the Hub side, the Network Service Provider (NSP) is composed of the Home Agent (HA) and the Authentication, Authorization and Accounting (AAA) server. The HA stores information about mobile nodes whose permanent address is in the NSP. AAA is in charge of controlling access to computer resources, enforcing policies, auditing usage, and providing the information necessary to bill for services. These combined processes are considered important for effective network management and security.

The Hybrid system performs as follows: When a Remote User (RU) in the core network directs traffic to the UE, this traffic will be forwarded to the HA, which knows the location and the FA the UE belongs to. Therefore, the HA will tunnel the traffic (IP packet as payload of a new IP packet) to the FA. The ASN-GW strips the IP header and forwards it to the BS. In case a UE would like to be registered into a specific coverage area, the BS (ASN-GW) checks if enough resources are available. If the resources are available, the UE is registered into the FA after authentication and authorization procedure. Afterwards, the RRM of the ST receives also a request for service from the ASN-GW, in case the user asks for service.

3.3.3 QoS Model

The main challenge when trying to perform a QoS mapping between DVB-S2/RCS and WiMAX sub-networks, both QoS models are previously defined in this chapter, is related to the real-time services. In IEEE 802.16e systems, there are several scheduling algorithms designed for supporting real-time services; Unsolicited Grant Service (UGC), real-time polling service (rtPS) and extended-rtPS (ertPS). In [30], it is shown that ertPS can support more than 21% and 35% of voice users compared with UGS and rtPS respectively. UGS always allocate the same amount of uplink resources to each user. rtPS incurs in additional access delay and MAC overhead due to bandwidth request process. On the other hand, ertPS can be used efficiently in WiMAX in order to support variable data rates and silence suppression, which fits perfectly for VoIP adaptation (addressed in Chapter 6). Moreover, this algorithm has been recently proposed and accepted in IEEE 802.16e standard.

Packets of different classes must be transmitted through the whole network, by interacting between WiMAX and Satellite terminals. Despite this field is yet open to

modifications, we propose the QoS mapping to CoSs given in Table 3.4. Let us stress that the WiMAX-RRM performs the QoS matching.

Chapter 4

Cross-Layer Design at System Level

4.1 Objective: Capacity Maximization

As demands of group-oriented and wireless-connection based communication services with the diverse QoS requirements increase explosively, mobile multicast has become a critically important component in the next generation wireless networks. One of the most attractive characteristics of mobile multicast is the highly efficient group communication through wireless broadcast channels with significantly reduced wireless resource consumption. The broadcast nature of the satellite network makes it the natural choice for multicasting services, however, the changing nature of the satellite channel has a great impact on adaptive satellite and hybrid systems such as DVB-S2/RCS and DVB-SH, specially in mobile scenarios, where fading needs to be mitigated in order to avoid link degradation and therefore capacity decrease. This is a challenging problem when we try to optimize a multicast transmission, where each user experiments different channel condition.

One solution, based on the broadcast scheme, is to transmit data considering the worst case user, but it causes inefficient use of radio spectrum and also a saturation of the system throughput when number of users increases. Therefore, the problem is how to satisfy different users without sacrificing efficiency. ACM was introduced to increase the efficiency of the satellite system, allowing to differentiate users according to channel conditions. However, one of the main drawbacks of ACM is that the physical layer only adapts to slow fading conditions. Although in Chapters 5-6, the fundamental idea is to design efficient upper-layer mechanisms that are able to follow the adaptability of the physical layer, in this chapter, the main problem is how to maximize multicast capacity from the physical layer point of view for fast changing conditions.

To address the above problem, we aim at optimizing the multicast transmission over the satellite and the terrestrial link of current hybrid systems. Our efforts have been focused on the following items:

1. To study the air interfaces in multibeam mobile broadband satellite/hybrid systems. First, we propose to optimize the multicast transmission by means of a Hierarchical Allocation (HA) algorithm based on SC, which is able to adapt to fast channel variations. The problem includes a capacity maximization but limited to fulfill a minimum rate to specific users, or multicast groups. Such rate is assumed to be

cross-layer information from upper-layers, requesting for a minimum service rate.

2. To compare HA-SC with an improved time-slot based algorithm, known to perform opportunistic multicasting, which aims at optimizing the number of users to be served at each time slot.
3. Finally, to apply the problem to a hybrid satellite-terrestrial system model, based on DVB-SH. We justify that a cooperative transmission scheme between the OFDM and TDM air interfaces is more efficient when trying to cover all mobile users. Moreover, based on the standard recommendations of DVB-SH [31], we propose to use HA-SC in combination with the ACM of the TDM air interface. We argue that the transmission scheme is suitable for managing multicast transmissions over different mobile scenarios.

4.2 Methodology: Hierarchical Multicast Transmission

It is known that the propagation delay is a strong limitation in satellite network (based on ACM) since it does not allow the physical layer to adapt to fast channel conditions, i.e. the NCC receives Channel State Information (CSI) from satellite terminals with a delay around 250ms (for a Geostationary Earth Orbit), and thus adaptation is not feasible. In order to overcome this problem, in this chapter, we propose another approach, termed the hierarchical multicast transmission. This strategy facilitates reliable transmission rates adapted to the actual channel conditions, without providing any feedback information from the receiver to the NCC. Our hierarchical allocation algorithm benefits from the SC scheme, which along with Dirty Paper Coding (DPC) are the theoretical transmission configurations that achieve the capacity region of wireless communications systems.

We consider a satellite multicast network in a quasi-static fading environment with additive Gaussian noise. The objective is to maximize the overall aggregated rate of a network where different multicast groups coexist. We constrain the scenario to serve a minimum QoS (in terms of rate) that should be allocated to those users with a minimum channel gain, meaning that users with channel gain above the minimum should receive the minimum rate service. We assume the users within the same group are independent and identically distributed (i.i.d.) following typical channel distribution for LMS (open or suburban areas, different tree shadowing conditions, etc.). However, since each group experiences different channel distribution, we assume the overall channel state is non-i.i.d..

Note that the CSI of each terminal is assumed to be known only at the terminal receiver. In this case, when no CSI is available at the transmitter, a multiple-antenna fading channel with either one receive or one transmit antenna can be modeled as a degraded broadcast channel where each channel realization corresponds to a user [32]. And the optimal transmission strategy for the degraded broadcast channel is SC, where the fading channel realization allows the layers up to a certain level to be decoded at the receiver, while considering the undecodable layers as interference. Therefore, the problem is to find the power allocations that minimize the expected interference and maximize the rate.

We combine the broadcast strategy proposed in [33], which maximizes the average rate over an i.i.d. channel distribution, with our multicast scenario in order to find an optimal distribution among users. This approach is very interesting since due to delay constraints, transmission allocation is performed with non-CSI available at the transmitter. Optimization of the SC transmission schemes has been also addressed for the case that the transmitter has partial CSI in [34, 35], as extended works of [33]. Regarding studies that assume non-CSI at the transmitter, [36] proposes a multilevel approach for joint source-channel coding, where the expected distortion is minimized by optimally allocating the transmit power among the source layers.

Enabling a minimum QoS allocation per group allow us to obtain a fair power allocation, while in [33], since the scenario is not constrained, the allocation problem not only tends to benefit users with best channel conditions, but also increases the outage probability of worst users. Some examples of the minimum QoS criteria can be find in the literature, however, main works assume CSI is known at the transmitter. In [37], for a single user fading channel, given a service outage constraint for a real-time application, the average rate is maximized for a non real-time application sent on top of it. An adaptive variable rate code is proposed and shown to be optimum in that scenario. Similarly, a minimum rate constrained capacity measure is defined for broadcast channels in [38], where it is shown that the minimum rate region is the ergodic capacity region of a broadcast channel, with an effective noise determined by the minimum rate requirements. Using similar assumptions for the CSI availability, a more general case is considered in [39], where each user specifies its rate constraints in a triplet of maximum rate, minimum rate, and a so called shortage probability.

4.3 Optimization Problem and Solution: HA-SC

Let us consider a transmitter that sends common data to N receivers organized in S multicast groups. The signal received y_n is defined as $y_n = cx_n + n_n$, where x_n is the transmitted signal with average power constraint $E[|x|^2] \leq P$, $n_n \sim CN(0, 1)$ is the additive noise, and c is the channel coefficient, which is assumed to be a circularly symmetric complex Gaussian random variable, with channel gain $g = |c|^2$. We consider a degraded broadcast channel since it allows SC to reaching the capacity region. The degraded broadcast channel is theoretically seen as a transmitter sending an infinite number of layers of coded information, where the receivers can only decode a set of layers depending on its channel conditions. This model is the general case of code layering. In the approach with a finite number of code layers, presented in [40], only a finite set of ordered receivers is required. Obviously, the approach has lower decoding complexity, however, it is a broadcast suboptimal approach. Each channel realization g_n has associated a rate fraction dR_n , and therefore the rate achieved by user n is the sum of all fraction rates associated to lower channel coefficient (worse received signal):

$$R_n = \sum_{i=1}^n dR_i \quad (4.1)$$

Considering the transmit power distribution as $\rho(g)$ and the interference for a fading

power g as:

$$I(g) = \int_g^\infty \rho(z) dz \quad (4.2)$$

One can determine the rate fraction that every gain coefficient g adds as:

$$dR(g) = \log \left(1 + \frac{g\rho(g)dg}{1 + gI(g)} \right) = \frac{g\rho(g)dg}{1 + gI(g)} \quad (4.3)$$

And for each realization of g , there is an achievable rate $R(g)$:

$$R(g) = \int_0^g \frac{z\rho(z)dz}{1 + zI(z)} \quad (4.4)$$

Note that the information layers transmitted to receivers with gain bigger than user g play the role of additional interfering noise since it can not be decoded by this user. Since we pretend to obtain an overall capacity measure to be maximized, we focus on the aggregated capacity of each multicast group, instead of the overall average achievable rate of (4.4) assumed in [33]. In our case, if we assume different groups, which are non-i.i.d. between them, then it means that we consider a different probability distribution function (pdf) $f_i(z)$ of the channel gain for each multicast group i , and the corresponding cumulative distribution function (cdf) $F_i(z)$. Thus, let us state that the rate averaged over the fading realization of a channel associated to multicast group i is:

$$R_{ave,i} = \int_0^\infty f_i(z)R(z)dz = \int_0^\infty (1 - F_i(z)) \frac{z\rho(z)dz}{1 + zI(z)} \quad (4.5)$$

In particular, we associate (4.5) with an average rate among the users within a group. And thus, the maximization of the overall capacity can be stated as:

$$\begin{aligned} \max_{I(g)} \quad & \frac{1}{S} \sum_{i=1}^S R_{ave,i} = \\ \max_{I(g)} \quad & \frac{1}{S} \sum_{i=1}^S \int_0^\infty (1 - F_i(z)) \frac{z\rho(z)dz}{1 + zI(z)} \end{aligned} \quad (4.6)$$

However, (4.6) might tend to an unfair resource allocation. It is straightforward to see that power allocation to worst users is costly, and thus it is more efficient to not allocate power to worst users (i.e. higher outage probability) in order to reach higher average overall rate. As a solution to avoid unfair allocations, we constrain the rate allocation to $R(g) \geq R_{min}$, which can be interpreted as a guaranteed service (specified by the upper-layers), meaning that at least all users will receive this rate. In [41], authors use a similar soft constraint, but based on an assumption of the worst channel pdf (statistical independent), which does not allow to control the parameters. However, in our case, we use an accurate boundary for the minimum guaranteed service, which let us:

- Define a minimum rate allocated R_{min} , which is specified by upper layers by means of a cross-layer design.

4. Cross-Layer Design at System Level

- Specify a low bound channel gain g_{min} , meaning that users with channel gain $g \geq g_{min}$ will receive at least R_{min} .
- Indicate the outage channel gain g_{out} , i.e. users below this channel level receive no rate. This is very helpful in case we need to discard the worst groups in order to avoid heavy capacity reduction. In that case, the optimization can be recomputed and so the power redistributed among groups, reaching higher rates.

Let us state our final problem optimization as:

$$\begin{aligned}
 & \max_{I(g)} \quad \frac{1}{S} \sum_{i=1}^S R_{ave,i} \\
 & s.t. \quad R(g_{min}) = R_{min} \\
 & s.t. \quad \int_{g_{out}}^{\infty} \rho(g) dg \leq P
 \end{aligned} \tag{4.7}$$

The second constraint specifies the maximum power P that must be distributed. Note that the set (R_{ave}, R_{min}) should be jointly optimized according to the interference (4.2). This problem can be solved by using the Lagrange multiplier method for functionals subject to integral constraints [42, 43]. For the sake of simplicity, we assume only one multicast group, i.e. $S = 1$. The functional J to be extremized is given by:

$$J = \int_{g_{out}}^{\infty} (1-F(z)) \frac{z\rho(z)dz}{1+zI(z)} - \lambda \left(\int_0^{g_{out}} \frac{z\rho(z)dz}{1+zI(z)} - R_{min} \right) - \mu \left(\int_{g_{out}}^{\infty} \rho(z)dz - P \right) \tag{4.8}$$

where λ and μ are the Lagrange multipliers, which for this particular case are a function of the independent variable z [42]. Note that integral boundaries must be equal in each terms, thus we organize (4.8) in two functionals as follows:

$$J = J_{g_{min}} + J_{\infty} + \lambda R_{min} + \mu P \tag{4.9}$$

$$J_{g_{min}} = \int_{g_{out}}^{g_{min}} \left((1-F(z)) \frac{z\rho(z)}{1+zI(z)} - \lambda \left(\frac{z\rho(z)}{1+zI(z)} \right) \mu \rho(z) \right) dz \tag{4.10}$$

$$J_{\infty} = \int_{g_{min}+\epsilon}^{\infty} \left((1-F(z)) \frac{z\rho(z)}{1+zI(z)} - \mu \rho(z) \right) dz \tag{4.11}$$

This can be interpreted as follows; the first term $J_{g_{min}}$ is in charge of assuring the minimum rate R_{min} , and the second one J_{∞} aims at maximizing the throughput by means of the remaining resources. The necessary conditions to maximize (4.7) over g is a zero variation of the functional [44], which corresponds to the Euler-Lagrange equation $\partial J / \partial I - d(\partial J / \partial I') / dz = 0$. The result of performing Euler-Lagrange to both functionals $J_{g_{min}}$ (4.12) and J_{∞} (4.13) is observed here:

$$\frac{1 - F(z) - \lambda(z)}{(1+zI(z))^2} - \frac{z(f(z) + \lambda'(z))}{1+zI(z)} = 0 \tag{4.12}$$

$$\frac{1 - F(z)}{(1 + zI(z))^2} - \frac{zf(z)}{1 + zI(z)} + \mu'(z) = 0 \quad (4.13)$$

where λ' and μ' are $\partial\lambda/\partial z$ and $\partial\mu/\partial z$ respectively. As a result, the power interference can be expressed as:

$$\rho(g) = \begin{cases} P, & 0 \leq g_{outage} \\ \frac{1 - F(z) - zf(z) - (\lambda(z) + z\lambda'(z))}{z^2(f(z) + \lambda'(z))}, & g_{outage} > z \leq g_{min} \\ \frac{1 - F(z) - zf(z) - \mu'(z)}{z(zf(z) + \mu'(z))}, & g_{min} > z \leq g_{max} \\ 0, & z > g_{max}. \end{cases} \quad (4.14)$$

The problem is solved by means of the following boundary conditions; $I(g_{out}) = P$, $I(g_{max}) = 0$. For $I(g_{min})$ boundary, it is necessary to take profit of (4.3) and (4.4), so we can write:

$$R(x_{min}) = \log \left(1 + \frac{g_{min} \int_{g_{out}}^{g_{min}} \rho(g_{min}) dg}{1 + g_{min} I(g_{min})} \right) = \log \left(1 + \frac{g_{min}(P - I(g_{min}))}{1 + g_{min} I(g_{min})} \right) \quad (4.15)$$

Since $R(x_{min})$ is one of the specifications, $I(g_{min})$ can be easily obtained. Note that the equality in (4.15) can be justified thanks to (4.2), from which can be subtracted that $I'(g) = -\rho(g)$.

In our solution, since we express the result as a function of the channel distribution, it allow us to perform the optimization not only for typical Rice/Rayleigh distribution, but also for a big set of multicast scenarios such as LoS, intermediate shadowing or deep shadowing. Note that, in some cases, this might increase the computation complexity of integrals, e.g. a clear example of complexity is the Bessel function included in a typical Rice distribution. In order to simplify and solve the problem, we contemplate the approximation in [45], instead of the truncated Taylor series, due to its highest accuracy and because it does not diverge so fast for high values.

Finally, let us highlight that the definition of R_{min} can be generalized to several groups, increasing the number of constraints, and thus controlling the minimum rate allocated to each multicast group.

4.3.1 Time Scheduling (TS) reference

The following algorithm, in contrast to the HA-SC, aims at optimizing the time-slot based multicast transmission. This algorithm has been chosen from the literature and modified according to our multicast scenario for the sake of comparison with the HA-SC.

It is based on the opportunistic multicasting, which searches for the best subset of users to schedule at each channel realization. Some topics have been published about this topic. For instance, in [46], the users are divided in two subsets of the same size; when a packet has to be sent, the scheduler decides which one can support the highest rate, and send the packet only to this subset. Then, this packet is copied to the service queue

of the other subset, and when this subset has better channel, the scheduler sends them the packet. Therefore, the main idea is to send the same information to all users, but at different time-slots. However, we have focused on a TS algorithm used in [47], which is not limited to scheduling only to the best half of the users. Moreover, the strategy relies on sending erasure encoded blocks rather than original message blocks. Further explanation is founded at the end of this section. We consider a channel as the one defined for the HA-SC, which remains constant during a time-slot. We can express the capacity of user n in slot k as:

$$R_n[k] = \log_2(1 + SNR R_m[k]) \quad (4.16)$$

with an average $SNR R_m[k] = g_m[k]P$. The transmitter is constrained to choose the best L users in each slot ($L = 1..N$) according to its capacity ($R_1 \geq R_2 \geq \dots \geq R_L \geq \dots R_N$), and sets the channel rate equal to R_L . Since the probability of successful reception for all users is L/N , the throughput obtained is:

$$\bar{T}_{TS}(L) \equiv \frac{L}{N} E[R_L] = \frac{L}{N \ln(2)} \int_0^\infty \ln(1+x) dx \quad (4.17)$$

Lets denote $f_{X(L)}$ and $F_{X(L)}$ as the pdf and cdf of $X(L)$, which is the L th largest value of SNR, and $F_{X(L)}^c = 1 - F_{X(L)}$. Note that (4.17) only depends on L , which should be appropriately selected in order to optimize the throughput capacity. Applying integration by parts and considering Rayleigh distribution properties, the left-hand side equality of (4.18) can be demonstrated. In our case, in order to compare the TS and HA-SC algorithms, we find the solution for the channel gain ($G(L) = X(L)/P$) as a Rician distributed random variable, rather than the SNR Rayleigh distributed case obtained in [47]. As a result, we obtain the right-hand-side equality:

$$\bar{T}_{TS}(L) = \frac{L}{N \ln(2)} \int_0^\infty \frac{F_{X(L)}^c}{1+x} dx = \frac{LP}{N \ln(2)} \int_0^\infty \frac{F_{G(L)}^c}{1+zP} dz \quad (4.18)$$

Being $F_{G(L)}^c(g) = F_{X(L)}^c(g/P)$. Finally, $F_{X(L)}^c$ is computed by means of the binomial distribution [47]. After some manipulation, the total throughput expression is derived as:

$$\frac{LP}{N \ln(2)} \sum_{j=L}^N \int_0^\infty \frac{(1 - F_G(z))^j (F_G(z))^{N-j}}{1+zP} dz \quad (4.19)$$

One of the main goals of multicast is to assure to all users in the multicast application the reception of all content. Since only a subset of L users receives the information correctly during each time-slot, the remaining $N - L$ users are affected by a burst (of size equal to the time-slot) of block erasures. [47] intelligently proposes Reed-Solomon codes and Raptor codes as correction strategies for recovering the message in the Link layer. Therefore, information of the needed redundancy that allows correcting erasures is sent from physical layer as part of the cross-layer strategy. However, up to now, the erasure correction is out of scope of this chapter, and will be addressed further. Note that we study the opportunistic multicasting method for i.i.d. channels, which is unfair for non-i.i.d.. However, we compare HA-SC and TS for the same type of channel conditions, and thus, opportunistic multicasting fits perfectly for this aim.

4.4 Mobile Hybrid Scenario

In a Land Mobile Satellites (LMS) propagation channel, as the one analyzed in [15], three different shadowing types are distinguished. Very low variations corresponding to the events caused by the environmental features (open area, urban, tree shadowed), which are represented by the three states-Markov chain. Slow variations caused by non-uniformities of the obstacles (lognormal distributed). And fast variations due to multipath, which are represented by Rice distributions. It is assumed that in the direct-satellite link of DVB-SH each receiver can perfectly estimate the CSI and can reliably feed back to the sender in Round Trip Time (RTT/2). Therefore, the NCC, in charge of the Radio Resource Management (RRM) of the satellite bandwidth, can reallocate the resources with a delay of at least the RTT (around 500 ms), but depending on the network congestion, users' mobility, etc., it might be higher. This delay is prohibitive in some cases when the resource allocation algorithms try to adapt to channel conditions, mainly because they take advantage of the channel knowledge to optimize the multicast transmission. E.g. high speed mobiles (50-100 km/h) are affected by fast fading channels with coherence times in the order of a few milliseconds.

In order to define our mobile scenario, we need to take into account two issues defined in previous chapters. The first one is that one of the practical ways to perform the principle of SC is the Hierarchical Modulation scheme. The latter is the fact that DVB-SH specifies two physical layer configurations: OFDM implementing hierarchical modulation for both the satellite and terrestrial links, and TDM implementing ACM for the satellite link.

We distinguish three different scenarios depending on users' mobility (according to its speed) and transmission. We define our system model as in Figure 4.1, which consists of a GEO Multibeam Broadband Satellite link for the DVB-SH satellite link, with the following scenarios:

1. A direct link using TDM air transmission that feeds a pedestrian/fixed scenario, with slow channel variations. NCC is able to adapt thanks to CSI.
2. A direct link using OFDM air transmission that feeds a mobile scenario, with fast channel variations. HA-SC is able to serve mobile users efficiently with non-CSI.
3. Indirect link using OFDM air transmission that feeds a mobile scenario, with fast channel variations. In this case, the BS receives the CSI from the users with a delay below the coherence time of fast fading (e.g. in 3G Long Term Evolution standard, the latency is below 5ms). Multicast algorithms (including HA-SC) can be adapted according to channel conditions with or without CSI.

However, note that it is a shame that TDM transmission (usually associated to broadband connectivity, e.g. K_a band) could not serve mobile users, taking profit of more available bandwidth. Therefore, we propose to use HA-SC algorithm into the TDM mode, meaning that ACM and HA-SC must work together in order to improve transmission scheme performance. In fact, Hierarchical Modulation has been already included in the guidelines of satellite standards such as DVB-S2 [18]. In particular, it is proposed to use it as a backward compatible mode for DVB-S receivers. However, compatibility between ACM and HA-SC might be included in future research because:

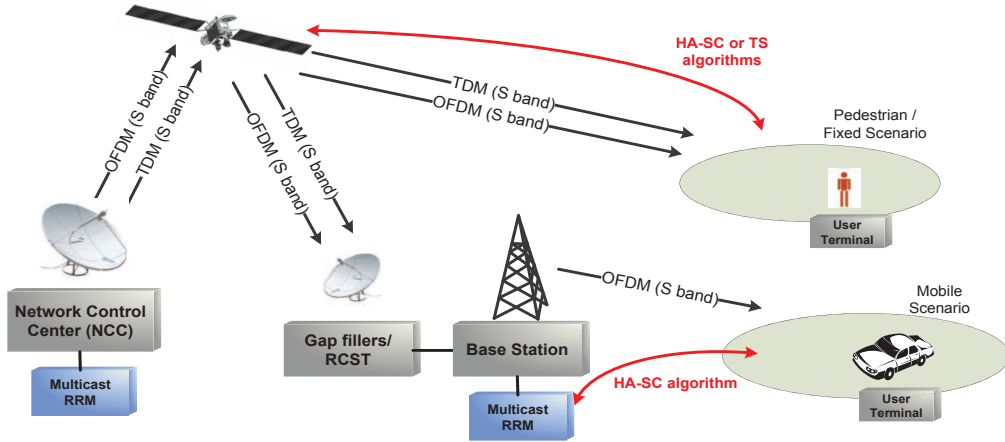


Figure 4.1: Hybrid DVB-SH Architecture with the proposed RRM for Multicast Transmission

- Broadband connectivity could be used in mobile scenarios.
- The ACM is able to efficiently adapt to slow channel variations and weather conditions, while HA-SC maximizes the capacity within each ACM mode by adapting to fast channels.

We must highlight again that, in this chapter, we focus only on the physical layer and the access schemes. Further chapters shows how upper-layers are able also to counteract with slow and fast channel conditions.

4.5 Results

In this section, we analyze the performance of the algorithms for different channel conditions. In order to fairly compare both algorithms, we normalize the total throughput obtained by the throughput achieved when only the best user is served, i.e. $R(g_{max})$ and $T_L(1)$ for HA-SC and TS respectively.

In our case, we propose to study the performance of the constrained scheme following a Rice distribution, for different direct signal's amplitude and diffuse multipath, in order to simulate different scenarios. Moreover, we have considered the effect of different channel environments for LMS at S-band, in particular, those in [15], previously shown in Figure 3.2. The total power constraint to be allocated is $P = 100$ and among $S = 5$ multicast groups.

Figure 4.2 shows the trade-off to be considered in HA-SC algorithm between the minimum rate allocated to the group with worst shadowing conditions (group 5), and how this affects the other multicast groups, where the groups reception quality is sorted as follows: $(SNRg1 > SNRg2 > \dots > SNRg5)$. We can observe that, when minimum expected rate of group 5 is increased around 30%, the capacity of the best group in terms of SNR is reduced around 5%. Following with HA-SC, in Figure 4.3, it is shown the performance as the number of users increases, considering infinite coding layers. Here,

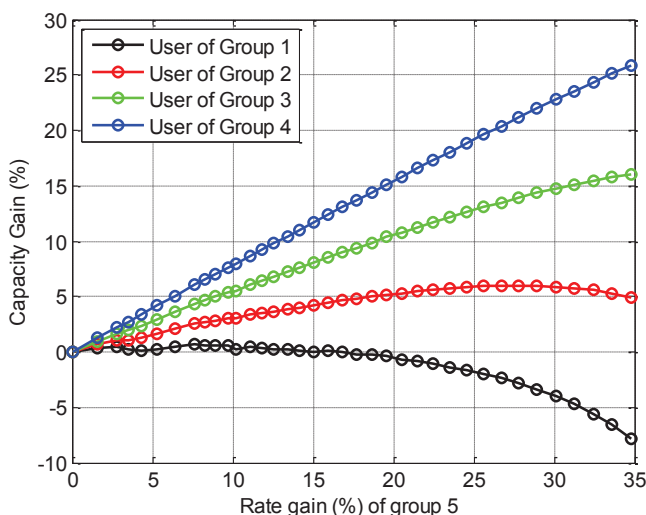


Figure 4.2: Capacity gain of multicast groups when increasing the minimum rate allocated the worst group (group 5), using HA-SC algorithm

users are equally distributed among groups, where users 1-4 remain to group 1, 5-8. to group 2, etc. The effect of channel conditions is clearly differentiated in terms of capacity. We can observe that when all users are served, the bad channel conditions can make the throughput reduce in 30%. The results obtained in Figure 4.3 are considered optimal since it has been used an infinite number of SC layers. Practical transmission modes based on SC such as hierarchical modulation uses 2-3 coding layers. E.g., in DVB-SH standard, the hierarchical system maps the data onto the 16-QAM in such a way that there is effectively a QPSK stream (high priority) buried within the 16-QAM stream (low priority). We have developed the algorithm in order to simulate the capacity achieve for a system based on hierarchical modulation. Results achieved are shown in Figure 4.4, where it is represented the throughput achieved as the number of coding layers (SL) increases. Note that the capacity achieved with HA-SC algorithm is reduced to 25% using only two layers, however, for $SL = 5$, the theoretical capacity is almost achieved. We also examine the TS algorithm, which is analyzed for different channel conditions (Figure 4.5), changing the subset of users served (L) among the total number of users ($N=20$). Unlike HA-SC algorithm, where throughput increases always with the number of users, in the TS algorithm only a subset of users achieves the maximum throughput. It should be stressed that the loss of direct signal amplitude not only makes decreasing the maximum throughput achieved but also the optimal number of users in the multicast group (L) is reduced. We show that the optimal solution can strongly improve the efficiency gain of the system with respect to the worst case user (i.e. where all users are served). However, comparing both algorithms (Figure 4.4 and Figure 4.5) for the deep shadowing. It is shown that HA-SC outperforms the TS algorithm, since in TS, users are constrained to the worst user (L) of the chosen subset, while HA-SC takes advantage of the SC transmission to improve the performance. Note that the main differences between the two proposals are found when increasing the number of users served.

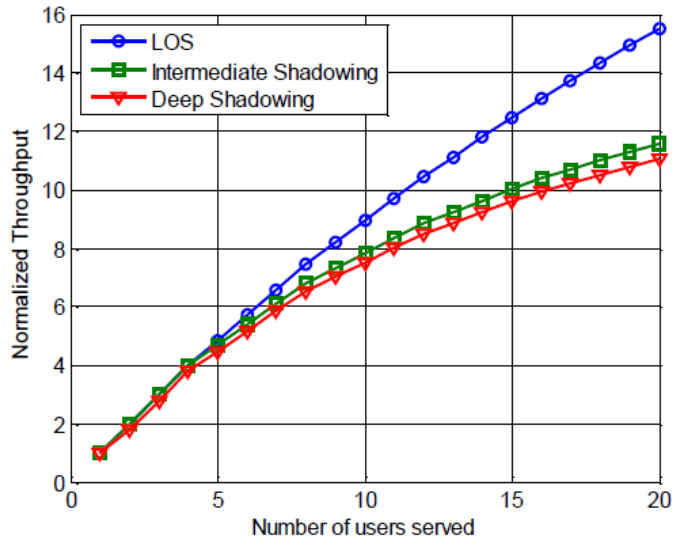


Figure 4.3: Normalized throughput achieved with HA-SC algorithm depending on the number of users served and for different channel condition

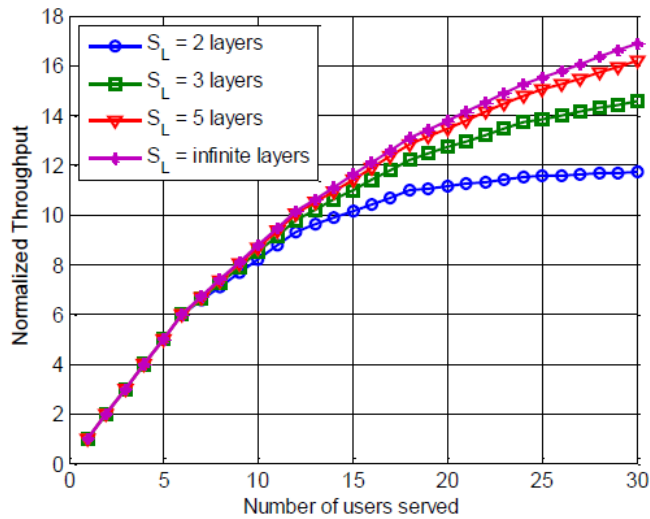


Figure 4.4: Normalized throughput achieved with HA-SC algorithm depending on the number of users served and for different number of SC layers.

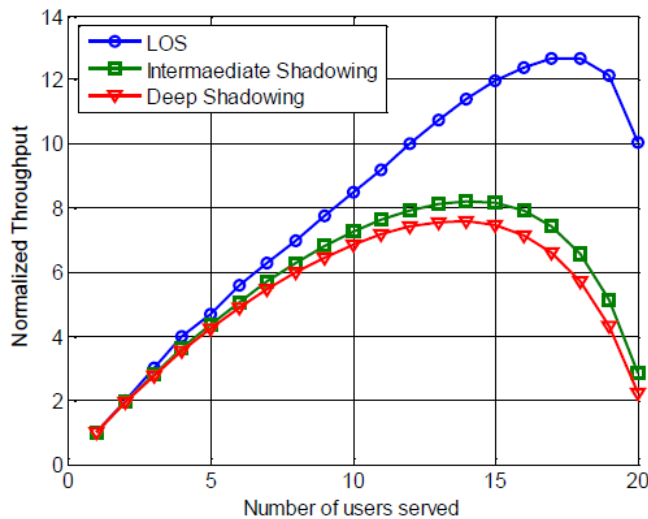


Figure 4.5: Normalized throughput achieved with TS algorithm depending on the number of users served (L) and for different channel condition

4.6 Conclusions

4.6.1 Conclusions on the methodology

The cross-layer interaction defined in this chapter is the typical one used in the already defined top-down approaches, where higher layers dictates the MAC parameters and strategies (in our case, the minimum allocated rate). A summary of the proposed methodology is detailed next.

1. Objective: Maximize the capacity of multicast transmission constrained to serve a minimum rate to the worst group of users, and when non CSI is available at the transmitter. The minimum rate allocated to the rest of groups can be also a constraint of the rate/power allocation problem. A mobile scenario is assumed.
2. Layers involved: PHY, MAC and upper layers (IP, Transport or APP), as depicted in Figure 4.6, where blue lines represent the input flow parameters of the cross-layer design, while the red-dashed lines are the allocation/output/result flows. Upper layers specify minimum rate allocated to MAC layer, which performs the optimization and the transmission according to the obtained solution, which is based on SC.
3. Mathematical tools: A NUM that allow us to state the power allocation problem as a function of the interference and different channel distributions. It is solved using the Lagrange multiplier method for functional subject to integral constraints.
4. Formulation: The constraints are the maximum available power to be distributed and the minimum rate for the group under worst channel conditions.

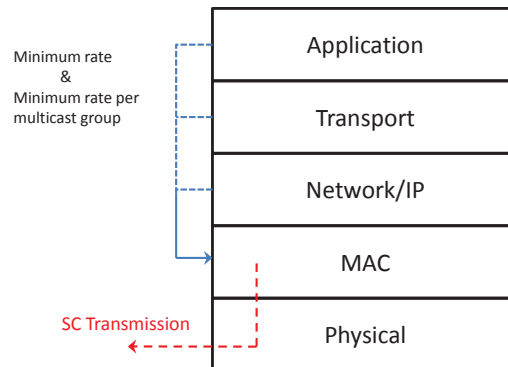


Figure 4.6: Cross-layer interaction and methodology conclusions: HA-SC

4.6.2 Conclusions on the performance gain

The results show that the hierarchical allocation designed problem maximizes the capacity when non CSI is available at the transmitter.

We have demonstrated that although TS algorithm increases overall throughput with respect to the worst case user, the HA-SC algorithm greatly outperforms TS by means of the hierarchical allocation transmission scheme.

We have finally justified the design of a hybrid systems able to give service to different mobile scenarios. We have argued also about the possibility of efficient cooperation between ACM and the HA-SC algorithm for mobile users (fast channel conditions) under different weather conditions (very slow fading). The problem of weather conditions adaptivity by means of ACM and the cross-layer interaction with higher layers is addressed in the following chapters (Chapters 5 and 6). The Hierarchical Modulation scheme performance (at the PHY layer) will be also analyzed within a cross-layer design of Chapter 6, in particular, it is included in a joint unequal protection scheme that aims at adapting to different LMS scenarios.

Chapter 5

Cross-Layer Design at PHY/MAC/IP Level

5.1 Objective: Capacity Maximization with QoS provision

The objective of the chapter is to design an efficient cross-layer scheme capable to deal with the capacity drop when satellite link is affected by different weather conditions. In particular, the chapter includes a cross-layer design that allows the resource allocation (identified as a scheduler) at MAC/IP layers to follow the adaptivity of the physical layer. Such interaction with different IP services makes the QoS requirements a challenging constraint when maximizing the capacity and bandwidth allocation problem, since we have to face not only with channel diversity but also with content differentiation. In this case, control over fairness is a key issue for managing the overall network operation while maximizing user satisfaction. The design including PHY/MAC/IP layers will be extended to the Transport and APP layers in Chapter 6, where real higher mechanisms will be considered.

Note that this design is implemented for the forward link of a broadband satellite system. Since the return link design is a relative straightforward procedure, in the second part of the chapter, we not only proceed to maximize the capacity of the return link, but we also propose to partially decentralize the resource allocation by means of signaling information available in the return link. With such decentralization, we seek for a new proactive-way performance of user terminals, instead of the typical reactive properties of totally centralized networks.

The chapter also shows how the bandwidth allocation problems, and in particular, those that need to fulfill some QoS requirements, can be modeled by using two of the most important and well-known mathematical optimization frameworks; NUM and Game Theory, which are used for the forward and return links respectively.

5.2 Forward Link

5.2.1 Methodology: Adaptive Fairness

5.2.1.1 Cross-layer related work

In wireless communications, most cross-layer works aim at improving the overall protocol stack performance in order to counteract with the channel dynamics. Such problem is frequently coordinated by an implicit scheduler at the MAC layer [48, 49]. In particular, in [48], authors propose a cross-layer packet scheduling that priorities services according to a combined delay-rate criteria algorithm and optimizes the transmit power at the PHY layer. In order to satisfy the variation of statistical delay-bound QoS requirements, [49] make use of a cross-layer scheme that dynamically assigns power-levels and timeslots for heterogeneous real-time mobile users.

The rate/delay trade-off has been also addressed in cross-layer architectures employing ACM-based transmission schemes [50, 51], where cross-layer design must include multiuser scheduling at the MAC layer in order to guarantee QoS. In such case, the perceived QoS per user or per multicast group depends on the channel state. Alternatively, multicasting diversity can be also understood at the APP layer, by means of Scalable Video Coding (SVC). For example, [52] focus on maximizing the decoded video quality of multiple users engaged in simultaneous real-time streaming sessions users, but explicitly considering the delay constraints impact. The interplay of SVC at APP layer and hierarchical modulation at PHY layer, both allowing multi-user diversity, has been demonstrated to improve transmitted video quality [53], but no delay constraints are considered in these works.

As previously introduced, ACM and layered video coding are suitable mechanisms to implement the multicast transmission strategy, however the interplay of both has not yet been solved. In this dissertation, we propose a joint cross-layer design that includes both mechanisms, interconnected by an adaptive MAC scheduler, which is able to fairly and efficiently allocate resources, while considering the requirements of constrained video application in terms of video quality and delay. The MAC scheduler and the interaction with ACM are defined in this chapter, and the interplay with SVC is aimed in the next chapter.

5.2.1.2 Network Utility Maximization related work

Extended research has been carried out in the area of rate allocation, and many research works have put into focus the [54] study, a well-known contribution for elastic traffic. Latter research has led to a more general optimization framework known as NUM, where elasticity can be modeled by smooth and strictly concave utility functions. Such utility function makes the problem convex and thereby tractable for optimality analysis. Moreover, it allows obtaining a good compromise between efficiency and fairness by implementing different priorities and pricing functions. Therefore, the NUM framework looks the right tool to trading delay and rate.

NUM has been used as an analytic tool for TCP congestion control, and in particular, it has been proved that it can be seen as a distributed primal-dual algorithm to maximize aggregate utility [55]. The basic idea is to regard source rates as primal variables and

congestion measures (such as delay) as dual variables. Recent literature [56] on distributed solutions has also shown that decomposition methods provide helpful mathematical tools, allowing to represent distributed dual-algorithms and making them converge to a global optimum. Basically, dual decomposition is appropriate when the problem has coupling constraints such that, when relaxed, the optimization decouples into several subproblems. In addition, the framework of layering as optimization decomposition [57] has allowed to integrate the various protocol layers into a single coherent theory (congestion control, routing, random access, power control, etc). However, it is also known that duality and decomposition have major drawbacks when the application is inelastic and utility function are non concave, leading to divergence on congestion control. A solution has been proposed in [58], where, despite the non concavity of the utility functions, the price-based distributed algorithms can still converge to the globally optimal rate allocation by looking at specific conditions.

New delay-constrained and rate-sensitive applications such as real-time or video streaming (i.e. inelastic traffic), make this type of utility function non-concave, resulting in non-convex NUM [59], which are specially difficult to analyze. Note that if algorithms for elastic traffic are now applied to inelastic traffic, this may result in instability, excessive congestion and low QoS. Hence handling of inelastic traffic remains an open problem for NUM. Recent studies have addressed inelastic applications through modeling the traffic characteristics. In [60], a content aware-distortion fair networking framework with joint video source adaptation and network resource allocation is developed. The main drawback is that they do not consider explicit utility functions. However, they take profit of the special characteristics of video content such as dependency between frames to build a content-aware and time varying utility function. More recently, in [61], they propose a multimodal sigmoid approximation in order to deal with the non-concavity of the staircase utility functions used for scalable applications, achieving a suboptimal efficient solution.

To overcome this problem, we assume a NUM that is able to adapt the rate allocated to each video application according to dynamic priorities. Our goal consists of formulating those priorities according to cross-layer information that includes delay measured and QoS requirements of each application. Moreover, in a second design detailed in Chapter 6, we modify the priority, not only allowing a rate-delay balancing, but also using a fair criteria to allocate the rates to be transmitted to users according to channel conditions.

5.2.1.3 Contributions

We propose a cross-layer queuing and scheduling architecture at the MAC/IP layers for a QoS-aware transmission scheme (DiffServ).

- We first focus on a NUM to formulate the rate/delay balancing when different video application flows are feeding the MAC scheduler. The utility of each flow is prioritized according to the corresponding video quality and delay requirements. The solution happens to be a water filling load across the scheduler architecture, which is known to enjoy low-complexity implementation.
- We prove that the proposed NUM-based rate/delay trade-off is feasible thanks to a cross-layer design including PHY, MAC and IP layers. It is straightforward to see

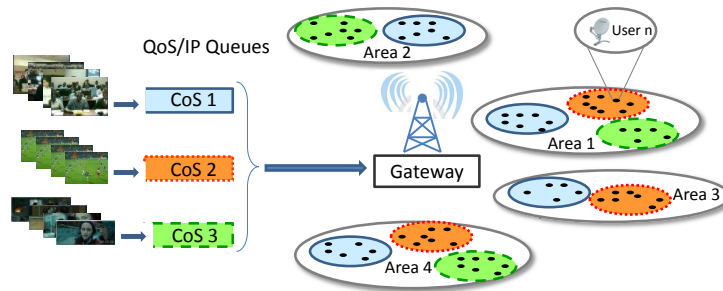


Figure 5.1: Video transmission scenario over a wireless/satellite link, multicast groups organized per video application requested and per geographical area.

that our design is backward compatible with any wireless transmission scheme that implements ACM. Finally, we contribute with a detailed protocol design suitable for the standard specification of multi-beam broadband satellite communications, i.e. DVB-S2.

- The final NUM scheme is evaluated for realistic satellite channel conditions and compared to other scheduling solutions. More experimental results are included in Chapter 6 for a multicast scenario and according to video content requested.

5.2.2 Satellite/wireless network model

We consider a generic air interface, as the one depicted in Figure 5.1, where independent video applications flows are transmitted by the Gateway (or base station) through the wireless link. For the sake of simplicity, we suppose that the video encoder is directly connected to the Gateway, where each video content is classified into a CoS queue depending on the QoS requirements; assuming that traffic requirements (e.g. in terms of maximum delay, packet loss or jitter) of each CoS must be fulfilled. At the receiver side, users are organized in multicast groups, according to geographical areas (i.e. channel conditions) but also depending on the type of video content requested (colored ellipses).

5.2.3 Architecture, physical and link layers

We assume the adoption of ACM at the physical layer, which takes profit of the users diversity (as in Figure 5.1), i.e. the channel conditions experienced by each user (time and location-dependant), which increases the overall transmission efficiency. We assume M different ACM modes, each one defined by an appropriate coding and modulation, whose spectral efficiency, transmission time and buffer length are denoted by η_m , T_m and L_m , respectively. However, despite such efficiency increase, information sent to users under severe channel conditions still need to be more protected (either adding redundancy or reducing modulation efficiency) in order to keep PER within requirements. Therefore, when users are affected by rain/clouds, the Gateway needs to increase the transmitted information in order to send the same useful information as they were sending before being affected by weather/channel conditions, overloading wireless link and causing a capacity drop. In a satellite scenario using K_a band, capacity drop becomes very significant; up

to 40% and 60% in temperate and tropical areas respectively, causing larger delays. This implies that a fair and efficient allocation must be designed in order to distribute the available resources.

At the MAC layer, the system implements a TDM access scheme, each time slot contains a PHY layer packet, with a constant amount of coded symbols but an ACM mode-dependent number of information bits and symbols: the consequence is a variable packet transmission.

5.2.3.1 IP Traffic model

In wireless communication networks, opposite to Internet, it is imperative to guarantee various QoS levels and a minimum per user bandwidth. The video application flows can be classified in, without loss of generality, three CoS types of differentiated quality. EF to provide premium service to those applications that require a robust network treatment such as Videoconferencing. AF, which are suitable for less constrained traffic in terms of delay and jitter, such as (High Definition) HD video streaming. And finally, BE traffic flows with non QoS or low QoS requirements, e.g. video file transfers, etc. It is clear that we can assume a DiffServ QoS model, thus enabling easy integration with IP external networks.

5.2.4 QoS PHY/MAC/IP-layer scheduler: DiffServ

We propose a DiffServ architecture fed by a layered video encoder (at APP), which serves C different video application flows. Each video flow is enqueued into the CoS queues (at IP layer) according to the quality requirement/treatment that best matches. The Fast Round Robin (FRR) schedulers classify flows coming from CoS queues into the ACM buffers, according to the Channel State Information (CSI) of destination users. We assume that each buffer is associated to an adaptive physical layer (ACM mode). Finally two levels of PHY/MAC schedulers, detailed next, are able to serve the time-varying ACM as depicted in Figure 5.2. In α -fairness Weighted Round Robin (WRR) and the Adaptive-WRR (AWRR), priorities are correlated to the channel conditions and the delay constraints, respectively.

5.2.4.1 α -fairness WRR: channel conditions

Although the notion of proportional fairness for elastic traffic was first introduced [54], and the fair bandwidth sharing has been also extended, it only focuses on the link's capacity, while in our case, we propose a fair pricing correlated with the channel diversity.

First, let us assume that the scheduler policy is a WRR. Let us define $W = (w_1, w_2, \dots, w_M)$ as the weights (i.e. priority) assigned to each physical layer (or ACM mode), with $\sum_{m=1}^M w_m = 1$. Assuming that physical layer m can provide a maximum capacity of $\eta_m R_s$ bits/s where R_s is the carrier symbol rate. The expression of the throughput S_m of ACM mode m is:

$$S_m = \frac{w_m L_m}{\sum_{j=1}^M w_j T_j} = \frac{w_m L_m}{\sum_{j=1}^M w_j L_j / \eta_j} R_s \quad (5.1)$$

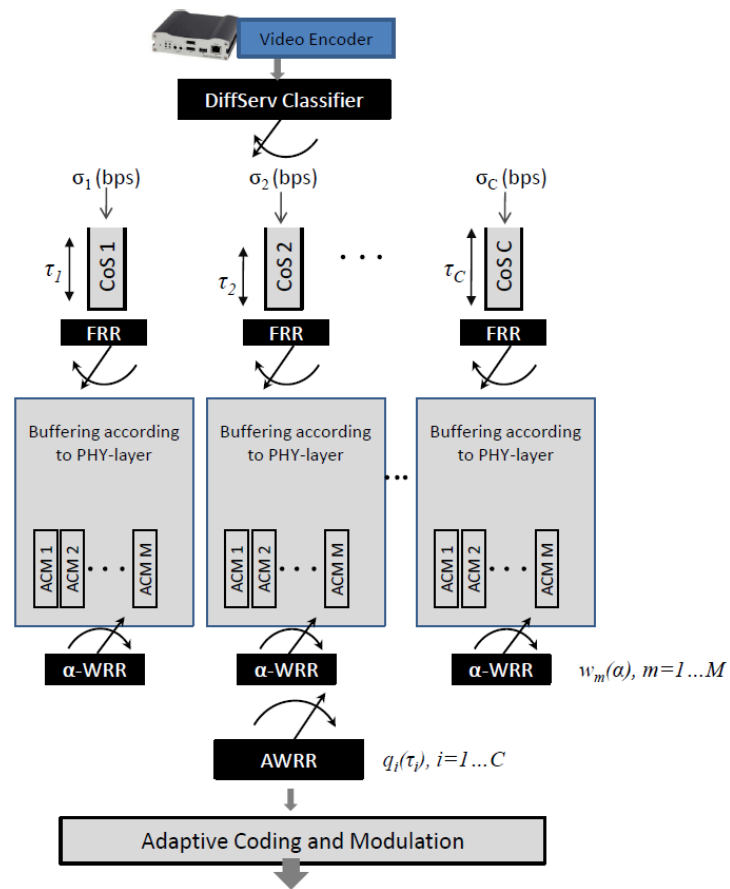


Figure 5.2: Cross-layer scheme of the adaptive PHY/MAC/IP schedulers

The features of each ACM mode, i.e. η_m , T_m and L_m , are the spectral efficiency, transmission time and buffer length respectively, previously introduced in Section 5.2.3.

According to [62] and the European Space Agency (ESA) Project [63], the weight can be expressed as a function of the ACM mode features, making it possible to dynamically adapt the bandwidth requests allocation depending on the channel conditions (including capacity drop) in order to guarantee fairness. As a result, the weight of a user i using the ACM mode m can be expressed as follows:

$$w_m(\alpha) = \frac{\frac{1}{T_m \eta_m^\alpha}}{\sum_{i=1}^M \frac{1}{T_i \eta_i^\alpha}} \quad (5.2)$$

Note that the weight policy is controllable by α . By setting different policies, we can give fairness in terms of throughput sent when $\alpha = 1$, which leads to $S_m = S_n \forall m, n$, or time shared when $\alpha = 0$, leading to $w_m T_m = w_n T_n \forall m, n$. Negative values of α are more efficient but also unfair in terms of transmitted throughput, since more priority is given to users under good channel conditions.

5.2.4.2 Adaptive WRR: delay aware

The AWRR scheduler is in charge of guaranteeing the delay requirements per CoS according to the weight q_c allocated to each CoS queue. q_c is defined as the percentage of time the queue c is accessed by the AWRR, which transfers the packet to the PHY-layer. Intuitively, it seems possible to design a policy that allocates weight q_c based on the delay τ_c measured at CoS queue c , in order to compensate queue balancing produced by load and capacity variations. We propose the scheme presented in (5.3). The basic idea is that an increase of premium service's (i.e. EF) weight must cause the same amount of decrease in other CoS. To guarantee the EF requirements, the rule applied is to first shift the weight of BE to EF, and if it is not enough, part of AF's weight will be shifted to EF. However, once the delay of EF service backs down, the weights taken will be returned following also (5.3), creating a relationship between DiffServ weights and the queue delay.

$$q_{ef}(\tau_{ef}) = \begin{cases} q_{ef}^{max}, & \tau_{ef} > \tau_{ef}^{max} \\ q_{ef}^{int} + (q_{ef}^{max} - q_{ef}^{int})\Lambda_{AF}, & \tau_{ef}^{int} \leq \tau_{ef} \leq \tau_{ef}^{max} \\ q_{ef}^{min} + (q_{ef}^{int} - q_{ef}^{min})\Lambda_{BE}, & \tau_{ef}^{min} \leq \tau_{ef} \leq \tau_{ef}^{int} \\ q_{ef}^{min}, & \tau_{ef} < \tau_{ef}^{min}. \end{cases} \quad (5.3)$$

where Λ_{BE} and Λ_{AF} are the logarithmic slopes of the shifting functions (from BE to EF and from AF to EF, respectively):

$$\Lambda_{BE} = \frac{\ln(\Gamma_{BE}/q_{ef}^{min})}{\ln(q_{ef}^{int}/q_{ef}^{min})} \quad (5.4)$$

Similarly, Λ_{AF} is obtained. Note that a logarithmic slope makes a smoother queue balancing than the staircase case, and in our case its performance is better than others (such as exponential or linear). Finally:

$$\Gamma_{BE} = \frac{(q_{ef}^{int} - q_{ef}^{min})(\tau_{ef} - \tau_{ef}^{min})}{\tau_{ef}^{int} - \tau_{ef}^{min}} + q_{ef}^{min} \quad (5.5)$$

where q_c^{min} and q_c^{max} are the minimum and maximum weights of CoS c , with c equals to ef , af or be . The intermediate weight $q_{ef}^{int} = (q_{be}^{max} - q_{be}^{min}) + q_{ef}^{min}$, and τ_{ef}^{int} are reached when EF takes the maximum allowed weight from BE. τ_{ef}^{max} and τ_{ef}^{min} represent the maximum acceptable delay and maximum desired delay of EF. Our delay-based design and the logarithmic slope make it more accurate than in [64] when fulfilling QoS requirements of EF and AF. It should be stressed that AF can take weight from BE following a scheme equivalent to (5.3). Finally, since such weights are normalized so as to satisfy $\sum_{c=1}^3 q_c = 1$, the BE's weight is $q_{ef}^{int} = 1 - (q_{ef} - q_{af})$.

5.2.5 Optimization: Rate-Delay balancing across MAC /IP queues

For the first approach of the optimization, and in order to focus on the rate-delay trade-off, we only consider the delay aware AWRR depicted in Figure 5.2 together with video encoder feeding the IP/CoS queues. Although the PHY buffers (and ACM modes) and the α -WRR are also considered, and they affect the performance of the system, their related parameters (e.g. w_m , α , etc.) are no included in the optimization. In fact, in the optimization, we only use these particular parameters to compute the available capacity:

$$C_{link} = \bar{\eta}(t)R_s = R_s \sum_{m=1}^M w_m \eta_m \quad (5.6)$$

where, $\bar{\eta}$ is the average spectral efficiency of the transmission. The α -WRR and the ACM modes will be included in the optimization performed in the Chapter 6, where they will interact with a layered video multicasting at the APP layer.

Based on the cross-layer framework between the MAC scheduler and the APP layer video encoder, we aim at maximizing the video quality sent by each flow constrained to keep the delay requirements.

5.2.5.1 Formulation Framework

First of all, let us assume that a video application flow c is sending at rate σ_c (bps) as a feasible strategy in order to compete for C_{link} . Let S_c denote the feasible strategy set of flow c , then $\sigma_c \in S_c$. And the feasible strategy space is $S = S_1 \times S_2 \dots \times S_C$. Then, a feasible strategy tuple is a C-dimension vector $(\sigma_1, \sigma_2, \dots, \sigma_C)$. The objective of each player is to maximize its utility U_c , which determines the level of satisfaction of each flow by adjusting σ_c . The resulting solution $\forall c$ is denoted as:

$$\sigma_c = \operatorname{argmax}_{\sigma_c \in S_c} U_c(\sigma_1, \dots, \sigma_c, \dots, \sigma_C) \quad (5.7)$$

In our request game, let us consider an increasing, strictly concave and continuously differentiable utility function that has been widely proposed in the literature [65], when $\sigma_c > 0$:

$$\tilde{U}_c(\sigma_c) = \begin{cases} \log(\sigma_c), & z = 1 \\ (1 - z)^{-1} \sigma_c^{1-z}, & \text{otherwise.} \end{cases} \quad (5.8)$$

In particular, we use the case $z = 1$ (logarithmic) [54], which maximizes the utility function and is proportionally fair.

In order to keep a relationship between the traffic model considered and our utility function, we introduce a QoS priority weight, termed $\Omega_c > 0$, in order to implement different CoS, which is suitable for our cross-layer QoS-aware design. As a result, our utility function can be expressed as $U_c(\sigma_c) = \Omega_c \log(\sigma_c)$. Since our objective is to maximize the satisfaction of all users, the utilities can be assumed to be additive:

$$\max_{\sigma_c} \sum_{c=1}^C U_c(\sigma_c, \Omega_c) \quad (5.9)$$

subject to the amount of available capacity in the wireless link C_{link} , i.e. $\sum_{c=1}^C \sigma_c \leq C_{link}$. Note that the NUM is convex due to the logarithmic form of the utility function [66].

Priority weight Ω_c makes each video application dependent on the belonging CoS. However, video quality does not only depend on the rate received but also on delay and jitter among others. In order to fulfill delay requirements, our problem must guarantee a delay τ_c below a maximum delay τ_c^{max} . Since the AWRR scheduler is in charge of maintaining $\tau_c < \tau_c^{max}$, intuitively, one can take advantage of the AWRR weight definition in order to balance not only the queue delay per CoS, but also the video rate sent through each one. Thus, priority weight $\Omega_c(q_c(\tau_c))$ will be subject to q_c and τ_c . Moreover, we constrain each application c with a minimum and maximum required video qualities ($v_{min,c}$ and $v_{max,c}$), resulting in the following NUM:

$$\begin{aligned} \max_{\sigma_c} \quad & \sum_{c=1}^C \Omega_c(q_c(\tau_c)) \log(\sigma_c) \\ s.t. \quad & \sum_{c=1}^C \sigma_c \leq C_{link} \\ & v_{min,c} \leq \sigma_c \leq v_{max,c} \end{aligned} \quad (5.10)$$

From the utility function of each video stream $\Omega_c(q_c(\tau_c)) \log(\sigma_c)$, we can observe the cross-layer interaction between layers, since the bit rate σ_c allocated to each video at APP-layer depends on weight q_c at the MAC-layer scheduler, which is dynamically adapted according to τ_c measured at IP-layer. Although in (5.10), constraint $\tau_c < \tau_c^{max}$ is not detailed, it is implicit in the q_c definition.

5.2.5.2 NUM Solution: Water Filling Rate/Delay balancing

The Lagrangian of our NUM (5.10) is denoted as:

$$\begin{aligned} L = \quad & \sum_{c=1}^C \Omega_c(q_c(\tau_c)) \log(\sigma_c) - \lambda \left(\sum_{c=1}^C \sigma_c - C_{link} \right) \\ & - \mu_c (\sigma_c - v_{max,c}) + \gamma_c (\sigma_c - v_{min,c}) \end{aligned} \quad (5.11)$$

The value of σ_c that solves $\partial L / \partial \sigma_c = 0$ is optimal when Karush Khun Tucker (KKT)

conditions [66] are satisfied. Such solution is expressed as:

$$\sigma_c = \begin{cases} v_{max,c}, & v_{max,c} < \frac{\Omega_c(q_c(\tau_c))}{\lambda} \\ \frac{\Omega_c(q_c(\tau_c))}{\lambda}, & v_{max,c} \geq \frac{\Omega_c(q_c(\tau_c))}{\lambda} \geq v_{min,c} \\ v_{min,c}, & \frac{\Omega_c(q_c(\tau_c))}{\lambda} < v_{min,c}. \end{cases} \quad (5.12)$$

where the value of λ is chosen to satisfy the C_{link} constraint. This solution is termed as *water-filling*, which is, among others, a well-known procedure for optimal power allocation [67, 68]. Herein, water-filling happens to be the optimal solution for the rate-delay cross-layer balancing. Moreover, opposite to other power allocation studies, our design allows us to set more constraints, which is suitable to better re-create real scenarios. However, since 3 possible sets of solution values, 3^C hypothesis should be checked, which is prohibitively large. In [69], they propose a low-complexity algorithm that provides the exact solution of water-filling with an overall complexity of $C \ln(C)$, in the same order of the best sorting algorithms.

5.2.5.3 Interpretation of Water Filling Solution

The following section aims at explaining how water-filling becomes the optimal solution of our NUM, and how to take profit of it as a rate-delay balancing. As illustrated in Figure 5.3, the problem can be seen as 3 flows of containers (one for each video application) arriving to the IP queues (EF, AF and BE). Considering water-filling solution is obtained every time-slot T_{wf} , then, each container is filled with an amount of water corresponding to the rate coming from each video application during each T_{wf} . The volume of water is equal to $\Omega_c \times \frac{1}{\lambda} \times 1$, and note that the level of water ($\frac{1}{\lambda(T_{wf})}$) is equal in all containers belonging to the same T_{wf} , however the volume is different due to the priority weights, $\Omega_c(q_c(\tau_c))$, which allow to follow different policies according to $q_c(\tau_c)$ and delay measured at each CoS queue. At IP queues, we are able to compute the volume of information per queue at time slot kT_{wf} that has not been yet served by the AWRR:

$$Q_c(kT_{wf}) = \sum_{t=0}^{kT_{wf}} \max \left\{ \frac{\Omega_c(q_c(\tau_c(t)))}{\lambda(t)} - q_c(t)C_{link}(t), 0 \right\} \quad (5.13)$$

From (5.10), we know that if C_{link} is constant, then $Q_c = 0 \forall t$. However, if traffic from other applications (apart from the 3 video streams) is transmitted via the wireless link, queues are fulfilled with new packets, and network capacity reduced. Similar behaviour happens in an ACM-based network, where link capacity might be reduced due to fading variations (at time t). Assuming the latter case, let us consider $C_{link} = \bar{\eta}(t)R_s$, where $\bar{\eta}$ is the average spectral efficiency of the transmission and R_s the symbol rate. If C_{link} is reduced during time-slot T_{wf} , then $Q_c(kT_{wf}) \geq 0$, and AWRR scheduler might need to increase EF weight in order to fulfill delay requirements.

The interpretation of such solution help us to understand the priority term Ω_c , which greatly impacts on the performance of the design. We propose a simple and flexible solution with different fairness policies δ_c :

$$\Omega_c(q_c(\tau_c)) = (q_c(\tau_c))^{\delta_c} \quad (5.14)$$

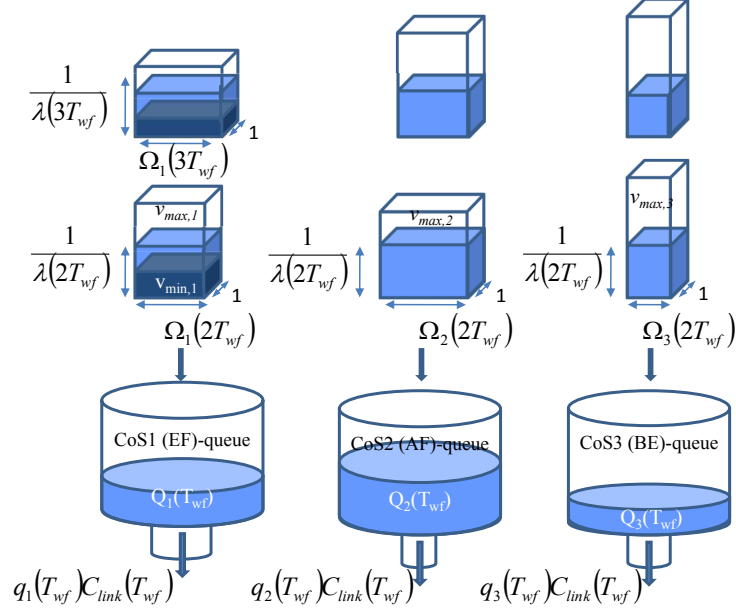


Figure 5.3: Queue balancing interpretation of the water-filling solution method.

For the sake of clarity, let us consider different scenarios where EF, AF and BE have hard, soft and non delay constraints respectively, and equal initial weights $q_c = 1/C$:

1. $\delta_c = 0 \forall c$: No priorities depending on delay constraints, i.e. available bandwidth is equally distributed.
2. $\delta_c = 1 \forall c$: If C_{link} is reduced, AWRP performs weight balancing in order to keep delay constraints, i.e. $q_{ef} < q_{af} < q_{be}$ and more rate is allocated to EF.
3. $\delta_c > 1$ only for $c \in G$: Given that $q_c < 1$, priority Ω_c of CoS $c \in G$ is reduced, which is useful to penalize some classes among the others.

5.2.6 Standard, Protocol Design and Implementation: FairDB

The peculiarities of our design are presented in this section, including the cross-layer architecture and the detailed cross-layer information exchanged among layers in order to perform rate-delay balancing.

5.2.6.1 Transmission standard: DVB-S2

Though, many transmission standards implement adaptive physical layers, WiMAX, DVB-SH, DVB-S2, etc. In this chapter, we consider the forward link of the DVB-S2 [18] air interface for multibeam broadband satellite communication networks, controlled by the NCC. The satellite is assumed to be transparent, and it follows a GEO. We assume transmission in the K_a band (20-30 GHz), and thus rain is the most affecting atmospheric event.

5.2.6.2 Objectives of FairDB

As described before, our Fair Delay Balance (FairDB) cross-layer design, implemented at the NCC, aims at maximizing the video quality of users (grouped according to channel conditions and content requested) depending on traffic and channel dynamics. Given the solution of our NUM in (5.12), the roles of the cross-layer design are two:

1. To provide MAC layer with delay statistics and channel conditions of all Satellite Terminals (STs).
2. To inform upper layers (IP layer) about the bandwidth available at the MAC layer.
3. To make sure that the allocated video rates are correctly transferred to upper layers (APP layer).

5.2.6.3 SI-SAP Architecture and Flows

Our cross-layer architecture is based on standardized tools implemented at the Satellite Independent Service Access Point (SI-SAP) [9]. SI-SAP is the interface between the satellite dependent (SD) lower layers (composed of PHY and Data Link layer, including the MAC layer), and the satellite independent (SI) upper layers (composed of IP and higher layers). Briefly explained, the SI-SAP interface provides the Broadband Satellite Multimedia (BSM) with a layer of abstraction for the lower layer, and it allows the BSM protocols developed in the SI layer to perform over any BSM family.

The SI-SAP interface makes use of the SI-SAP primitives [10] to define the exchanged information between the SI upper layers and the SD lower layers. Primitives can be classified into functional groups [70] depending on whether they are related to data transfer, connection/session management, resource management or security. SI-SAP interface is also composed of Queue IDentifiers (QIDs), which should be seen as a generalization of the way IP queues and SD queues are associated, and a way of hiding specific SD layer implementations from the IP layer. Each QID offers a defined type of service for transfer of IP packets to the SD layers, as well as a means of forwarding packets to different BSM ST destinations. Note that each time a new QID is created, its QoS specifications should be defined: token bucket rate, token bucket size, data rate, maximum packet size, etc.

According to our design in Figure 5.4, we define four cross-layer control flows:

- XL_{IP-MAC} flow from IP to MAC layer containing the status of IP queues.
- $XL_{APP-MAC}$ flow from APP to MAC specifying the minimum ($v_{min,c}$) and maximum ($v_{max,c}$) qualities per video application.
- $XL_{MAC-APP}$ flow from MAC to APP containing the allocated rates per video application (σ_c), which are computed by FairDB block.
- $XL_{PHY-MAC}$ flow from each ST to MAC layer of NCC containing the Channel State.

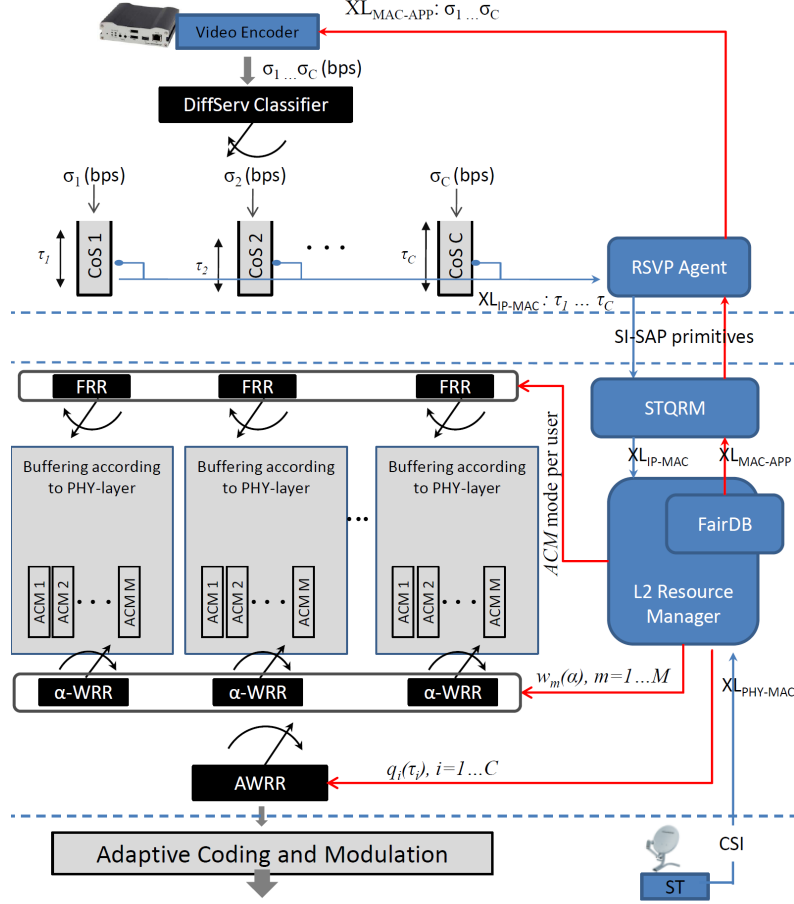


Figure 5.4: Architecture of the cross-layer design, including exchanged information and per-layer block description.

The new entities depicted are the RSVP (Resource ReserVation Protocol) Agent [71], which is aware of IP resources (queue status, availability) and manages the new admissions, negotiates specifications and policies. It is in charge of translating new requests and information flows (such as the queue status of XL_{IP-MAC}) to and from primitives across the SI-SAP. The STQRM (ST QID Resource Manager) [10] is mainly responsible of translating primitives arriving at the SI-SAP. Finally, the L2 Resource Manager is responsible of computing the weights for AWRR and α -WRR schedulers by using the XL_{IP-MAC} and $XL_{PHY-MAC}$ cross-layer flows. Moreover, it performs the MulFairDB.

5.2.7 Performance Results

In this section, we aim at showing the performance of our FairDB cross-layer design, including the trade-off between delay and rate received solved in (5.12).

5.2.7.1 Per-layer parameters

Realistic channel conditions at K_a Band, of 20-30 min. duration, are used in order to get a good estimation of the delay experienced and the rate received by 100 users. In particular,

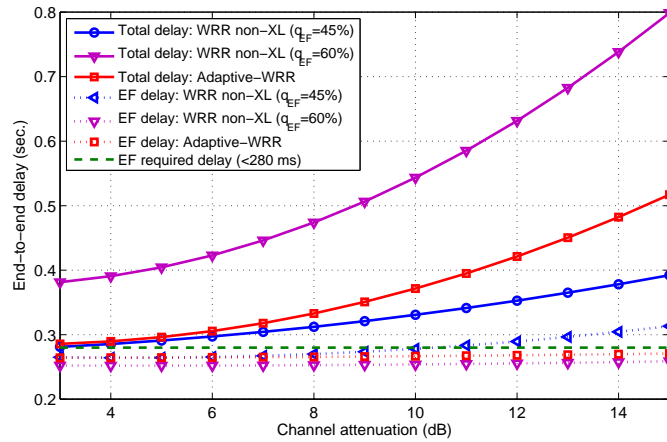


Figure 5.5: Comparison between AWRR and WRR in terms of EF delay and total delay for different maximum rain attenuations.

we assume users diversity in terms of channel statistics, where they are affected by deep rain (maximum 15 dB), low rain (maximum 9 dB) and clearsky attenuation. Users are equally distributed among the weather areas.

Delay requirements at the IP-QoS model for EF and AF are $\tau_{ef}^{max}=280$ ms and $\tau_{af}^{max}=400$ ms respectively, in terms of maximum allowed delay. Note that the end-to-end measured delay takes into account: the delay introduced in the MAC/IP queuing and scheduling, the transmission delay, the propagation time (RTT/2, around 250 ms). One different video application will be sent per CoS: Videoconferencing (EF), TV-On line (AF), and Video Streaming (BE). Additional Poisson-distributed traffic is used to simulate heavy congestion.

5.2.7.2 Simulation Results

In order to understand the goals of our work, let us first introduce the advantage of using our cross-layer AWRR scheduler, which performs the load balancing of MAC/IP queues. A comparison between the AWRR and common WRR schedulers can be observed in Figure 5.5 in terms of delay experienced by users receiving EF and by all the users (termed as total delay). WRR, which does not benefit from cross-layer (XL), is no able to adapt to traffic dynamics and need to allocate a static weight to EF. Such strategy happens to be not efficient, nor underweighting ($q_{EF}=45\%$) neither overweighting ($q_{EF}=60\%$). In the first case, because it does not fulfill the delay requirements (lower than 280 ms), and the latter one because it penalizes other applications unnecessarily. On the other hand, AWRR is not only able to work as a trade-off between both, but it also controls the rate allocated to each CoS at the APP.

Further results are included in Chapter 6.

5.3 Return Link

5.3.1 Methodology: Adaptive Fairness

5.3.1.1 Centralized vs. Decentralized Satellite Networking

Designing fair and efficient bandwidth-on-demand allocation schemes in order to meet current heterogeneous traffic requirements are of paramount interest. Though many transmission standards have been designed to follow a centralized model, the debate between centralized and decentralized networks is still open. Moreover, new intelligent/interactive distributed mechanisms has led to a renewal interest in decentralized networks. In particular, Internet is considered itself the largest decentralized existing network. Main pros of decentralization are scalability (avoiding bottlenecks at the central entity, which limits number of users) and signaling overhead reduction. Moreover, robustness was the main design criteria in the early 1980s, focused on designing a network communication protocol less vulnerable to attacks/failures. Trust and fairness are also important, since if decisions are decentralized, users can participate in the final allocation.

In the case of wireless networks, they can be sorted from totally centralized to distributed. Wireless mesh networks are defined distributed while wireless sensor networks can operate in both ways. Cellular networks and local area networks, such as WiFi/WiMAX were initially designed to work in a centralized manner, however, recent research has shown that operating in a distributed manner is also possible in terms of time slot allocation [72] and resource management [73].

Traditionally, satellite communications networks are totally centralized. A centralized approach seems to be more appropriate due to the simplicity of its implementation and robustness compared to distributed control. Specially in satellite networks, due to the possible mobility and time varying channel conditions, connectivity and capacity of up/downlinks are changing dynamically and randomly. Satellite transmission must be adaptive to nonuniform and time-varying traffic and fading channel states, while meeting QoS requirements. In the literature, different solutions proposing to perform centralized allocations are available; [74] aims at evaluating the impact of channel dynamics at K_a Band and proposes control mechanisms based on a trade-off between call blocking probability and guaranteed capacity to compensate rain fade, while in [75], authors are focused on the dynamic time-slot allocation by means of a cross-layer framework, which optimizes the overall system efficiency and minimizes signaling overhead.

On the other hand, the large propagation delay inherent in satellite link due to Round Trip Time (RTT) makes centralized algorithms inefficient for handling delay-sensitive services. In fact, in the centralized case, the controller is in charge of computing the bandwidth allocation, which is broadcast to the ST after their requests have been received, causing a request-allocation delay of at least two RTT. In [76], authors highlight that a distributed demand control assignment for Time Division Multiple Access (TDMA) schemes can outperform a centralized one in terms of end-to-end delay. The large and increasing number of users to be served by satellite link is also a drawback, since the central entity needs full knowledge of the network conditions (i.e. varying traffic requests and channel conditions per user), the signaling overhead through the network renders

remarkable.

However, changing the access scheme to distributed is a rough task for already defined satellite standards, which include a complete description of signaling, request and allocation procedure. For instance, in the return link of broadband satellite networks, the most common bandwidth allocation belongs to the class called resource reservation on-demand [77] or Demand Assignment Multiple Access (DAMA). In DAMA, the control center performs the bandwidth allocation algorithm according to the STs requests received, and sends back to the STs the resulting allocation (with at least one RTT after the request was sent). In such a scheme, STs are unaware of the allocation policy and the overall system resources, and hence their bandwidth requests are performed according to incoming traffic requirements, which can be considered to follow a fully selfish behaviour, unaware of the future bandwidth assignment. To overcome this problem, we propose to partially decentralize the bandwidth allocation by means of bandwidth requests games performed by STs. We assume STs compute the bandwidth requests according to truly available standardized signaling from the central entity; satellite link capacity and/or the overall requests depending on the game carried out. Our solution allows STs to adapt to capacity changes according to the bandwidth allocation granularity, improving the management of the incoming traffic. Moreover, due to the responsible-bandwidth requests (i.e. not selfish), we will show that one of our goals is a signaling overhead gain reduction from the central entity to the users.

Therefore, in order to summarize, since potentially there are many STs distributed in the network which will be competing for the use of the available bandwidth, there are several issues which arise and must be dealt with: 1) efficient bandwidth requests; 2) the crucial notion of fairness among users (correlated or not to capacity drop); 3) the ability to implement the requests in a distributed manner for different satellite network state knowledge; and 4) appropriate pricing and constrained models capable to adapt to channel and traffic dynamic according to 1) and 2).

5.3.1.2 Game Theory related work

In distributed resource allocation scenarios, each terminal performs its requests aiming at satisfying its traffic requirements. They may even perform overestimated requests for the selfish sake of their own expectations. In any of these cases, the overall limited bandwidth allocation will become suboptimal and unfair. In the broadest sense, game theory [78] has demonstrated to be a natural framework with a collection of mathematical models formulated to study situations of conflict and cooperation. Therefore, game theory looks the right tool to control the selfish behaviour of greedy users, and to help aiming at fulfilling the four issues explained above. There is a substantial literature of distributed game theoretic approaches to solve capacity drop and resources allocation problem.

In wired networks, explicit game theory tools, such Nash Equilibrium (NE), has been proposed in [79, 80] in order to evaluate the selfish behaviour of users and to propose priced utility function in terms of delay and loss rate. However, NE does not assure that distributed users reach the optimal operating network point, and thus happens to be Pareto inefficient in many cases. In contrast to this, the interest for the Nash Bargaining concept has been recently increased thanks to its efficient and fair properties. For instance,

in broadband networks games [81], where authors present a game theoretic framework which puts into focus [55] and [54] works, and goes further showing how a joint allocation and a pricing policy can be efficient and present nice fair properties. Distributed solutions have been also addressed for TCP, and in particular, it has been proved [55] that TCP congestion control is seen as a distributed primal-dual algorithm to maximize aggregate utility. However, it is also known that duality have major drawbacks when the application is inelastic and utility functions are nonconcave, leading to divergence on congestion control.

Regarding wireless networks, game theory efforts have been focused on physical layer and MAC schemes in order to counteract channel dynamics. At the physical layer, many examples can be found in the literature proposing Nash Bargaining Solutions (NBS), such as spectrum sharing for ad hoc networks [82]), channel allocation [83] or power allocation in the satellite communications case [84].

In satellite networks, main efforts have been focused on bandwidth allocation. In [85], the authors propose a payoff utility function based on a Nash bargaining model to solve the bandwidth allocation problem for emergency services over satellite networks. They consider both QoS requirements and weather conditions, but based on giving more priority to those users that access the system first. In [86], a two level hierarchical bandwidth allocation is proposed at beam level and at user level. At user level, both non-cooperative (converging to NE) and a cooperative (whose solution is Nash Bargaining) frameworks are compared. [87] consider the allocation problem over broadband satellite network, they propose centralized multi-objective optimization to design an allocation mechanism that results in an equilibrium with desirable properties for both the service provider and the users. The authors propose a flexible Stackelberg game which provides and equilibrium between the proportionally fair allocation (fair in terms of channel conditions but bandwidth inefficient) and the opportunistic allocation (maximize throughput but no longer fair).

In our work, we focus on a specific aspect of resource allocation of satellite interactive networks. In satellite systems, STs request bandwidth according only to the traffic at the IP/MAC queues of the ST and future predictions, meaning that they are careless of the available resources at the satellite link. Such type of requests can saturate the DAMA scheme at the NCC in heavy loaded scenarios, unable to fairly allocate resources over all terminals (each in different channel conditions) requesting for heterogeneous traffic (with different QoS requirements). The addressed problem is to decentralize the STs requests, allowing each terminal to estimate suitable capacity requests (in terms of fairness and signaling savings) to be sent taking advantage of available NCC signaling. The distributed bandwidth request game case has not yet been addressed in satellite communications, where is particularly challenging compared to other wireless terrestrial networks (such as 3G, WiFi, etc.), given that the big RTT delay and the big amount of receivers make the network difficult to coordinate.

5.3.1.3 Contributions

In this chapter, we propose to partially decentralize the bandwidth allocation of the DAMA scheme on the return link of a multibeam broadband satellite air interface. In

our scenario, STs are able to estimate previous bandwidth allocations by means of signaling information available in the return link, in order to perform responsible bandwidth requests.

The contributions with preciseness are:

- We study the couple price function vs. network knowledge. Using different network knowledge information (capacity unaware, capacity aware and requests aware) in the pricing, we model each scenario as a non-cooperative game. We demonstrate that the first two scenarios converge to a Pareto inefficient NE. On the other hand, for the requests aware scenario, the STs are able to reach the optimal network operating point (Pareto Optimal solution), and thus performing efficient bandwidth-aware requests.
- We address also the α -fairness performance as a proportional pricing in the games proposed. We consider a tunable fairness, controlled by parameter α , which depends on the adaptive physical layer. We show that the solutions to the bandwidth requests problem are correlated to channel conditions of STs. This is particularly attractive since different policies can be chosen for different scenario: equal throughput, equal time sharing, to favour users in good or bad weather, etc. Moreover, our solution happens to be a NBS for specific values of α .
- We contribute with a partially distributed DAMA scheme, including protocol and cross-layer design (at physical, MAC and Internet Protocol layers), called Distributed Fairness Capacity Request (DistFair). We prove that the distributed non-cooperative request game is feasible thanks to the DistFair design. Moreover, we argue that DistFair is suitable for the standard specification of interactive satellite communications, i.e. DVB-RCS.
- Realistic temperate/tropical channel conditions at K_a band allow us to test the behaviour of our distributed design in terms of efficiency, fairness and signaling savings. We show that DistFair happens to be suboptimal (only by a maximum error of 0.1%) due to the delay (a RTT) of receiving the information of the available resources, during which channel and traffic conditions might change. However, the signaling overhead reduction with respect to a centralized solution, makes the return satellite capacity increase. We demonstrate also that STs adaptation rate to traffic dynamics is faster than in centralized solutions.

5.3.2 Interactive Satellite Network Model

5.3.2.1 Architecture, physical and link layers

We consider a GEO interactive broadband satellite network offering services to a big variety of terminals and covering large areas, from urban to rural and isolated areas. The broadband connectivity offered by the satellite, with greater bandwidth availability at the satellite, is suitable for current high traffic demands. We assume multibeam coverage, i.e. the satellite transmits to several beams to the earth, covering ideally non-overlapping geographical areas. This is particularly interesting since it allows increasing the capacity

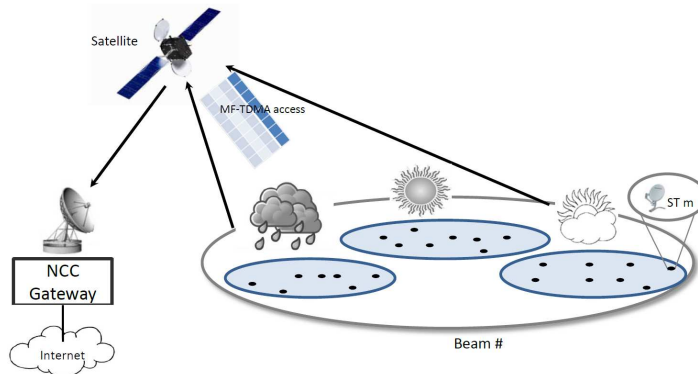


Figure 5.6: Star-like topology of and interactive multi-beam satellite network. Users are organized according to weather conditions

by frequency and polarization reuse over the full coverage. The total available power is distributed among the beams according to traffic and system requirements. For the sake of simplicity, we assume beam power is assumed pre-allocated and constant, and apart from differences in the link budget, all beams are conceptually identical. Next, we consider the air interface of one of the beam; this does not entail any loss of generality since the interference of other beams over the one under consideration is taken into account.

In our network model, we can assume a transparent or regenerative satellite [88]. The first one (depicted in Figure 5.6) is suitable for star-like architecture around gateways, where each gateway, connected to Internet core, supports the users for a given number of beams. Therefore, all transmissions from satellite terminals need to be processed by gateways. The latter case is designed as a response to cover the growing demand in multimedia broadband and real-time services, since its mesh-based topology is able to reduce in one satellite hop the connecting services delay. In this case, the satellite terminals are connected to each other via the regenerative satellite, where packets are processed on-board (OBP), and there is no terrestrial control center anymore.

For the return link design, we also assume K_a and therefore, we adopt ACM (or Adaptive Coding for the return link) at the PHY layer to counteract bad weather conditions, taking profit of the users diversity (see Figure 5.6). As in the forward link design (Section 5.2.3), we assume M different ACM modes, and the related features: spectral efficiency, transmission time and buffer length denoted by η_m , T_m and L_m respectively.

At the MAC layer, we focus on the return channel of an interactive satellite air interface. The access scheme implements MF-TDMA, and thus the return link is segmented into portions of time and frequency, which are organized in Super-Frames (SF), with each SF containing a number of frames, which are divided in time slots. The traffic bursts sent towards the time slots are carrying either ATM cells or MPEG-TS packets. Therefore, the main point is how to allocate the available time slots among the STs. Traditional satellite networks perform the allocation by means of a centralized entity or NCC, which collects the requests and necessary channel conditions from STs, and then, the entity broadcasts the time slot planning to all STs. As highlighted in the introduction, it is suitable in order to maximize throughput but entails some drawbacks such as the signaling and the RTT delay. In this chapter, we solve the bandwidth request / bandwidth allocation problem

using a partially decentralized network, which works as follows. Requests from STs are sent using signaling information attached to traffic ATM/MPEG bursts, the NCC collects the requests and the channel state information from each ST to perform the bandwidth allocation, and broadcasts the slot-schedule allocation and the physical layer features. At that point, a centralized network would begin again the cycle. However, in our distributed approach, STs are allowed to individually compute the bandwidth requests according to the signaling information broadcasted by the NCC (slot-schedule and physical layer). In such case, STs can do bandwidth-aware requests, given that they have knowledge of the network state. Since STs are capable to know previous bandwidth allocation, and thus they can approximately estimate the bandwidth allocation that will receive from NCC, three are the main advantages:

- Avoiding the excess in the number of capacity requests. Particularly in heavy loaded scenarios (such as in rain conditions), where signaling savings can increase the satellite capacity.
- Signaling information savings if we consider that channel or traffic is static with in a time window.
- STs are aware about the future allocations before receiving slot planning from NCC, which increases the adaptation rate gain, faster than centralized solutions.

The latter one is of big interest, since it means that, at the ST, delay or packet loss constrained application can adapt faster (proactive), than in a reactive case, where in order to be adapted, end-to-end parameters need to be measured and send back to the ST.

The signaling and adaptation rate studies will be aimed further, however the bandwidth request game is addressed next, where ST solve the capacity allocation in order to request for the optimal bandwidth.

5.3.2.2 Interactive traffic model

In satellite interactive network, opposite to Internet, it is imperative to guarantee various QoS levels and a minimum per user bandwidth. Thus, our problem is constrained to allocate at least the necessary bandwidth needed by Premium service applications, which require a constant rate during all the time, reserved for real-time and high constrained application such as VoIP and Videoconferencing. The way the rest of the available capacity is managed is the main issue faced in the chapter.

The available capacity is composed of, without loss of generality, two traffic types of differentiated quality. Silver services, which are suitable for less constrained traffic in terms of delay and jitter, such as video streaming and web traffic, but highly constrained in terms of packet loss. In contrast to Premium, Silver allows for statistical multiplexing among users. And finally, Bronze services traffic with non QoS or low QoS requirements, e.g. e-mail, web browsing, file transfers, etc. It is clear that we can assume a DiffServ QoS model, thus enabling easy integration with IP external networks.

Formally, in the system, there are N STs competing for the capacity of the satellite link (C_{sat}). Each ST i has a guaranteed minimum flow rate $x_{Premium,i}$ (for Premium

traffic), and a variable excess flow rate $x_{req,i}$ (for Silver and Bronze services), with a total request $x_{t,i} = x_{Premium,i} + x_{req,i}$. We consider the satellite capacity reserved for Premium traffic $C_{Premium} < C_{sat}$, and thus $C_{sat} = C_{Premium} + C$, assuming that the capacity drop will never reach $C_{Premium}$. Where C is the available capacity to be allocated according to our bandwidth request game presented in the following section.

5.3.3 Justification of the Game Formulation

The two main frameworks of game theoretical models for wired and wireless optimizations are cooperative and non-cooperative. However, cooperative games require additional signaling or agreements between the players, which is specially difficult in satellite communications due to large propagation delays, and thus not relevant works can be found in the literature. In non-cooperative games, players have potentially conflicting interests, and in order to maximize their payoff, players act according to their strategies as in [85, 86, 87, 89]. Particularly, in [86], authors compare both types of games, despite showing that cooperation achieves higher throughput, they do not solve the signaling among users. Games can be also classified in static or dynamic games. In static games, players make simultaneous moves without knowing what the other players do, as in [85] which uses Nash Bargaining problems, or in [87] using a leader and a follower (Stackelberg) game competing for bandwidth. On the other hand, in dynamic games, the players have sequential interaction, meaning that the move of one player is conditioned by the move of the others, this idea is exploited in [89], where authors model the bandwidth allocation problem as a dynamic buyer(satellite)-seller-terminal game.

We analyze here the bandwidth requests performed by users willing to compete for the available bandwidth that is allocated by the NCC of an interactive multibeam satellite network. In order to do so, it is obvious that STs cannot agree by exchanging signaling messages due to large propagation delays. Instead of that, in our scenario, STs are able to estimate the previous bandwidth allocation knowing the last time-slot planning (of one SF) and channel state of all STs, acquirable from the NCC. Given the allocation obtained, STs can perform responsible bandwidth requests to the NCC, in charge of next time-slot allocation. For the same reason, i.e. large delays, STs perform bandwidth requests per SF at the same time, since a sequential interaction is unfeasible. Therefore, we assume that the problem can be modeled as a static N-user (or N-ST) non-cooperative game.

5.3.4 Optimization: Distributed Bandwidth Request Game

5.3.4.1 Modeling the satellite selection game

First of all, let us denote the STs as the players, where a player i chooses a capacity request $x_{req,i}$ (in bits/s) as a feasible strategy in order to compete for C . Let S_i denote the feasible strategy set of player i , then $x_{req,i} \in S_i$. And the feasible strategy space of this game is $S = S_1 \times S_2 \dots \times S_N$. Then, a feasible strategy tuple is a N-dimension vector $\mathbf{x}_{req} = (x_{req,1}, x_{req,2}, \dots, x_{req,N})$. The objective of each player is to maximize its utility \tilde{U}_i which determines the aggressiveness (or willingness to request) of each player by adjusting

$x_{req,i}$. The resulting solution \mathbf{x}_{req}^* is defined $\forall i$ as:

$$x_{req,i}^* = \operatorname{argmax}_{x_{req,i} \in \mathcal{S}_i} \tilde{U}_i(x_{req,1}^*, \dots, x_{req,i}, \dots, x_{req,N}^*) \quad (5.15)$$

In our request game, let us consider an increasing, strictly concave and continuously differentiable utility function that has been widely proposed in the literature, when $x_{req,i} > 0$:

$$\tilde{U}_i(x_{req,i}) = \begin{cases} \log(x_{req,i}), & z = 1 \\ (1-z)^{-1} x_{req,i}^{1-z}, & otherwise. \end{cases} \quad (5.16)$$

In particular, we use the case $z = 1$ (logarithmic) [54], which maximizes the utility function and is proportional fair for all users.

In order to keep a relationship between the traffic model considered and our utility function, we introduce a QoS priority weight, termed $\Omega_i > 0$. We define $\Omega_i = \sum_{j=0}^2 q_j h_i(j)$, where q_j is the weight allocated to each type of traffic (Silver or Bronze), and $h_i(j)$ is the percentage that user i has of type of traffic j . Meaning that users requesting more Bronze than Silver traffic will be penalized in front of others that do not. As a result, our utility function can be expressed as $\tilde{U}_i(x_{req,i}) = \Omega_i \log(x_{req,i})$.

Modeling the STs with such utility functions makes them selfish and not socially responsible in the sense that they only care about their own benefit and are not aware of available resource. This is a consequence of the lack of network state knowledge. In our work, we present three illustrative N-user non-cooperative games, where the STs maximization problems are of the form $U_i = \tilde{U}_i(x_{req,i}) - P_i(x_{req,i})$, in order to investigate the impact of the network bandwidth information in the pricing P_i and the constraints, for three different scenarios:

1. Network capacity unaware.
2. Use of limited information: Network capacity aware.
3. Capacity requests (of other STs) aware. The bandwidth requests are decided on static optimizations each time the ST receives the time-slots allocation.

We address the trade-off between throughput efficiency and fairness for each scenario.

5.3.4.2 N-user non-cooperative game 1: Capacity unaware

The first game addressed here in is a non-cooperative game, where STs do not have any knowledge about the satellite capacity and the requests performed by other users. In order to model the game, we assume a typical pricing function easily found in the literature, which defines the cost of accessing the network for user i as $p_i x_{req,i}$, bounding the price of anarchy. In addition, we constrain the capacity allocated per user as follows; each user request $x_{req,i}$ will request a minimum flow rate of $x_{min,i}$, and a maximum $x_{max,i}$. Finally, we denote the game as the following distributed maximization problem:

$$\begin{aligned} \max_{x_{req,i}} \quad & \Omega_i \log(x_{req,i}) - p_i x_{req,i} \\ \text{s.t.} \quad & x_{min,i} \leq x_{req,i} \leq x_{max,i} \end{aligned} \quad (5.17)$$

Opposite to previous works based on TCP [80, 90], where p_i has been modeled according to packet loss, buffer size, etc, herein, we address a fair pricing. In particular, it is the pricing used for the forward link (5.2), correlated to the channel diversity, with a slight modification to include the number of STs. First, let us assume that the request policy of the NCC is a Weighted Round Robin (WRR). The expression of the throughput requests S_m to be allocated from STs using ACM mode m is:

$$S_m = u_m \frac{w_m L_m}{\sum_{j=1}^M w_j T_j} = u_m \frac{w_m L_m}{\sum_{j=1}^M w_j L_j / \mu_j} \quad (5.18)$$

where $W = (w_1, w_2, \dots, w_M)$ are the weights (i.e. priority) assigned to each physical layer (or ACM mode) with $\sum_{m=1}^M w_m = 1$, and u_m are the number of STs using ACM mode m . The features of each ACM mode, i.e. η_m , T_m and L_m , are the spectral efficiency, transmission time and buffer length respectively. And the weight of a user i using the ACM mode m is expressed as follows:

$$w_i(\alpha, m) = \frac{\frac{1}{T_m \eta_m^\alpha}}{\sum_{j=1}^M \frac{1}{T_j \eta_j^\alpha}} \quad (5.19)$$

In (5.19), more weight means more priority in the WRR, but for pricing, more price means more restriction. According to this, we propose the pricing $p_i = \beta/w_i$, where $\beta > 0$ is a scaling factor.

The value that solves (5.17) is:

$$x_{req,i}^* = \left[\frac{\Omega_i}{\beta/w_i} \right]_{x_{min,i}}^{x_{max,i}} \quad (5.20)$$

where $[y]_b^a$ means that $y \in [a, b]$, and if $y > a$ or $y < b$ then $[y]_b^a = a$ or $[y]_b^a = b$ respectively. It is straightforward that, due to the lack of traffic requests and capacity information (from other STs) the solution of the (5.20) problem tends to a Pareto inefficient solution, which might not converge depending on p_i . Although the pricing introduced can correlate requests with channel conditions, STs are totally blind regarding the current available capacity, and thus they perform inefficient requests. To overcome this inefficiency, we propose a pricing that includes a regulatory function using network information.

5.3.4.3 N-user non-cooperative game 2: capacity aware

In the second model, we assume that STs are aware of the total satellite link capacity C to be shared among all users. Nevertheless, they don't have information about the traffic requests and channel status (i.e. physical layer) of the rest of STs. Taking advantage of C , we use a regulatory function $C - (x_{req,i} + x_{a,-i})$ addressed in [91] in order to bound the price of anarchy in our scenario. Let us define $x_{a,-i}$ as the allocated bandwidth to all other STs, i.e. $\sum_{j \neq i} x_{a,j}$. We adapt the function to our scenario as:

$$\begin{aligned} \max_{x_{req,i}} \quad & \Omega_i \log(x_{req,i}) - \frac{\beta(1/w_i)x_{req,i}}{C - (x_{req,i} + x_{a,-i})} \\ \text{s.t.} \quad & x_{min,i} \leq x_{req,i} \leq x_{max,i} \end{aligned} \quad (5.21)$$

Note that we still consider the correlation with the adaptive physical layer (i.e. weight value w_i) in (5.19). In (5.21), via the denominator term, the network state works regulates the requests. As the sum of users request approach the available capacity C , the denominator approaches to zero, and hence the price increases without bound. This preserves the network resources by forcing the users to decrease their requests.

The N-user non-cooperative distributed game defined in (5.21) has the property to reach a Nash Equilibrium (NE) point when the joint strategy space S is convex and compact, and the objective function that each player wants to maximize is concave ($\partial^2 U_i / \partial x_{req,i}^2 < 0$). Both conditions are fulfilled in (5.21).

Definition 1: At NE, none of the users can unilaterally change his strategy to increase his payoff. The NE point $\tilde{\mathbf{x}}_{req}$ can be formally defined as:

$$U_i(\tilde{x}_{req,i}, \tilde{x}_{req,-i}) \geq U_i(x_{req,i}, \tilde{x}_{req,-i}) \quad \forall x_{req,i} \in S_i \quad (5.22)$$

where $\tilde{x}_{req,-i}$ is the vector of other players strategies. We compute the equilibrium point explicitly by applying Lagrange multipliers, which leads to the solution:

$$x_{req,i}^* = [(C - x_{a,-i})\Phi_i(\Omega_i, w_i)]_{x_{min,i}}^{x_{max,i}} \quad (5.23)$$

where $\Phi_i = 1 + ((\beta/w_i) - \sqrt{(\beta/w_i)^2 + 4\Omega_i(\beta/w_i)})/2\Omega_i$. Note that the solution is proportional to the available capacity $(C - x_{a,-i})$ term. However, since $w_i > 0$, then $\Phi_i < 1$, which means that the capacity requested by user i never reaches the available bandwidth, i.e. $x_{req,i}^* < (C - x_{a,-i})$. As a result, the equilibrium point of (5.21) happens to be Pareto inefficient for this scenario. Therefore, although this scenario is aware of the total satellite link, and performs better than the (5.17) in terms of bandwidth efficiency (as we will see later), it is not capable to reach network optimality. Convergence and uniqueness of solutions of the type (5.23) are proved in [91]. It is straightforward to see that in order to achieve a Pareto improvement, STs must know the real bandwidth requests and the available capacity. Such problem is addressed next.

5.3.4.4 N-user non-cooperative game 3: Capacity requests aware

In this model, we assume that all STs are aware of the previous requested bandwidth of the overall network and the transmission efficiency (or ACM) used by each ST. This assumption is realistic for current and future satellite systems, where the NCC is able to periodically broadcast the time slots allocation and the channel dynamics (in terms of ACM mode used) per ST.

Taking the problem (5.17) as a starting point, which has been shown to be Pareto inefficient, our purpose is to look how this problem can be easily derived to an efficient request allocation. In particular, in [81], the authors show how the prices p_i of a problem like (5.17) are determined such that the corresponding rate allocations lead to a centralized NBS rate allocation. In particular, the authors state that considering a convex Global Optimization Problem (GOP) (5.24), if \mathbf{x}_{req} is the unique NBS of the GOP and μ denotes the implied cost associated with satellite link obtained from the solution of GOP, i.e. it is

the Lagrange multiplier associated with the constraint $\sum_{i=1}^N x_{req,i} \leq C$. Then, if $p_i = \mu$, the solution to the (5.17) is not only Pareto Optimal (PO) but also a NBS.

$$\begin{aligned} \max_{x_{req,i}} \quad & \sum_{i=1}^N \Omega_i \log(x_{req,i}) \\ \text{s.t.} \quad & \sum_{i=1}^N x_{req,i} \leq C \\ & x_{min,i} \leq x_{req,i} \leq x_{max,i} \end{aligned} \quad (5.24)$$

In order to clarify the above concepts, the Pareto optimality concept is defined below. *Definition 2:* The strategy profile \mathbf{x}_{req} is PO, if for any other strategy \mathbf{x}'_{req} :

$$U_i(\mathbf{x}_{req}) \geq U_i(\mathbf{x}'_{req}) \quad \forall i \quad (5.25)$$

In other words, there exists no other strategy that leads to strictly superior performance for all players simultaneously. From this definition, it is clear that an optimal network operating point should be a PO point. However, since a PO point might be not unique, a way to select the suitable operating point might take into account the fairness criteria. The NBS solution is able to include both requirements: being PO and by satisfying the axioms of fairness [92, 93]. Let us denote the NBS request game as $Dist_{NBS}$. However, this concept of fairness is defined from the network's point of view, since it maximizes its total revenue based upon charging μ units of bandwidth to ST i , whose utility is also maximized.

In this chapter, thanks to (5.19) and opposite to [81], the fairness concept is correlated to the Channel State Information (CSI) of STs, also included in the NCC signaling. In other words, instead of assuming a constant price $p_i = \mu$, each ST has associated a tunable price $p_i = \beta/w_i$. The full problem, solved in a decentralized way by each ST, allow us to compute the optimal capacity to be requested:

$$\begin{aligned} \max_{x_{req,i}} \quad & \sum_{i=1}^N \Omega_i \log(x_{req,i}) - \beta(1/w_i)x_{req,i} \\ \text{s.t.} \quad & \sum_{i=1}^N x_{req,i} \leq C \\ & x_{min,i} \leq x_{req,i} \leq x_{max,i} \end{aligned} \quad (5.26)$$

Thus, the pricing adapts to the weather conditions experimented by the whole satellite network. By applying Lagrange, we note that the solution $x_{req,i}$ that solves this bandwidth request game problem is optimal when Karush Khun Tucker (KKT) conditions [66] are satisfied leading to:

$$x_{req,i} = \begin{cases} x_{max,i}, & \lambda < \frac{\Omega_i}{x_{max,i}} - \frac{\beta}{w_i} \\ \frac{\Omega_i}{\lambda + (\beta/w_i)}, & \frac{\Omega_i}{x_{min,i}} - \frac{\beta}{w_i} \geq \lambda \geq \frac{\Omega_i}{x_{max,i}} - \frac{\beta}{w_i} \\ x_{min,i}, & \lambda > \frac{\Omega_i}{x_{min,i}} - \frac{\beta}{w_i}. \end{cases} \quad (5.27)$$

This solution method receives the name of *water filling*. The problem can be graphically seen as N different containers (one per ST) during one Super-Frame. Each container is

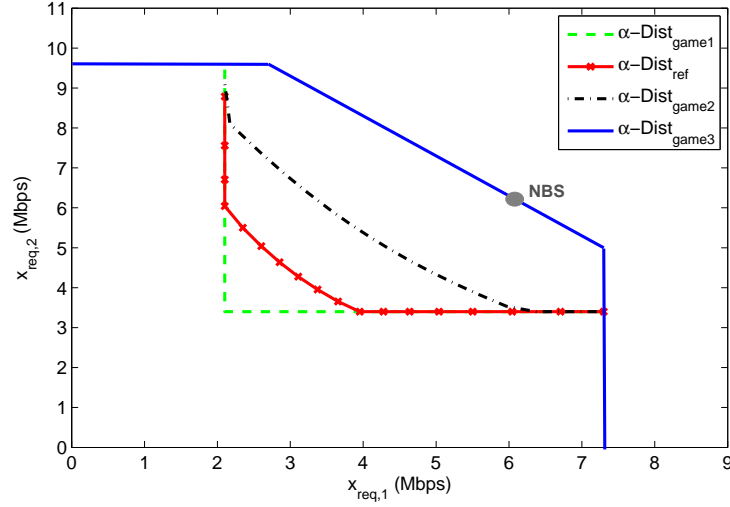


Figure 5.7: Set of feasible solutions (in terms of Mbps) for different Dist Optimization problems

filled with a volume of water representing the capacity request, which is equal to $\Omega_i \times \frac{1}{\lambda + (\beta/w_i)} \times 1$. For the sake of clarity, since from (5.19) $w_i \propto 1/\eta_i^\alpha$, then the water level, or in other words, the bandwidth request of user i is:

$$x_{req,i} \propto \frac{1}{\lambda + (\beta/w_i)} \propto \frac{1}{\lambda + (\beta\eta_i^\alpha)} \propto \frac{1}{\eta_i^\alpha} \quad (5.28)$$

Meaning that, for example, in case $\alpha = 1$, more priority to request bandwidth is given to those STs transmitting with low efficient ACM modes. Such example demonstrates that bandwidth request performed by ST i is correlated with its channel conditions.

The Pareto efficiency of the scenarios addressed herein can be observed in Figure 5.7, where two users are competing for the network resources in terms of capacity requests (Mbps), we can observe how our solution (5.27) called $\alpha\text{-Dist}_{game3}$ is PO, since it reaches the threshold of the network operating area, for different values of w_i (i.e. different values of α). Moreover, this solution not only allows to define different policies according to channel conditions (i.e. weight allocation), but also includes the NBS for the particular value when $\beta/w_i = \mu$. The rest of the games depicted are denoted as $\alpha\text{-Dist}_{ref}$, $\alpha\text{-Dist}_{game2}$ and $\alpha\text{-Dist}_{game1}$ for the requests solutions in [91], (5.21) and (5.17) respectively. Although all these solutions lead to different NE boundaries, in Figure 5.7, it is demonstrated that they are Pareto inefficient.

5.3.5 Cross-Layer Architecture Design: DistFair

The peculiarities of our design are presented in this section, including the cross-layer architecture, the detailed cross-layer information exchanged among layers and the intra-network signaling between the STs and the NCC, in order to perform the Distributed Fairness Capacity Request (DistFair) solved by (5.26).

We focus on the particular case of a DVB-RCS air interface, which has been designed as the return link of the widely accepted second satellite generation standard (DVB-S2) [18] for broadband satellite communications.

As we already described, the objective of the DistFair cross-layer design, implemented in the STs, is to adapt the bandwidth requests performed by STs according to traffic and channel dynamics, specially when capacity drops drastically. Given the optimal bandwidth requests solution addressed in the non-cooperative game (5.26), the roles of the DistFair design are three:

- To provide the needed knowledge network state to each ST: previous bandwidth allocation and ACM mode used per ST
- To make sure that the resulting bandwidth requests are correctly transmitted from the ST to the NCC.
- Inform upper layers about the bandwidth available at the MAC layer.

The first two roles are provided through a Physical-to-Medium Access Control (PHY-MAC) layers signaling flow between the NCC and the ST, termed intra-network cross-layer. The latter one is carried out thanks to a MAC-to-Internet Protocol (IP) cross-layer flow at the ST, which follows an explicit methodology due to the fact that the exchange of information between layers and the adaptation is carried out online.

In order to detail the cross-layer flows, let us first describe the frame description and the capacity requests types of DVB-RCS.

5.3.5.1 DVB-RCS framing description

In a MF-TDMA scheme, as the one presented in Section 5.3.2.1, a frame is composed of traffic burst (TRF), ACQuisition (ACQ), SYNChronization (SYNC) and Common Signaling Channel (CSC). Considering the examples provided in the guidelines of the standard [22], we assume fixed frame duration is 26.5 ms, with 23 carriers per frame. Each carrier is divided in 26 time-slots, 24 of which are TRF slots (with 1 ATM cell per TRF), and 2 remaining slots for transmitting 4 SYNC.

The SYNC contains the Satellite Access Control (SAC) field composed of signaling information added by the ST for requesting capacity purposes. The SatLabs recommendations [94] recommends a SAC length of 14 bytes (when used in the SYNC burst with Turbo Code encoding), 7 for identifications and other MAC information, and the 7 remaining are used for bandwidth requests. Since each request requires 2 bytes (specifying type of requests, scaling factor and up to 8Mbits/s requests), each SAC might contain up to 3 different requests from the ST (e.g. one for each type of traffic). SYNC slots are assigned with a 32 frames period. Therefore, each terminal transmits a SYNC burst every 848 ms.

5.3.5.2 DVB-RCS capacity request mapping

Let us also propose the following mapping between the different types of capacity requests of the DVB-RCS DAMA scheme and the types defined in Section 5.3.2.2. Taking into account the characteristics of each traffic type, such as delay/jitter tolerance, packet loss tolerance and other QoS requirements, we assume that Constant Rate Assignment (CRA), Rate-Based Dynamic Capacity (RBDC) and Volume-Based Dynamic Capacity (VBDC) are Premium, Silver and Bronze respectively.

5.3.5.3 Intra Network Cross-layer: *XL-net*

The signaling considered in the intra-network cross-layer has been implemented in the Dynamic Bandwidth Allocation (DBA) scheme of the DVB-RCS air interface, where the NCC periodically broadcasts two signaling frames:

- The Terminal Burst Time Plan (TBTP), which updates the timeslot allocation within a SF between every competing ST. DBA allocates transmission bandwidth according to the time-variant bandwidth requests (SYNC signaling) coming from the STs. Traffic capacity is assigned on a frame basis, meaning that TBTP is distributed every frame.
- The Timeslot Composition Table (TCT) message with physical layer features of all STs: Symbol rate, modulation and coding, roll-off and time duration. It must be broadcasted at least every 10 seconds to reflect system status changes.

According to the TBTP, STs can obtain, in a distributed way, an estimation of instantaneous available capacity in order to request bandwidth in a responsible way. Moreover, thanks to TCT, each ST can not only know the STs affected by a rain event (which are going to increase bandwidth requests), but also is able to compute the fair pricing (5.19) to be used in the game problem (5.26).

Moreover, as it is shown in (5.26), STs need to know the request type of others in order to define the QoS parameter Ω_i . However, no specific signaling is defined in the DVB-RCS standard for this purpose. Assuming that the CRA capacity is initially negotiated, we only need 1 bit to specify the type of request (RBDC or VBDC). We have identified two different fields of the TBTP table (see [21]) that might be suitable to specify the type of request assigned to each time-slot or group of time-slots:

- *Assignment_Type*: In particular, the "Assign type for MF-TDMA time slots" option, where the "one time assignment" might be matched to the VBDC traffic type and the "repeating assignment" might be matched to RBDC traffic type.
- To use one of the bits of the *Logon_ID* field.

5.3.5.4 MAC to IP Explicit Cross-layer: *XL_MAC-IP*

Our cross-layer architecture is based on standardized tools implemented in the SI-SAP [9].

Let us define the primitives addressed in the chapter, i.e. those included in the Resource Reservation functions of the Control C-plane:

- *SI-C-QUEUE_OPEN-xxx*: IP queue manager request to open a new QID.
- *SI-C-QUEUE_MODIFY-xxx*: IP queue manager request to modify an existing QID.
- *SI-C-QUEUE_CLOSE-xxx*: IP queue manager request to close an existing QID.
- *SI-C-QUEUE_STATUS-xxx*: Exchange of queue status information between the IP queue manager and the lower layers.

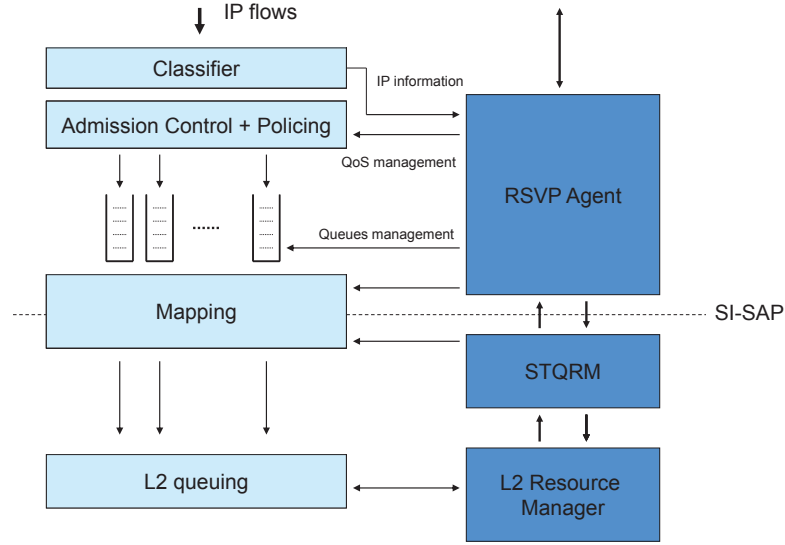


Figure 5.8: BSM Architecture of MAC to IP Explicit Cross-layer scheme

The responsibility of interaction is of the RSVP Agent, the STQRM and the L2 Resource Manager (depicted in Figure 5.8). The latter is responsible of performing the algorithm according to the information received from *XL_{net}* (i.e. TCT and TBTP). According to the results obtained at the L2 Resource Manager of the MAC layer, the information is transmitted to the STQRM, which interacts with the RSVP Agent using the SI-SAP primitives. This means that our DistFair approach can be essentially seen as a MAC-centric approach.

Next, we present the primitives flows to be used in our DistFair for different scenarios:

- *New IP flow request from ST terminal*: from RSVP to STQRM requesting for new QID, needed to open a new IP flow. The STQRM decides (considering the available capacity computed by the L2 Manager) if there are available resources and if it can increase bandwidth requests to the NCC, and confirm (or not) the availability to the RSVP by using the *SI-C-QUEUE_OPEN-cfm*.
- *Decrease of available capacity (due to rain event)*: First of all, L2 Manager estimates the available resources. If the ST needs to reduce the capacity requests to the NCC, then its own data rate should be decreased, the STQRM sends an *SI-C-QUEUE_MODIFY-xxx* from STQRM to RSVP ordering to change QID specifications (data rate, token bucket rate, etc).
- *Critical decrease of available capacity (due to rain event)*: Related to the last flow, however, if the reduction of data transmission or QoS (known thanks to the *SI-C-QUEUE_STATUS* information) does not fill IP flows requirements, the STQRM can decide to send *SI-C-QUEUE_CLOSE-ind* to RSVP ordering to close a QID, and therefore an IP flow.

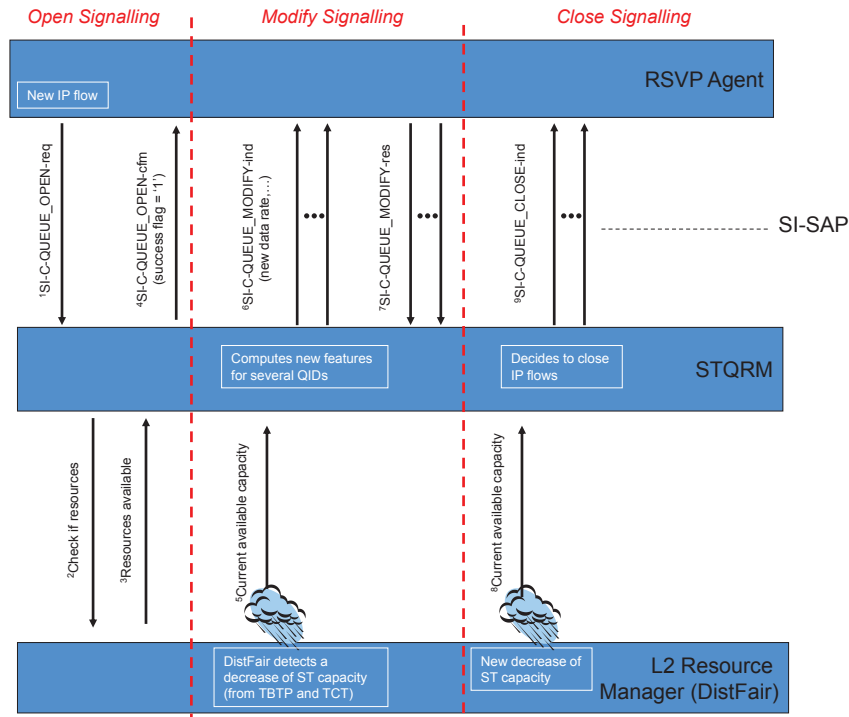


Figure 5.9: SI-SAP primitives signaling flows performing in different scenarios

Figure 5.9 shows the SI-SAP flows used for each of the scenarios presented above, by using the open, modify and close signaling flows. Note that the flows are classified in time from left (1) to right (9), for the first one and the last one respectively.

5.3.6 Results

In this section, we aim at showing the performance of our DistFair cross-layer algorithm, including our solved requests problem (5.26) denoted by α -Dist_{game}³. Realistic channel conditions at K_a Band are used in order to get good estimations of the capacity requests performed by STs. The study includes diversity in terms of channel statistics for N STs, each one experimenting different weather conditions: from deep to low rain attenuation. In particular, we assume 60 % of the STs are under rain conditions and the rest are under clear sky conditions.

Four are the main topics addressed in these results: requests-channel correlation, requests efficiency, fairness and signaling savings.

5.3.6.1 Bandwidth requests: channel correlation

The simulations are performed with N=100 STs during 40 minutes (2400 sec.). We assume that each ST requests the same amount of traffic under clear sky conditions in order to better observe the correlation between requests and channel. Let us first show an example of selfish-centralized traffic requests (i.e. not a distributed DistFair) performed by users, see Figure 5.10. Note that users under rain conditions need to request for more bandwidth

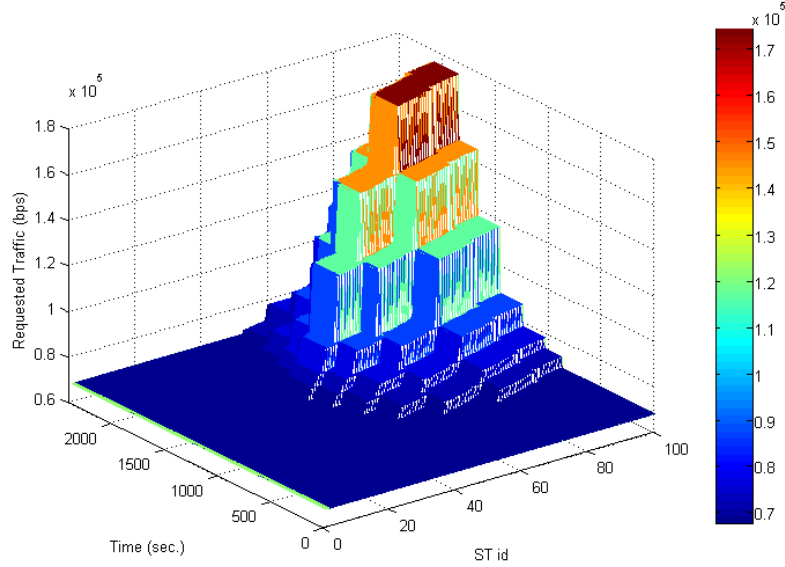


Figure 5.10: Centralized (selfish) traffic requests per user during a rain event affecting 60 % of users

when are affected by rain events in order to protect more (reducing the spectral efficiency) the traffic to be transmitted.

Two examples of traffic requests performed by STs using DistFair cross-layer design can be seen in Figure 5.11 and Figure 5.12, where each example follows a different policy in terms of α . The first case favours STs under rain conditions ($\alpha = 1$) since higher pricing is allocated to those STs under clearsky conditions. In the latter case, where $\alpha = 0$, only STs under rain conditions are affected and are not allowed to drastically increase the traffic requests, in other words, STs under clearsky conditions achieve isolation with respect to the capacity drop.

The evolution of the overall aggregated traffic demands of STs (in a centralized system) and the traffic demands of our DistFair requests solution (when $\alpha = 0$) are compared in Figure 5.13. Note that our solution is able to adapt the capacity requests during a 200 sec. rain event. The main improvements with respect to other solutions are that: 1) It avoids over-requesting capacity, e.g. unnecessary bandwidth requests over the current traffic requests can cause unnecessary congestion situations. 2) Channel adaptation, i.e. under heavy fading, "blind" methods, such as over-requesting 10%, are not able to adapt to the capacity drops.

The different behaviour of requests performed by users under rain conditions and clearsky conditions can be also observed in Figure 5.13. Since $\alpha = 0$ is used, only the requests of rain users are affected, while requests from clearsky users are maintained. Such performance is also observed in Figure 5.14, where we compare the following policies: DistFair ($\alpha = 0$), a proportional fair (where the request of terminal i can be denoted by $x_{req,i} * C / \sum_{i=1}^N x_{req,i}$) and a policy where rain users have guaranteed requests. The comparison is studied in terms of the ratio between aggregated demands and capacity requests performed. The scenarios considered are when rain users are using QPSK mod-

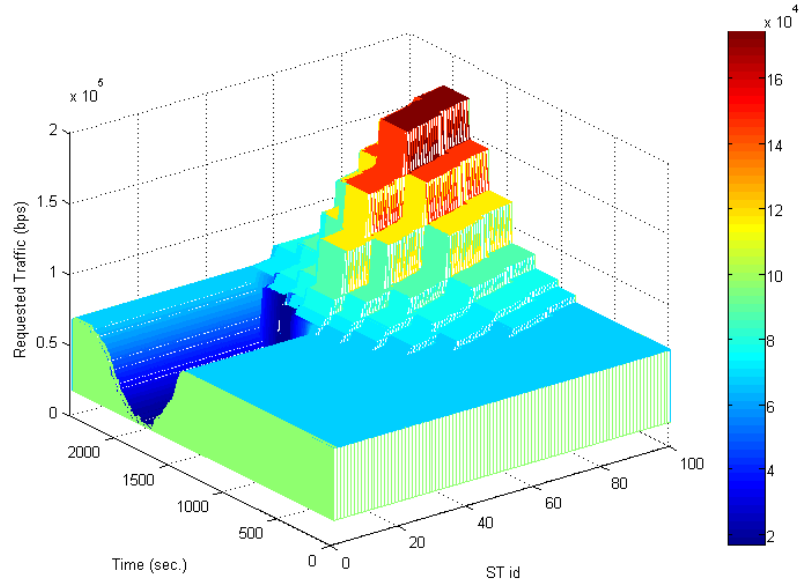


Figure 5.11: Traffic request per user during a rain event affecting 60 % of users, by using DistFair with $\alpha = 1$

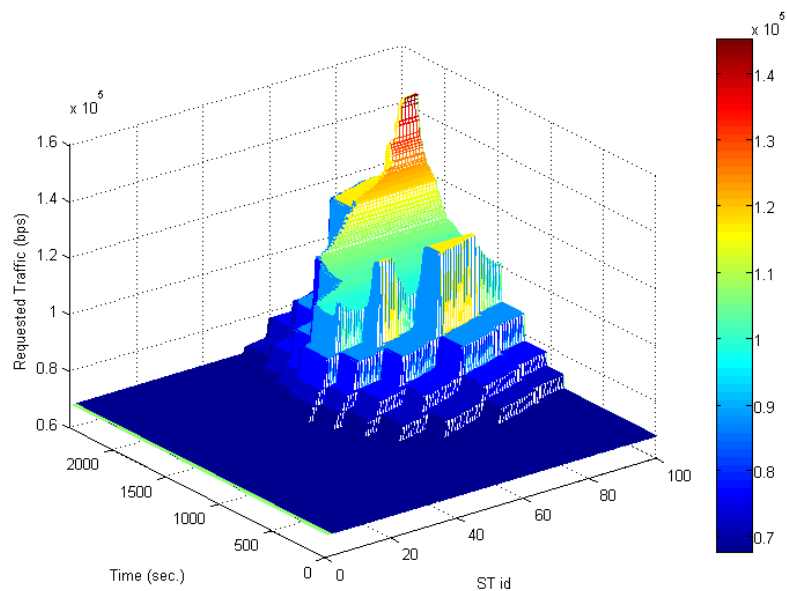


Figure 5.12: Traffic request per user during a rain event affecting 60 % of users, by using DistFair with $\alpha = 0$

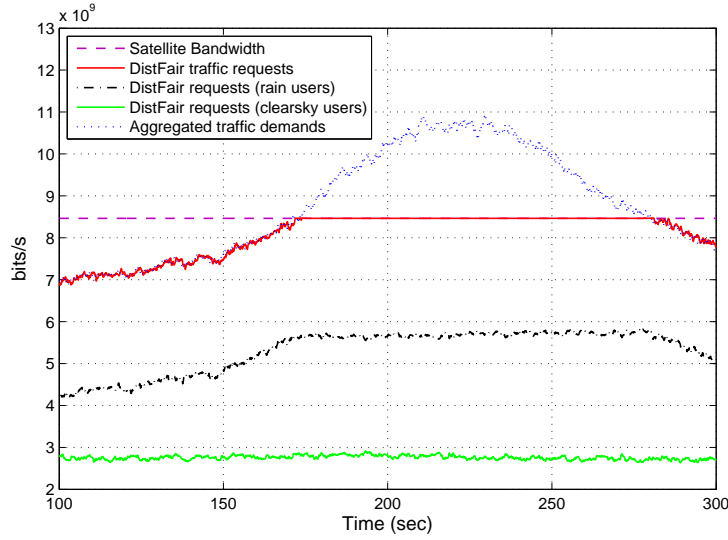


Figure 5.13: Aggregated bandwidth demands and DistFair capacity requests during a rain event

ulation with different adaptive coding: $6/7$, $4/5$, $3/4$, $2/3$, $1/2$, $2/5$ and $1/3$ are scenarios 1 to 7 respectively, i.e. sorted from low to deep rain attenuation. As it is expected, DistFair ($\alpha = 0$) allows clearsky users to be isolated from the rain event.

5.3.6.2 Efficiency and Fairness

The efficiency of the requests performed is an important quality measure for our decentralized requests. Not only compared to other distributed solutions, but also regarding the impact of the delayed (one RTT) network information (previous bandwidth allocation and ACM modes) used to perform the distributed requests.

In order to do so, we assume a diversity of $N=1000$ STs, where 60% of them are under rain conditions. We study the traffic requests performed during 24 minutes channel time-series for different values of α fairness (policy pricing). The three distributed games presented are evaluated: α - $Dist_{game1}$, α - $Dist_{game2}$ and α - $Dist_{game3}$. Moreover, they are compared with the distributed NBS solution (denoted as $Dist_{NBS}$), and the efficiency is normalized with respect to the real bandwidth availability at the moment the STs perform the requests.

As it is depicted in Figure 5.15, since $Dist_{NBS}$ and α - $Dist_{game3}$ are Pareto optimal solution, the two solutions maximize overall bandwidth allocation. Although it is barely noticeable, both solutions are suboptimal compared to the real available bandwidth due to the delayed network information. However, the efficiency error is only around 0.1%, which might be removed by using predictive traffic algorithms.

Regarding α - $Dist_{game1}$ and α - $Dist_{game2}$, both are shown to be Pareto inefficient. The first one, is very inefficient for negative values of α , and for positive values, it overtakes the maximum bandwidth efficiency since in this case, STs are not aware of the satellite link capacity. The latter one only reaches maximum efficiency around $\alpha = 2$.

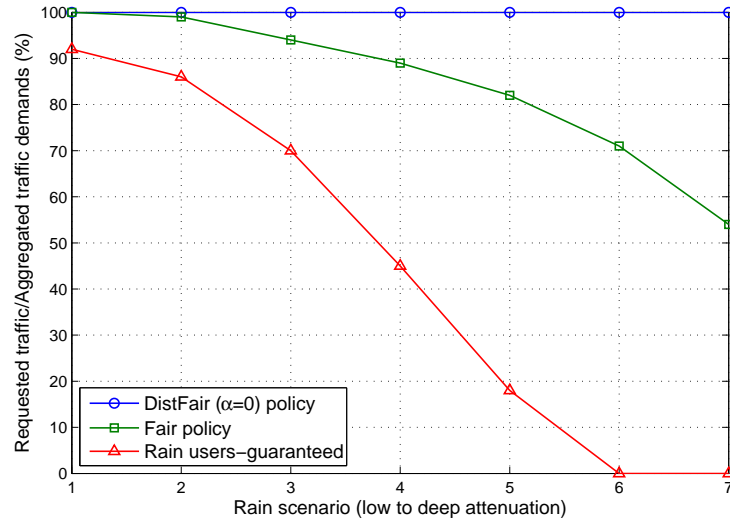


Figure 5.14: Ratio requested traffic vs. aggregated traffic demands in different rain scenarios

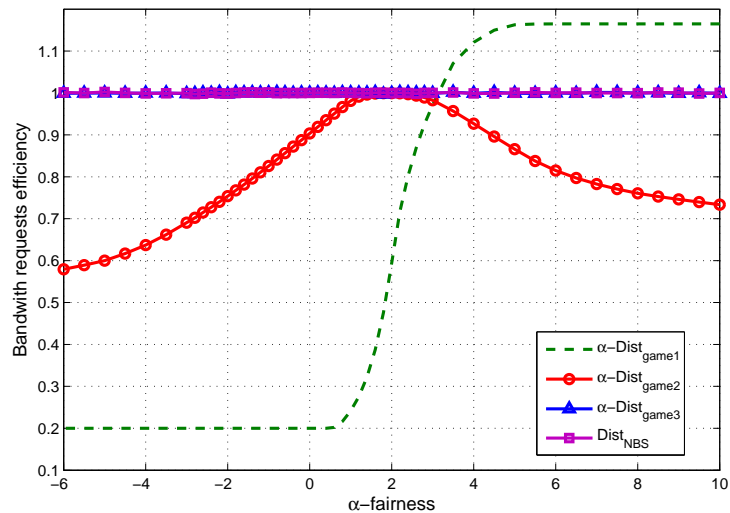


Figure 5.15: Bandwidth requests efficiency depending on pricing α for different distributed games, when satellite link is affected by rain

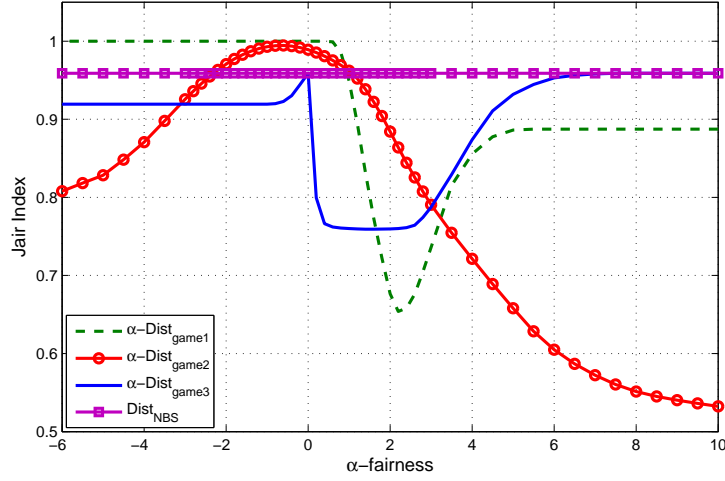


Figure 5.16: Jain’s fairness index depending on pricing α for different distributed games, when rain is dropping the satellite link capacity

A similar comparison is performed in order to observe fairness of DistFair. Although a measure of fairness is very subjective, several options can be found in the literature. In our case, we assume the concept of Jain’s index to observe the fairness of each request solution:

$$I_{fairness}(y_1, y_2, \dots, y_N) = \frac{(\sum_{i=1}^N y_i)^2}{N \sum_{i=1}^N y_i^2} \quad (5.29)$$

where $y_i = x_{req,i}/\eta_i$. The result ranges from $1/N$ if the system is completely unfair (worst case) to 1 if the system is completely fair (best case), and it is maximum when all users perform the same requests. Results are shown in Figure 5.16 for different pricing. As expected, the fair request by definition is the one solved by $Dist_{NBS}$, which does not depend on α . However, our $\alpha-Dist_{game3}$ solution allows to serve different fairness depending on α . When $\alpha \leq 0$, STs under good weather conditions are isolated from rain effects, when $\alpha \in [0, 6]$ these STs are penalized when capacity drops due to rain users. We must emphasize also the case when $\alpha \rightarrow \infty$, which makes the pricing (from (5.19)) $w_i \rightarrow 0 \forall i$, meaning that the requests of $\alpha-Dist_{game3}$ when $\alpha \rightarrow \infty$ are equal to the NBS of $Dist_{NBS}$. The above performances can be also observed in Figure 5.17, where the traffic requested by STs under clear sky conditions is shown.

Note also that the fairness of $\alpha-Dist_{game1}$ and $\alpha-Dist_{game2}$ are unfair and difficult to control.

5.3.6.3 DVB-RCS signaling savings

One of the main goals of our DistFair solution is the reduction of return link signaling, in particular the SYNC signaling. These savings can be divided in three *saving solutions*:

- a) To reduce SYNC framing.
- b) To reduce bandwidth request rate or increase SF size.

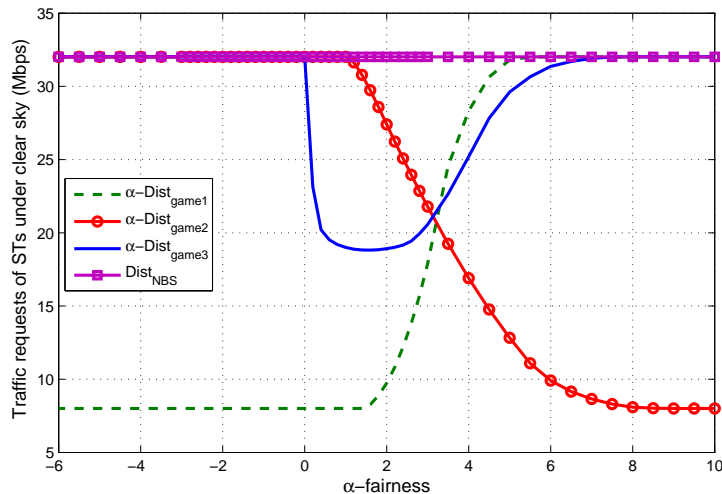


Figure 5.17: Bandwidth requested by users under clear sky conditions depending on pricing α and for different distributed games, when satellite link is affected by rain

c) The saturation threshold in heavy fading scenarios.

For this study, we assume $N=1000$ STs, and in order to simulate dynamic traffic requests, the traffic distribution to be transmitted by each ST describes the interarrival times of a Poisson process.

In the first case, to reduce SYNC framing (*a*), the DistFair solution does not need any more to reserve 2 time-slots among the 26 available per frame per carrier for SYNC signaling during all the SF (as proposed in the standard and explained in Section 5.3.5.1). Since STs perform the capacity requests when the TBTP is received (only once per SF), then only N SYNC slots per SF are enough.

Decreasing the bandwidth request rate (*b*) also allows to reduce signaling overhead in the DistFair. Given the fact that the bandwidth request game can be solved for any amount of available bandwidth, STs can compute the bandwidth to be requested during 1, 2, ... or K SFs. Meaning that STs only need to send the capacity requests once every 1, 2, ... or K SFs, and we can decrease the bandwidth request (i.e. SYNC) rate. Similarly, it is also possible to increase the SF size (i.e. length). The drawback of such solution is that the bandwidth allocation estimated at the ST will be less accurate when the bandwidth request period increases.

In heavy fading condition (*c*), we can demonstrate that many STs might work at a requests saturation threshold as shown in Figure 5.18, termed x_{thr} . Since every ST and the NCC are able to compute x_{thr} , terminals can avoid using unnecessary SAC transmissions in SYNC bursts. Note that this is true in the particular case of VBDC traffic. For RBDC, in order to prevent a terminal anomaly resulting in a hanging capacity assignment, the last RBDC request received by the NCC from a given terminal shall expire after a timeout. The time out can be configured between 1 and 15 SF lengths. Meaning that there are still signaling savings, but less than in VBDC.

Thus, we are considering that the STs in the x_{thr} threshold do not send their SYNC burst. This assumption allows reducing overhead slots in heavy conditions, which can be

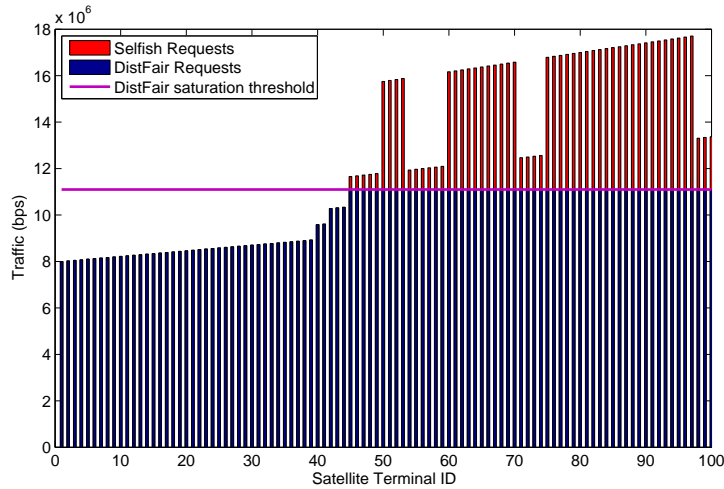


Figure 5.18: Comparison between Selfish requests in a centralized network and the DistFair requests

assigned to TRF bursts.

The results of the three *saving solutions* are summarized in Figure 5.19. Note that the behaviour of the three savings are cumulative, meaning that, for the heavy fading scenario, all the signaling savings can be applied (*a*, *b*, *c*), allowing to reduce the signaling up to 90%. While in low fading only *a* and *b* solutions are performing. In Figure 5.19, the results are expressed in terms of bandwidth gain (a better choice to demonstrate the design strengths). The reduced SYNC framing (*a*) allow us to reach up to 5.5% of bandwidth gain (case of low fading when 1 SYNC per ST per SF is sent). In low fading scenarios, the gain can reach 8% when 1 SYNC per ST is sent every 8 SFs thanks to *a* and *b* savings. However, by substantially reducing the SYNC rate, the accuracy of the requests might decrease and the bandwidth efficiency might be reduced, and thus a trade-off must be considered. Finally, in heavy fading scenarios (*c*), when the bandwidth gain is vital, DistFair can increase the bandwidth up to 8% with a 1 SYNC per 3 SFs rate.

5.4 Conclusion

5.4.1 Conclusions on the methodology

In a broadband adaptive satellite-based scenario such as the one considered, where resources are usually enough to efficiently distribute bandwidth demands, the main challenge is the capacity decrease due to weather conditions. Such challenge is the main impact on the system design, and thus a bottom-up approach, where PHY-layer informs upper-layers about the ACM mode to be used is the most suitable way to proceed. In particular, in our designs, PHY layer informs MAC layer about the used ACM mode per ST.

Although a simple MAC to PHY flow would be efficient and would not entail high cross-layer interaction complexity, it does not take into account high layer system parameters. Such design is not optimal for multimedia applications, due to the incurred delays

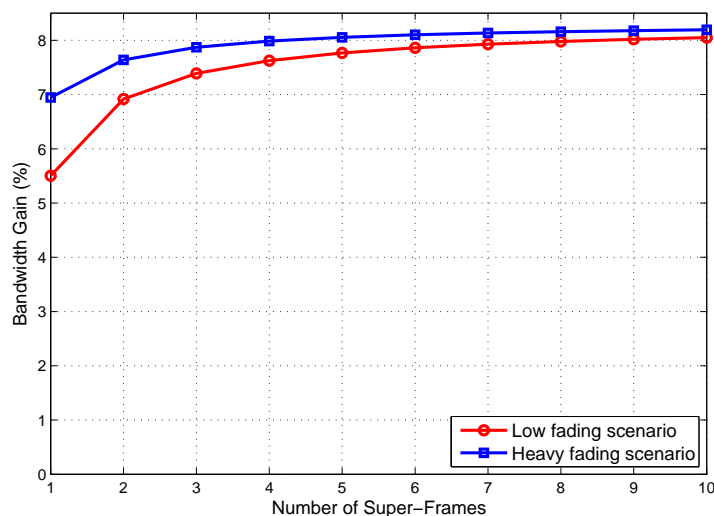


Figure 5.19: Bandwidth gain (%) depending on the SYNC rate in terms of SFs

and throughput reductions. In order to do so, we argue that taking into account higher layer parameters in order to design the allocation policies is a fair compromise between lower and higher layers. Therefore, in the PHY/MAC/IP designs presented herein, we have proposed a hybrid approach with a bottom-up PHY to MAC flow, and an IP to MAC flow of priority/bandwidth requirements.

Since the optimization is focused on the bandwidth allocation and we are only focusing on the lower layers, the performance criteria considered is the bit rate (with QoS provision). In the following chapter, we will see that a subjective parameters criteria performs better when considering higher layers.

According to the study and the analysis performed, we are able to propose a methodology for each cross-layer design (both forward and return links).

5.4.1.1 Forward link

1. Objective: Maximize bit rate allocated per video application (one per CoS), but considering the different delay requirements per video application, for adaptive satellite systems (although useful for other satellite/terrestrial systems).
2. Layers involved: PHY, MAC and IP are considered, as depicted in Figure 5.20. The cross-layer entity is placed at the MAC layer of NCC (centralized), which control the scheduler responsible of resources allocation. From PHY layer, the design needs for the available capacity of the satellite link. In particular, in case of ACM-based transmission, it needs the ACM mode used per ST. From IP layer, we need the delay measured per CoS at IP queues. Additionally, information from APP layer might be used regarding the minimum and maximum quality allowed per video application.
3. Mathematical tools: We assume NUM, which looks the best tool to define convex problems due to the versatility and the adaptive compromise between efficiency and

fairness (one can define different priorities and pricing). However, delay-constrained applications can make utility function non-concave. We propose dynamic priorities, which means that the priorities of the utility function vary depending on the measured delay at IP layers. Water-filling (low complexity) is used to interpret the solution, a rate/delay queue balancing at the scheduler.

4. Formulation: The priorities are designed as a function of queue delays. The constraints are the available capacity, minimum and maximum video quality allowed and delay requirements (implicit).

5.4.1.2 Return link

1. Objective: Perform responsible bandwidth requests (in a partially decentralized bandwidth allocation of DAMA scheme) that maximizes the bandwidth allocation for adaptive satellite systems, considering QoS provision, fairness and efficiency as design parameters. STs are able to adapt faster to system dynamics (channel and traffic) because they are able to estimate optimal bandwidth requests that will fit with future NCC allocations.
2. Layers involved: PHY/MAC and IP are considered (see Figure 5.21). The cross-layer entity is placed at the MAC layer of STs (distributed). Intra network cross-layer between the PHY/MAC layer of the NCC and the MAC layer of the ST is needed to compute available capacity: TBTP and TCT signaling tables. The information extracted from tables is the ACM mode used per ST, previous allocated bandwidth per ST and CoS of the bandwidth allocated. Explicit cross-layer from MAC to IP of STs in order to decrease/increase IP queues rates in case of reduced bandwidth, or from IP to MAC to inform about new IP flows.
3. Mathematical tool: We assume game theory, which is suitable to model non-cooperative scenarios/games where the distributed nature and the big satellite delay do not allow direct cooperation. Game theory mathematical tools allow us to model the couple price function vs. network knowledge (capacity unaware, capacity aware and requests aware). Moreover, concepts as NE, PO solution and NBS are very useful to interpret the different available solutions: system optimal point, efficiency, fairness, etc.
4. Formulation: The priorities are designed as a function of the weight allocated to each CoS. The price function depends on a tunable fair parameter that takes into account the channel conditions. The constraints are the available capacity (previous bandwidth allocation) and the maximum and minimum requested bandwidth per ST.

5.4.2 Conclusions on the performance gain

The rate/delay trade-off designed for the forward link outperforms non-cross-layer solutions thanks to the novel adaptive scheduler. It is able to control the rate allocated to

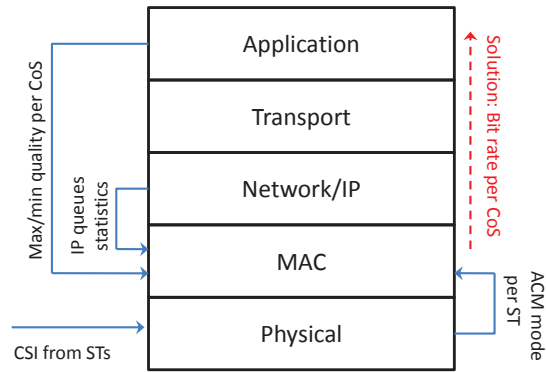


Figure 5.20: Cross-layer interaction and methodology conclusions: Forward link

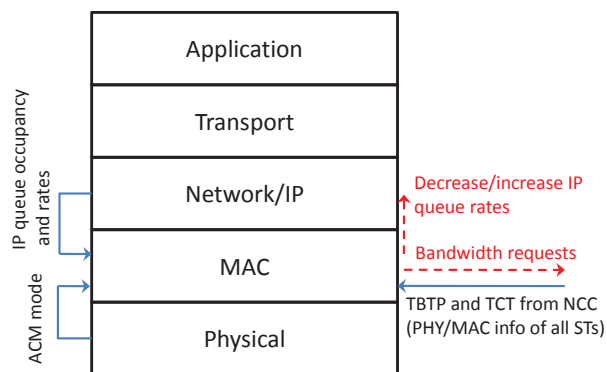


Figure 5.21: Cross-layer interaction and methodology conclusions: Return link

each CoS at the APP, to fulfill QoS requirements of premium services and, at the same time, minimize the average delay of all services: to reduce delay up to 200 ms for 12 dB rain attenuation.

The return link design is evaluated for different statistics: bandwidth efficiency, fairness and bandwidth gain. Regarding the efficiency of the bandwidth requests with respect to the real bandwidth allocation, we demonstrate that the distributed solution is almost optimal (efficiency around 99.9%). Second, we show that a our solution can define different fairness degrees depending on a controlled parameter. Finally, our design is able to reduce the signaling overhead when heavy channel conditions are affecting the satellite link, obtaining a bandwidth gains up to 8%.

Furthermore, we demonstrate that the proposed designs are feasible thanks to a detailed protocol and cross-layer design. We have shown that the forward link design can be implemented in any ACM-based standard with DiffServ QoS provision. Regarding the return link, the design is suitable for the DAMA scheme of a DVB-RCS air interface.

Chapter 6

Cross-Layer Design at PHY/APP Level

6.1 Objective: adaptive scalable QoS and error protection (VoIP and Video applications)

As we already mentioned, extended research has been carried out in the topic of higher layer mechanisms that aim at delivering different quality services. In particular, many works have been focused on multimedia applications such as VoIP, Videoconferencing or Videostreaming. However, the interaction of these applications with physical layer mechanisms, and in particular the previously introduced adaptive physical layer and hierarchical modulation, has not yet been addressed in depth.

In this chapter, we propose several cross-layer solutions that allow the interaction between PHY and Transport/APP layer mechanisms for two different methodology topics: adaptive scalable QoS and adaptive error protection, for different multimedia applications. In particular, the solutions addressed are classified as follows:

1. Bit rate adaptation (Scalable QoS): The designs focus on improving/optimizing quality of the transmission in terms of rate/delay/jitter.
 - (a) *Layered Video Multicasting*: We present a solution to optimally distribute resources across MAC, IP and APP layers to deal with the ACM at the PHY layer of a multi-beam broadband satellite communications. We use the bit loading balancing across the IP queues (taken from the Chapter 5), and we solve the video layers distribution to be sent among the adaptive physical layers, all constrained by delay requirements.
 - (b) *Adaptive VoIP Codecs*: Two different cross-layer solutions are analyzed depending on the entity that performs the adaptation, either the end-user or the control center:
 - i. *Distributed design - Rate control*: We present a cross-layer optimization of VoIP real time traffic transmission over satellite systems implementing DVB-S2/RCS. In particular, we aim at reducing total transmission delay

and jitter. We propose an end-user-centric scenario where codec rate adaptation is performed in a distributed way by the terminals involved in the call session. The design aims at individually improving the video quality of the users involved in the VoIP call.

- ii. *Centralized - Load control*: In order to maximize the overall VoIP quality of the network, we propose a centralized scenario where VoIP load control is performed through a physical-transport-application cross-layer information flow in a hybrid network (DVB-S2/RCS and WiMAX, detailed in Section 3.3) and we compare performance under temperate and tropical propagation conditions. Realistic modeling is considered for both the system and K_a band channel (and for both approaches).
2. Adaptive Error Protection: Based on the previous rate adaptation designs, we aim at improving/optimizing the protection schemes of satellite transmissions.
- (a) *High layer-FEC protection and PHY Layer optimization*: According to the obtained results of the *Adaptive VoIP codecs* designs in tropical areas and the problems of degraded reception conditions, we aim at providing different cost-efficient solutions for the channel impairments in these areas. We consider the hybrid architecture to extend services to isolated areas, similar to the one presented in Section 3.3), but based on DVB-S2/RCS + Wi-Fi networks. In order to avoid the QoS reduction, we focus on different Fading Mitigation Techniques (FMT), designing a multi-layer protection scheme, using Link Layer-FEC (LL-FEC) and Application Layer-FEC (APP-FEC) in higher layers, and analyzing the intervention threshold and super-frame length in the physical layer.
 - (b) *Unequal Error Protection (UEP) optimization*: Based on the *Layered Video Multicasting* design and the discussed conclusions, we aim at implementing a multi-layer unequal protection scheme, which is based on a Physical-Transport-Application cross-layer design. Hierarchical Modulation at the physical layer, and unequal erasure protection codes at the transport layer, are jointly optimized in order to enable recovering lost data in case the protection is performed separately.

6.2 Methodology: Bit rate adaptation (Scalable QoS)

6.2.1 XL Design 1: Layered Video Multicasting

As previously mentioned in Chapter 5, SVC can be understood as a Layered Video Multicasting at the APP layer. Next, we take advantage of it together with the adaptive physical layer to design a joint multicast transmission strategy by means of a cross-layer design. Both mechanisms will be interconnected by the MAC/IP Scheduler and queues detailed in Chapter 5).

The main challenge is to efficiently allocate the satellite resources considering tunable fairness policies. Moreover, note that our design must be capable to fulfill the video

quality and delay requirements of the transmitted video applications. Therefore, the contributions of the current section are:

- We solve not only the rate/delay balancing but we also propose an efficient and fair NUM that distributes video layers among the ACM modes, termed layered multicasting.
- We define the new cross-layer control flow that needs to be added to Figure 5.4 in order to perform the Layered Video Multicasting.
- The final NUM scheme is evaluated for realistic satellite channel conditions in a multicast scenario, where users are grouped according to video content requested and channel conditions.

6.2.1.1 Modeling the Scalable Video Coding

A Rate-Distortion (R-D) relationship is necessary to perform rate control at lower layers. To do so, we assume a layered SVC scheme, with L_c being the maximum number of video layers coded for video application c . Let us define the effective bit rate for video layer j as \bar{v}_j , and the bit rate of a user that receives up to layer j as:

$$\bar{v}_j = \sum_{i=1}^j v_i \quad (6.1)$$

which means that $\bar{v}_{1c} < \bar{v}_{2c} < \dots < \bar{v}_{\phi_c} < \dots < \bar{v}_{L_c}$, and where v_{1c} is the base layer and $v_{2c} \dots v_{L_c}$ are the enhancement layers. It is straightforward to see that the maximum achieved rate is $v_{max,c} = \bar{v}_{L_c}$. Let us also define $v_{min,c} = \bar{v}_{\phi_c}$ as the minimum quality (to reach a minimum satisfaction).

Assuming a sufficiently high rate environment, we propose the R-D model in [95], which considers that video rate distortion (D) when we decode up to layer j can be modeled by a decreasing exponential function of the bit rate:

$$D(\bar{v}_j) = S_v e^{-\beta \bar{v}_j} \quad (6.2)$$

where S_v is the source variance set to 100, and β is a constant.

6.2.1.2 Optimization: Rate-Delay Balancing across Layered Video Multicasting

6.2.1.2.1 Video layers and ACM: multicast mapping

In this section, we consider the full architecture (see Figure 6.1), assuming SVC in the video flows that feed the scheduler, and ACM to efficiently transmit according to users channel diversity.

In this section, we take advantage of adaptive physical layer properties. In particular, since ACM modes are sorted out from the lowest to the highest spectral efficiency and $SNR_1 < SNR_2 < \dots < SNR_M$, users that need ACM mode m are able to decode all

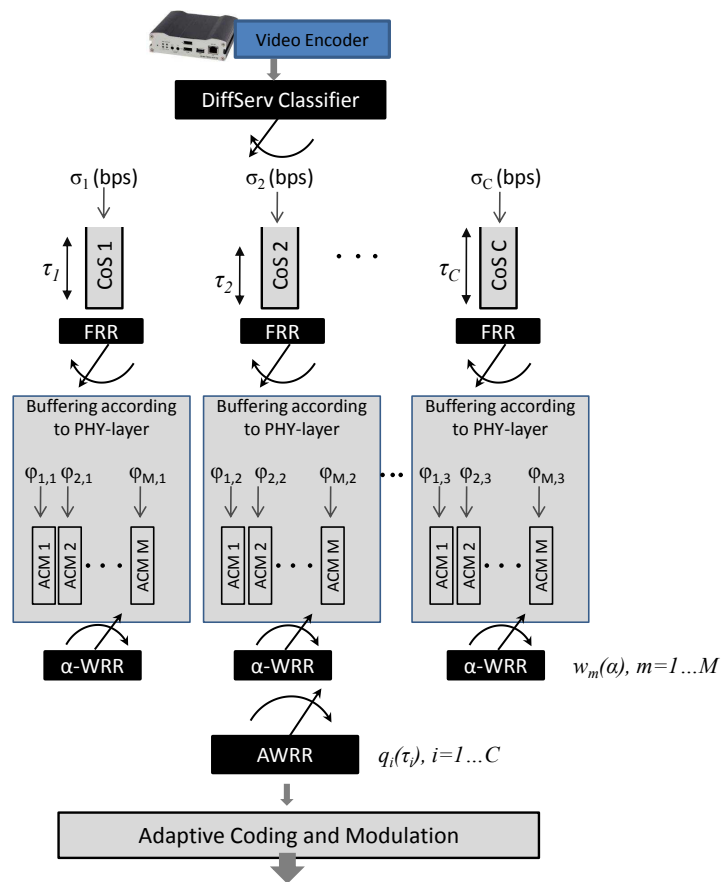


Figure 6.1: Cross-layer scheme of the adaptive PHY/MAC/IP schedulers and Layered Video

ACM modes below m (i.e. with lower SNR). We propose to distribute different video layers among the ACM modes as follows: minimum bit-rate layers are sent through the lowest ACM modes, and enhancement layers through higher ACM modes. Then, users under best and worst channel conditions are receiving the highest ($v_{max,c}$) and the minimum video quality ($v_{min,c}$), respectively. Moreover, each enhancement layer must have at least a minimum $\Delta_{min,c}$ rate, defined by the video encoder. Note that the way to distribute layers is the main challenge.

We propose to distribute video layers according to weight w_m (5.2) allocated to each ACM mode, which is able to follow different fair policies. The number of users decoding each ACM mode is also relevant when distributing video layers, since we assume that ACM modes with a big number of users associated will be prioritized to receive more video quality. Then, both parameters: weight w_m and number of users $N_{m,c}$ using ACM mode m in CoS c ; will take part of the new utility function $U_{m,c}(q_c, w_m, N_{m,c})$.

Since users are distributed according to channel conditions, let us define $X_{M_c,c}$ as the set of ACM modes (M_c among the total M modes) that are needed to transmit to users requesting video application c , and thus it is of size M_c . The elements of $X_{M_c,c}$ are sorted out from low to high efficiency modes. Let us assume $X_{1,c} \subset X_{2,c} \subset \dots \subset X_{M_c,c}$, where $X_{1,c}$ is the ACM mode with lowest spectral efficiency, and we denote $X_{M_c,c} \setminus X_{M_c-1,c}$ as the ACM mode with the highest spectral efficiency of the set $X_{M_c,c}$. Note that, we define the subset $A \setminus B$ as the relative complement of set B in set A . For the sake of simplicity, let us denote $X_{M_c,c}$ as X_c .

6.2.1.2.2 Formulation Framework

The objective of the following NUM is not only to trade delay and rate, but also to find the optimal distribution of layers to be sent through each ACM mode (to each user or group). Taking into account the rate achieved per video layer in (6.1), the previous NUM (5.26) and that video rate allocated to ACM mode m of CoS c is $\varphi_{m,c}$ (bit/s), we can express the new NUM optimization as follows:

$$\begin{aligned}
 \max_{\varphi_{m,c}} \quad & \sum_{c=1}^C \Omega_c(q_c(\tau_c)) \sum_{m \in X_c} w_m(\alpha) N_{m,c} \sum_{k \in X_{m,c}} \log(\varphi_{k,c}) \\
 \text{s.t.} \quad & \sum_{c=1}^C \sum_{m \in X_c} \varphi_{m,c} \leq C_{link} \\
 & \sum_{m \in X_c} \varphi_{m,c} \leq v_{max,c} \quad \forall c \\
 & \varphi_{m,c} \geq \begin{cases} v_{min,c} & m \in X_{1,c} \\ \Delta_{min,c} & \text{otherwise.} \end{cases}
 \end{aligned} \tag{6.3}$$

6.2.1.2.3 NUM Solution: Water Filling Layered Multicasting

Similarly to (5.12), we obtain:

$$\varphi_{m,c} = \left[\frac{\Omega_c(q_c(\tau_c))}{\lambda} \sum_{k \in \tilde{X}_{m,c}} N_{k,c} w_k \right]_{\Delta_{min,c}}^{v_{max,c}} \tag{6.4}$$

where $[y]_b^a$ means that $y \in [a, b]$, and if $y > a$ or $y < b$ then $[y]_b^a = a$ or $[y]_b^a = b$ respectively. Moreover, the subset $\tilde{X}_{m,c} \subseteq X_c \setminus X_{m-1,c}$, or in other words, $\tilde{X}_{m,c}$ includes all the ACM modes used to transmit those enhancement layers that are not transmitted using the lower ACM modes of subset $X_{m-1,c}$. Note also that when $m \in X_{1,c}$ (the lowest ACM mode), the minimum allowed rate is $v_{min,c}$ instead of $\Delta_{min,c}$.

Given that the solution happens to be also water-filling and thanks to (5.2), the volume of water, or in other words, the rate $\varphi_{k,c}$ is proportional to the α -fairness:

$$\varphi_{m,c} \propto q_c \sum_{k \in \tilde{X}_{m,c}} w_k \propto q_c \sum_{k \in \tilde{X}_{m,c}} \frac{1}{\eta_k^\alpha} \quad (6.5)$$

Two points must be highlighted from the above equation. First, in case $\alpha > 0$, within a group of users requesting CoS c , those users under bad channel conditions will receive more video quality than when $\alpha < 0$. However, in such case, the transmission efficiency is reduced (using lower ACM modes), and the congestion in IP queues increases. And second, at that point, q_c plays the role of rate-delay balancing in order to reduce congestion. Therefore, our solution can be seen as trade-off between rate, delay and multicasting.

6.2.1.3 Standard, Protocol Design and Implementation: MulFairDB

The new added features of our design such as new exchanged information between layers and the cross-layer architecture are included in this section.

First of all, let us term the new cross-layer design as Multicasting Fair Delay Balance (MulFairDB), presented in Figure 6.2, which is implemented at the NCC. The main objective of new features of the design is to make sure that the allocated video rates are correctly transferred to the APP layer. In this case, the control flow (from MAC to APP) previously named as $XL_{MAC-APP}$ not only contains the allocated rates per video application (σ_c), but also their corresponding layered rates ($\varphi_{m,c}$) to be distributed among ACM modes, both computed by MulFairDB block.

6.2.1.4 Performance Results

In order to estimate videoconferencing quality in terms of delay, we assume the ITU-T E-Model [96]. The effect of delay on highly interactive tasks quality can be estimated by the use of a curve derived from the Transmission Rating (or R-value). Then, R-value is mapped to the Mean Opinion Score (MOS), defined as the subjective methods mandated in [97] to measure user satisfaction.

At the application layer, H.264/AVC [98] is considered, with rates from 64kbps-2Mbps, obtained using one base layer and enhancement layers.

The performance of the layered mapping between the ACM and the SVC is well illustrated in Figure 6.3. The way MulFairDB distributes the video layers among the ACM modes is compared to a Fair Queuing (FQ) policy, which performs and equitable distribution. Note that in these results, AWRR interacts within MultFair (since it is part of our cross-layer design and the NUM), but AWRR is also working when FQ policy is used, although they do not interact by means of cross-layer. FQ allocates 1/3 of the video to each ACM modes, meaning that users under deep and soft rain conditions can decode

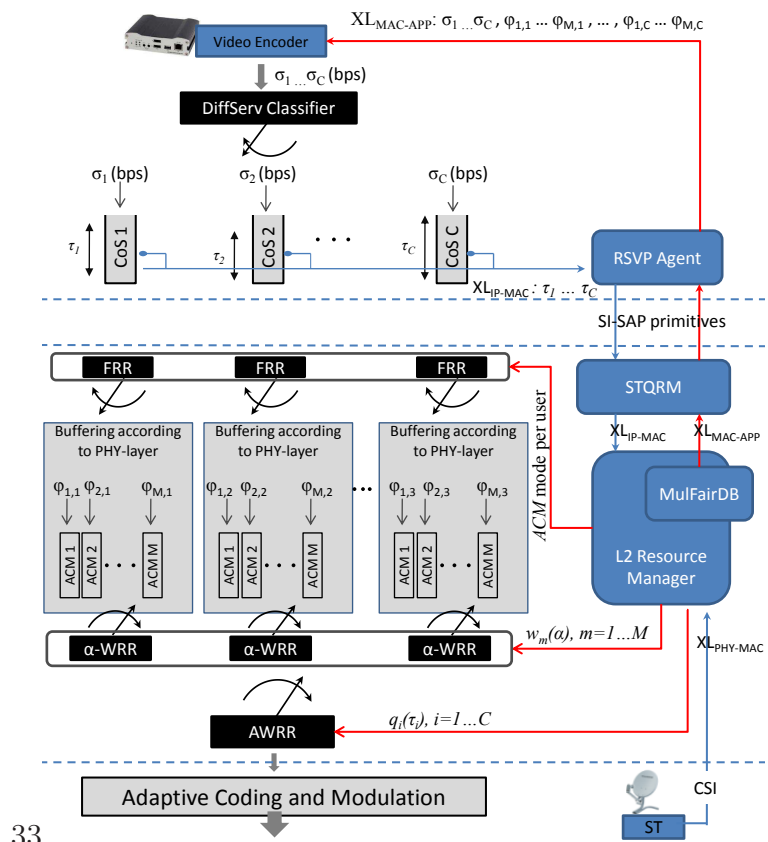


Figure 6.2: Architecture of the cross-layer design, including exchanged information and per-layer block description

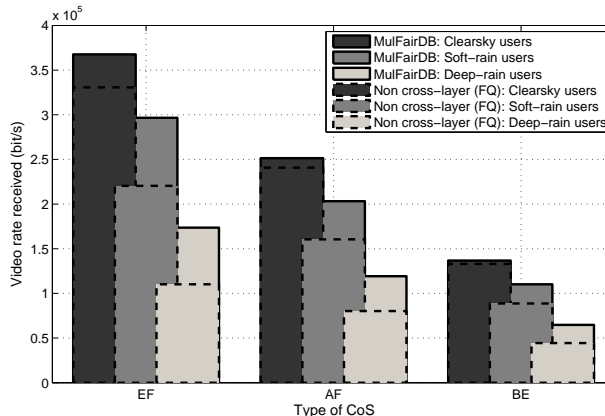


Figure 6.3: Comparison between MulFairDB (for $\alpha = 0$) and a non cross-layer approach in terms of video layers distribution among the users.

up to $1/3$ and $2/3$ of the video transmission respectively. Note that users under clearsky are able to decode all the video rate. It can be observed that our solution outperforms the non cross-layer approach in terms of video rate received by all users, even by the worst users, and per type of application (or CoS). That is because our solution distributes video layers according to the fairness criteria of the α -WRR scheduler.

A similar comparison is performed in Figure 6.4 in terms of average rate received per group depending on channel attenuation. In that case, we compare MulFairDB not only with FQ but also with a broadcast approach, where all video layers are sent via the worst ACM mode, i.e. all users requesting for the same video application receive the same quality. It can be observed that MulFairDB outperforms both approaches, up to 20% and 40% of rate gain compared to FQ. Regarding the delay experienced by users, depicted in Figure 6.5, we demonstrate that FQ achieves the best result since it is the one that transmit less video rate to rain users (see Figure 6.3). However, the difference with MulFairDB is only of up to 20 ms, and depending on the value of α , our solution can reach better values of delay, meaning that it works as a trade-off between rate and delay. Note that broadcast obtains the worst results in terms of both measures since it follows a very inefficient strategy.

Let us highlight that the good result, in terms of delay, of the FQ policy are thanks to the AWRR (which works as a rate-delay balance). If AWRR is disabled, the results become worse, and MulFairDB clearly outperforms FQ in terms of MOS, as observed in Table 6.1.

6.2.2 XL Design 2: Adaptive VoIP Codecs

VoIP is gaining popularity due to its cost benefits, moreover, it is evident from the thrust exhibited by enterprises and service providers to migrate towards a converged IP based network. Our aim is to focus on optimizing the VoIP transmission taking in to account the adaptability of the PHY layer. Our proposed cross-layer mechanism is based on the application layer adaptation across networks to capacity dynamics in the satellite link.

For a communications network to comply with the stringent QoS requirements of

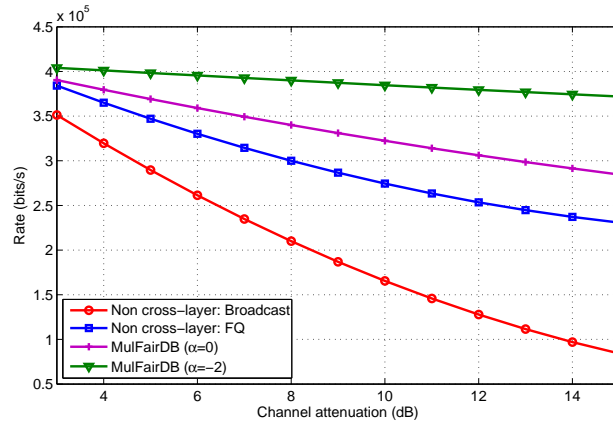


Figure 6.4: Comparison between MultFairDB and a non cross-layer approaches in terms of average rate received per group depending on channel conditions.

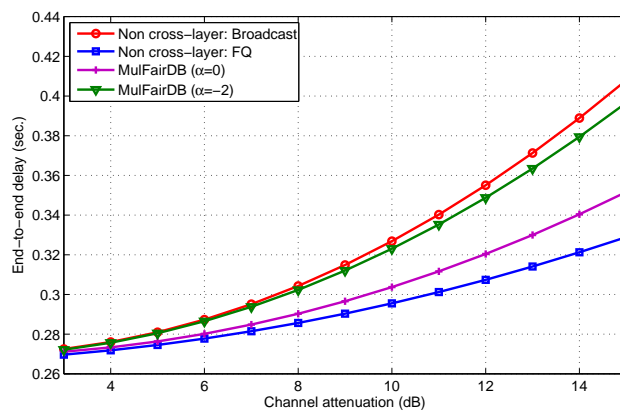


Figure 6.5: Comparison between MultFairDB and non cross-layer approaches in terms of end-to-end delay depending on channel conditions.

Table 6.1: Delay and MOS for different ways to distribute video layers, for fairness parameter $\alpha = 0$

	Max. Delay (s)			R-val (EF)	MOS (EF)
	BE	AF	EF		
Broad	4	1	0.380	69.1	3.55 (Many users dissatisfied)
FQ	1.3	0.43	0.294	79.9	4.02 (Some users dissatisfied)
MulFairDB	0.62	0.32	0.273	82	4.1 (Satisfied)

VoIP traffic, a number of cross-layer optimization techniques have been recently proposed for improving the overall performance [99, 100]. However, cross-layer has not yet been exploited for hybrid network design. Specifically, we focus on a cross-layer approach tailored to the hybrid terrestrial-satellite scenario. We propose that the terrestrial networks attached to the satellite network perform cross-layer load control in order to follow the satellite significant capacity dynamics (due to the adaptability of the physical layer to the propagation channel conditions). To this end, we compare two possible approaches: a distributed rate control and a centralized load control. Moreover, we also compare two different adaptation speech mechanisms. The first one based on taking advantage of the Real Time Control Protocol (RTCP) and the second one based on using adaptive wideband codecs, whose packets bear cross-layer signaling across the networks. Realistic modeling is considered for both the system and channel.

Note that we assume the hybrid architecture detailed in 3.3. As we introduced in Chapter 3, both standards, DVB-S2/RCS and WiMAX, are compliant with upper layer protocols and thus IP protocol in particular, allowing to support all multimedia services from Voice over IP to high speed Internet and video transmission.

Here we first explain the concept of capacity drop due to DVB-S2/RCS transmission adaptation to channel conditions. Second, we highlight the difference between the two methodologies used: centralized and distributed. Finally, we detail both designs, define the cross-layer interaction and show the results.

6.2.2.1 Capacity drop due to DVB-S2/RCS adaptation to channel conditions

ACM has been considered as a powerful tool to further increase system capacity in DVB-S2, allowing for better utilization of bandwidth resources. However, when atmospheric conditions (such as rain or clouds) affect the backhaul satellite link, data must be protected in order to avoid increase PER (Packet Error Rate), thus reducing the spectral efficiency. This reduction of the spectral efficiency can result in capacity drops and so a significant decrease of voice quality (bigger delay and jitter) and the active VoIP connections in the system should be dropped. The same applies for DVB-RCS but in a less significant way since overall bandwidth is lower than in the forward link and only Adaptive Coding (AC) is performed, i.e. less physical layer transmission modes are available.

Therefore, in order to maintain active connections and the quality requirements, we propose a cross-layer load/rate control based on VoIP rate adaptation.

6.2.2.2 Centralized vs. Distributed Approaches

VoIP rate adaptation is based on codec selection algorithms, which can be classified into two different types: network-centric or centralized approach and end-user-centric or distributed approach. In the first case, the network is responsible of controlling the delivery of the best possible speech quality to all the mobile terminals at any instant. In the latter case, the terminal has complete responsibility on the link quality measurements and the control of the speech coding mode.

We first propose a distributed approach, where the user terminal monitors the link quality measuring different statistics (such as delay, jitter). When it realizes the Channel State Information (CSI) changes due to weather conditions, it sends a message to the other end terminal specifying the codec to be used, i.e. to adapt to channel variations. A centralized approach seems to be more appropriate in a satellite system, where the network is controlled by the NCC, and even more appropriate in hybrid networks, where the architecture complexity increases. However, a distributed approach is more adequate to the Internet backbone network connected to the satellite system. Moreover, in the distributed case, the terminal can optimize its own quality call according to the measures.

The drawback of the distributed approach is that the full benefit can only be obtained if all the terminals implement the advanced mode selection algorithms. Moreover, the centralized approach is able to optimize the overall performance of network, sometimes benefiting some users while reducing the quality of others. Therefore, we also propose a centralized approach, since it can fully optimize the cross-layer load adaptation by making use of the available information at a central entity regarding the satellite link, the WiMAX network and VoIP connections. In particular, in our proposed hybrid architecture, the codec selection is performed by the NCC, which plays the role of load controller.

6.2.2.3 Distributed design: Rate Control

We propose two adaptation algorithms. First of all, we assume a bank of codecs to be available at both transmission ends. In this case, a codec switch is performed driven by the RTCP reports, and a transport-to-application layer cross-layer information flow is required. Second, we propose the use of Adaptive Multi-Rate Wideband (AMR-WB) codecs (inherently cross-layer across the hybrid network).

6.2.2.3.1 RTCP-based rate adaptation

Our first proposal of cross-layer distributed rate control for VoIP flows relies on the RTCP reports [101]. We propose a bank of codecs be available at both transmission ends and a codec switch is performed driven by the RTCP report. In this case a transport-to-application layer cross-layer information flow is required. The architecture is sketched in Figure 6.6, where the required information flow is also shown.

RTP receivers provide reception quality feedback using RTCP report packets which may take one of two forms depending upon whether or not the receiver is also a sender. The only difference between the sender report (SR) and receiver report (RR) forms, besides the packet type code, is that the sender report includes a 20-byte sender information section for use by active senders. The SR is issued if a user has sent any data packets

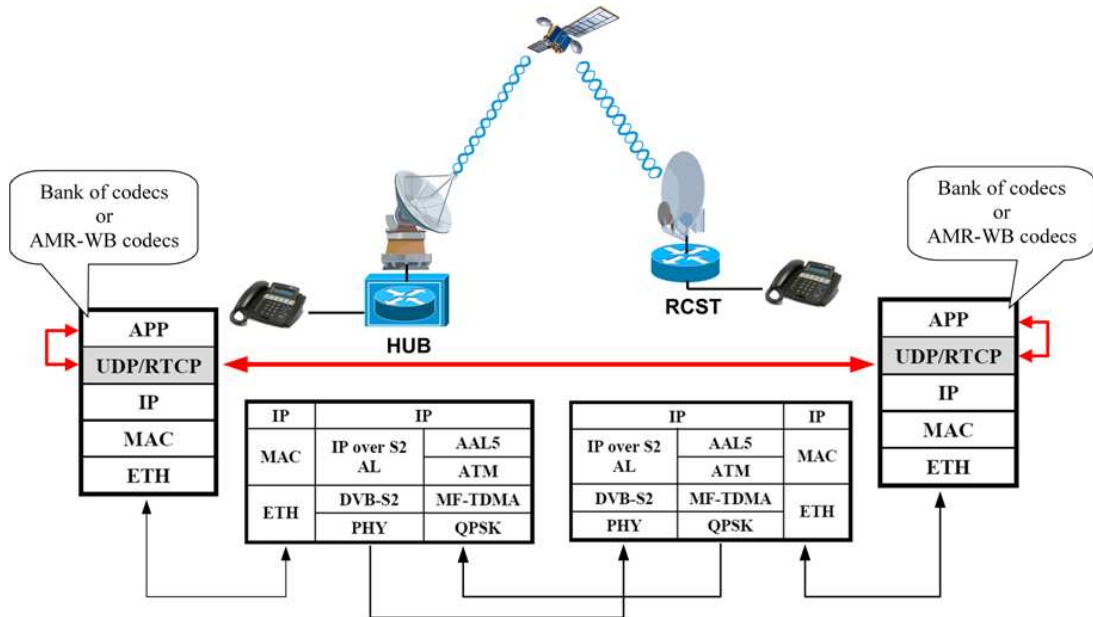


Figure 6.6: RTCP-driven and AMR-WB architectures

during the interval since issuing the last report or the previous one, otherwise the RR is issued.

The receiver checks the statistics and computes the desired codec rate (increases/decreases the VoIP codec if requirements are fulfilled /not fulfilled). This codec is sent back to the sender using RR (receiver reports), which changes its codec. All adaptive mechanism is performed with control messages between the two users without necessity of cross-layer interaction between SD and SI layers and then, without using primitives. This is called cross-layer across the network. However a small cross-layer signaling between the Transport and APP layers of the end-user is required. Same proceeding is allowed in the return link.

The adaptation performs as follows. A user sends RTP packets, which contains VoIP frames. The receiver is capable to extract the following information from these packets: jitter, delay and packet loss. Jitter is measured by RTCP and included in the RR messages sent by the receiver. As this value is measured in sampling units, in order to convert to time units, one must divide by the sampling rate of the media codec. Delay between two peers can be calculated with the difference among three times, such as, Delay since Last Sender Report (DLSR), Time of Last Sender Report (TLSR) fields in RTCP receiver report packets and report receiving timestamp. Moreover, inter-arrival delay jitter and packet loss rate are obtained from the inter-arrival jitter field and cumulative number of packet lost fields in RTCP receiver report packet, respectively. Inter-destination delay jitter can be calculated with the delay values received from all the other multicast group members.

We assume a bank of codecs is available at each transmission end, which is a realistic assumption for many of currently available VoIP software packages. When the sender receives the RR, we also assume that the application can switch to a different codec according to the information extracted from the RTCP reports. In particular, the switching

Table 6.2: Codecs considered for RTCP-driven scenario

ITU-T codec	Modulation Type	Bit Rate	Reason to be selected
G. 711	Companded PCM	64 Kbps	Narrowband, most commonly used, "Tool quality"
G. 722.1	Transform Coding	24/32 Kbps	Wideband
G. 729	CS-ACELP ^a	8 Kbps	Narrowband, most commonly used after G. 711

^a Conjugate-Structure Algebraic Code Excited Linear Prediction.

to a lower (higher) bit-rate codec takes place whenever the delay or jitter value reported by the RTCP is above (below) than the required ones. In that case, either the Transport layer can inform the APP layer about the measured statistics, or if it is aware of the available codecs and the current codec used, the Transport layer can send to the APP layer the desired codec to be used. We consider the codecs shown in Table 6.2, where the key technical characteristics along with the reason to be selected are presented. The methodology is applicable to any other set of codecs.

6.2.2.3.2 AMR-based rate adaptation

Our second cross-layer mechanism is based on the RTCP reports interpreted along with the signaling of AMR-WB (Adaptive Multi-Rate Wide Band).

AMR-WB coding algorithms are based on an Algebraic Code Excitation Linear Prediction (ACELP) technology. This same technology has been used in various speech codec standards, such as GSM enhanced full rate (GSM-EFR) (3GPP TS 06.51) and narrowband GSM-AMR (3GPP TS 26.071). The main novelty in AMR-WB is the sub-band structure, which enables significant savings in complexity and memory consumption. The audio band is split into two frequency bands so that the internal sampling frequency of the core is 12.8 KHz, having an audio bandwidth of 50-6400 Hz. Separate processing is performed for the frequency range from 6400 to 7000 Hz: more bits can be allocated to the perceptually important lower band.

The multi-rate approach provides the flexibility to achieve an optimal balance between error protection and source coding within a fixed bit-rate budget. When applying a lower source coding rate, the system can allocate more bits to channel coding; thus, enhancing the speech quality by improving robustness.

It should be stressed that the utilization of wideband speech in an IP network does not introduce additional system complexity compared to narrowband. Within the IP network the operation is by definition transcoder free, since compressed speech is transmitted in IP packets end-to-end.

We propose to use the cross-layer field of the AMR-WB codec called CMR (Codec Mode Request) for the VoIP flow rate adaptation. The CMR is included in the header of the RTP packets (containing VoIP frames), and therefore, no signaling is needed in this case. The CMR indicates to the end user the desired coding mode. The CMR must be computed (to be written in the IP voice payload) and extracted from the IP

Table 6.3: Codecs considered for the AMR-WB based scenario

Mode (Kbps)	Size of a speech frame (bits)	RTP payload size (bytes)
23.85	477	62
23.05	461	60
19.85	397	52
18.25	365	48
15.85	317	42
14.25	285	38
12.65	253	34
8.85	177	24
6.60	132	19
1.75 ^a	40	7

^a Assuming silence indicator (frames are continuously transmitted).

payload received (to select the appropriate codec for transmission). The CMR is computed based on quality measurements, which is assumed to be based on both the link state and the RTCP reports. The speech coder has nine different encoding rates (23.85 Kbps to 6.60 Kbps) in addition to low background noise encoding rate (Table 6.3), providing the necessary flexibility for VoIP load/rate control. The 12.65 Kbps mode is the lowest mode that can offer high quality wideband speech (the two lowest modes are intended to be used only under bad radio link conditions or high loss rate). Furthermore, the highest mode of the AMR-WB codec provides speech quality equal to the ITU-T G.722 wideband codec at 64 Kbps. The codec can change the encoding mode every 20 ms (i.e., every frame). The AMR codec is able to operate in source controlled rate mode, where the background noise is coded with a lower bit rate between two talk spurts. The end of a talk spurt is detected by the voice activity detector. During the silence periods the comfort noise generator functions produce parameters that describe the characteristics of the real background noise. At the receiving side, error concealment functions are used to hide the effect of missing speech frames.

The major differences between the adaptive RTCP-based scenario and the CMR based scenario are:

1. The frequency of the reports sent by the RTCP may not be synchronized with the speed of adaptation required by the PHY.
2. The highest bit rate of the AMR-WB codecs is lower (less than half) than the G.711 and hence, the same system load allows for a higher number of VoIP connections

Note that the inbound cross-layer signaling of both AMR-WB and RTCP allows for a fully distributed cross-layer VoIP bit rate adaptation. Note also that the algorithms to "measure quality" are open and undefined. In our case, signal quality is measured according to [11], while delay is computed as explained below. Finally, it is interesting also to note that this cross-layer solution is fully scalable and has no impact on the DVB-S2/RCS and WiMAX standards and architecture.

A simple algorithm for the VoIP adaptation is assumed based on two thresholds; one for decreasing and another for increasing codecs. A simple algorithm for the VoIP adaptation is assumed based on two thresholds; one for decreasing and another for increasing codecs. In particular, when delay measured by RTCP reports exceeds the maximum allowed, speech codec is decreased since it does not fulfil the requirements. On the other hand, if the delay is lower than maximum required (less a specific hysteresis range in order to avoid fluctuations between VoIP speech rates when channel experiences fast fading) speech codec is increased. Such example is valid for both RTCP-driven and AMR rate adaptation. The pseudo-code Algorithm for speech rate adaptation is shown below.

```
%%% VARIABLES %%%
% Ed and Ej are the estimated delay and estimated jitter respectively.
% Max_delay_up: delay threshold for decreasing the cod_type
% Max_jitter_up: jitter threshold for decreasing the cod_type
% Max_delay_down: delay threshold for increasing the cod_type
% Max_jitter_down: jitter threshold for increasing the cod_type
% Cod_type: actual speech rate code number
% Rb[ ]: array with all speech rates
% Change_code_rate(rate value): Changes the speech rate of VoIP users.

%%% ALGORITHM %%%
if (Ed >= Max_delay_up) || (Ej >= Max_jitter_up)
    if Cod_type > 1
        Cod_type--;
        Change_code_rate(Rb[Cod_type]);
    end
elseif (Ed < Max_delay_down) && (Ej < Max_jitter_down)
    if Cod_type < 7
        Cod_type++;
        Change_code_rate(Rb[Cod_type])
    end
end
end
```

In order to summarize, let us present Figure 6.7, which details the VoIP frames (RTP) and control packets (RTCP, RR and SR) flows for both mechanisms.

6.2.2.3.3 Delay Budget Model and Performance Model

We propose the following delay budget model for our scenarios:

$$T_{tot} = T_{codec} + T_{MAC} + T_{trans} + T_{prop} + T_{payout} \quad (6.6)$$

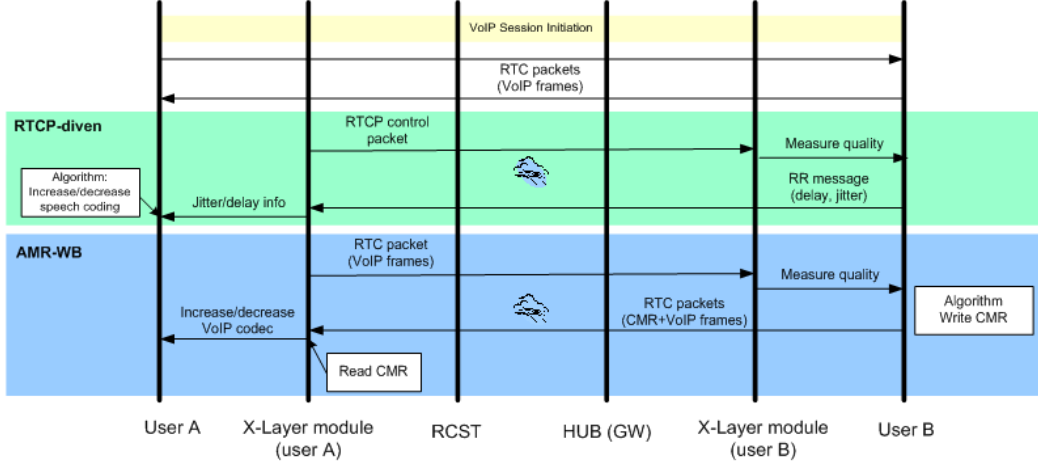


Figure 6.7: RTCP-driven and AMR-WB flows

Table 6.4: Relevant Codec Delays

Time (ms)	G.711	G.722.1	G.729	AMR
T_{la}	0	20	5	5
T_{payout}	10	20	10	≤ 10
T_{pack} (N=1)	0	0	0	0
T_{pack} (N=2)	20	20	10	20
T_{pack} (N=3)	40	40	20	40

$$T_{tot} = T_{codec} + T_{network} + T_{payout} \quad (6.7)$$

T_{codec} is the delay introduced by the codec, which we model as follows:

$$T_{codec} = T_{fr} + T_{la} + T_{proc} + T_{pack} \quad (6.8)$$

where T_{pack} is the delay introduced when encapsulating more than one voice packet per IP packet, T_{fr} is the framing delay, T_{la} is the look-ahead delay (for prediction purposes) and T_{proc} is the processing delay.

The rest of the values are: the delay introduced in the MAC queuing and scheduling (T_{MAC}), the transmission delay (T_{trans}), the propagation time (T_{prop}) and the playout delay to smooth out the jitter (T_{payout}).

Relevant time delays for the codecs of our cross-layer scenarios are showed in Table 6.4.

Note that a rate control such as the one presented below is compatible with the forward link solution presented in Chapter 5, and in particular, with the MAC/IP scheduler of the forward link detailed in Section 5.2.4.

Voice quality (clarity) can be measured by subjective methods such as MOS mandated in ITU-T Recommendation P.800 [97] and parametric estimation (objective) methods like Perceptual Speech Quality Measure (PSQM) (ITU-T Q.861 [102]), PSQM+ and Perceptual Analysis/Measurement System (PAMS). The subjective methods are time-consuming

Table 6.5: Relationship between R-value and MOS

R-value	MOS	User satisfaction	Quality
(lower limit)	(lower limit)		
90	4.34	Very satisfied	Toll
80	4.03	Satisfied	Toll
70	3.60	Some users dissatisfied	Below Toll
60	3.10	Many users dissatisfied	Unacceptable ^a
50	2.58	Nearly all users dissatisfied	Unacceptable ^b

^a Used only in very special conditions

^b Connections with R-value below 50 are not recommended

and expensive to use while parametric estimation can be done quickly and inexpensively on the voice codecs.

We will use the ITU-T *E-Model* [96, 103] for estimating the voice quality. The application of E-Model results in the Transmission Rating or the R-factor. The resultant value of the R-factor is known as the R-value. Table 6.5 summarizes the relationship between the R-value and MOS.

The E-Model contemplates the case of situations where the user is being provided certain "advantage". In this case, a downward adjustment in the R-value is provided with a factor "A". For instance, when providing access to hard-to-reach locations via multi-hop satellite connections, we will use this degree of freedom to adjust our results for both the transparent and regenerative scenarios. It should be stressed that the extension of the E-model to wideband codecs is still under development.

6.2.2.3.4 Cross-layer and Signaling

In a distributed approach such as the one presented herein, the cross-layer flows depend on the mechanisms used:

- RTCP-driven: In such case, the information flow is performed between the transport and the application layer of the end-user. The quality of the transmission is measured by the receiver as detailed before (using RTCP control packets), which sends back the RR with delay, jitter, packet loss statistics. At the sender, the transport layer decodes the message and decides if necessary to increase/decrease the codec rate. By means of the cross-layer flow, the transport layer informs the APP layer about the new codec to be used in the VoIP frames.
- AMR-based: Here, the receiver is able to compute the quality of the speech using the RTP (VoIP frames) packets. The receiver computes the desired codec rate, which is sent back to the sender using the CMR header of RTP packets. Therefore, the adaptive mechanism is performed with control messages between both users, and no explicit cross-layer is identified. The design is generally termed as inherently cross-layer across the network.

Therefore, we only need to define signaling flows between the Transport and APP layers for the RTCP-driven approach. We propose the following signaling flow for the

RTCP-driven scenario:

- Signaling flow name: *C – VOIP_INFO – * * **
- Function: To send the desired VoIP codec from the Transport layer to the APP layer
- Parameters:
 - *codec_mode*: field with the next VoIP codec to be used.
 - *Session_id*: id of the opened VoIP session (in case user has several opened sessions, such as multiple calls).

We are assuming that the Transport layer has all the available codecs to be used during the VoIP call, which is also considered as an offline cross-layer interaction, since a layer has use of mechanism properties of a higher layer.

6.2.2.4 Simulation Results

We assume a system with the following parameters. As introduced in the QoS model (see Section 3.2.2), each class of service has specific requirements. Cross-layer rate control is configured to guarantee a maximum delay of 270 ms (i.e. 20 ms excess w.r.t. propagation delay). On the other hand, delay requirements for AF and BE are 500 ms and 750 ms respectively.

We have chosen a users distribution where 75% of users are in good weather conditions (clearsky users), while 25% of users are in bad weather conditions (rain users).

Since channel conditions can vary depending on geographical areas, we have modeled a set of typical rain events (from 7 dB to 18 dB attenuation) to observe the attenuation dependence, and specially its effects on VoIP requirements (note that the rain event has been low-pass filtered simulating the effect of the channel estimation algorithm). Deep attenuations (18 dB) have been also considered to study the cross-layer mechanisms in the worst case scenario. Finally, system is loaded until 85% of its total capacity.

We first present results of the aggregated EF type of traffic for the reference scenario, where no cross-layer mechanisms are enabled, and VoIP is using the codec G. 711. Figure 6.8 shows the EF traffic delay and the aggregated throughput when a 3dB rain event is affecting the satellite link. It can be observed that channel attenuation is not noticeable since system is able to manage the traffic load. In the time axis of throughput, 1 sample corresponds to 1 second. On the other hand, Figure 6.9 shows the effect on the delay when the channel undergoes a deeper attenuation (12 dB). Note that, in the last case, the delay is above the guaranteed delay. Two different curves are displayed for the delay, the red refers to the satellite terminals undergoing bad weather conditions and the blue one refers to those under clear sky conditions, showing that the whole beam (not only rain users) is affected by rain in terms of bandwidth reduction. Such effect is due to the chosen scheduling policy (see Section 5.2.4). In the current results, the available bandwidth is equally distributed among users (clearsky or rain), however, we could use alternative policies, e.g. isolating the users in clear sky conditions, avoiding the effect of the rain in those users.

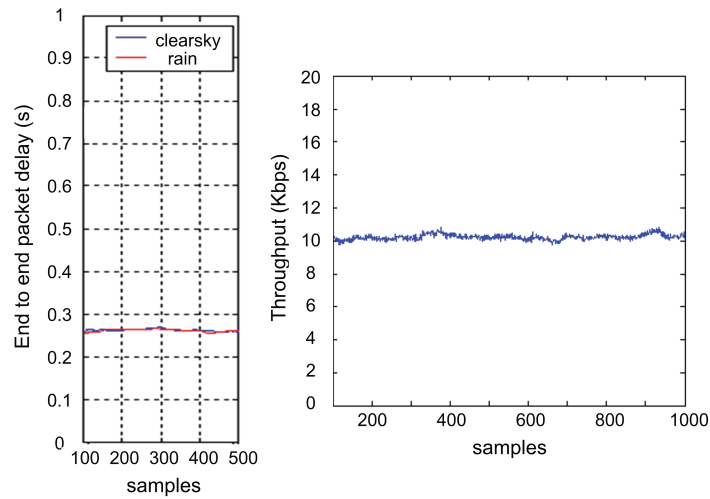


Figure 6.8: EF traffic delay and aggregated Throughput for a channel attenuation of 3 dB, VoIP using the G. 711 codec.

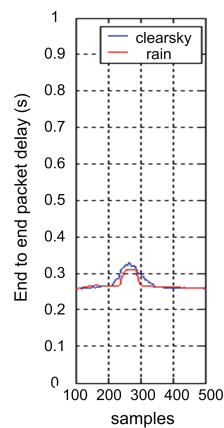


Figure 6.9: EF traffic delay for a channel attenuation of 12 dB, VoIP using the G. 711 codec.

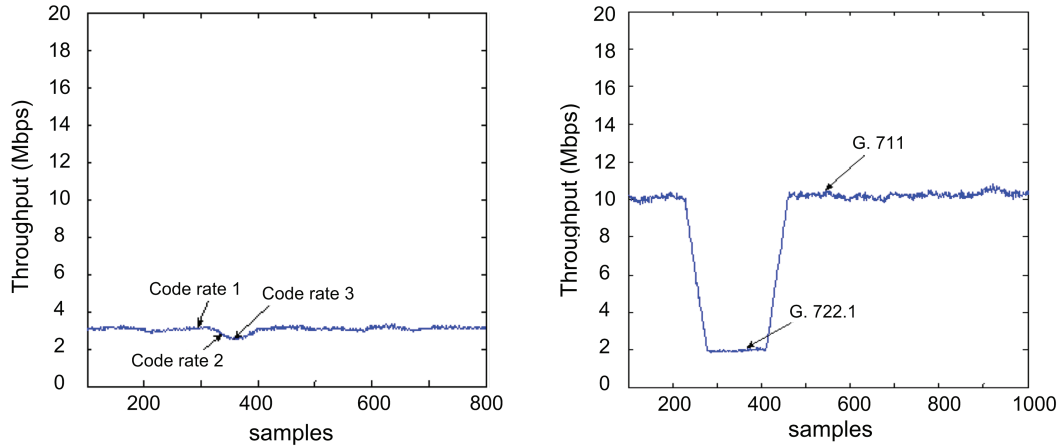


Figure 6.10: EF aggregated throughput for a channel attenuation of 12 dB for the two cross-layer approaches. AMR-WB (left) and RTCP-driven rate control (right)

Once we have checked the performance of the system without cross-layer design, we will observe the results using the two proposed approaches. Figure 6.10 shows the RTCP-driven adaptive scenario that counteracts such a load peak. Codec types are: G.711 / G.722.1 / G.729 VAD. Note the rate adaptation during the rain events (approx. five minutes long). The RTCP-based adaptive scenario can be therefore used to control VoIP bit rate to fulfil delay requirements when for some reason there is peak of VoIP traffic or a capacity drop. As a side effect of the VoIP adaptation results obtained, we conclude that it can be seen as system load control. The load control concept will be exploited in the centralized approach (see Section 6.2.2.5).

Figure 6.10 (left) shows the AMR-WB-based adaptation for the same system and channel conditions. It can be observed that this case outperforms the RTCP-driven based due to the following reasons:

- The cross-layer scenario does not add signaling overhead (RTCP reports) on the system.
- The AMR-WB codec has inherent higher voice quality (Wide Band).
- The AMR-WB highest bit rate is lower than the G.711 bit rate and therefore the system can admit a higher number of connections.
- Like in the RTCP-based adaptive baseline, VoIP adaptation can be seen as a system load control.

A comparison between the two approaches has been carried out in Figure 6.11. We have used as reference scenario the case where only a single VoIP codec is used (G.711). Figure shows that Bank of codecs performs better than the single 64 Kbps code, but it adapts slowly, and therefore probability of exceeding the requirement (270 ms) is already

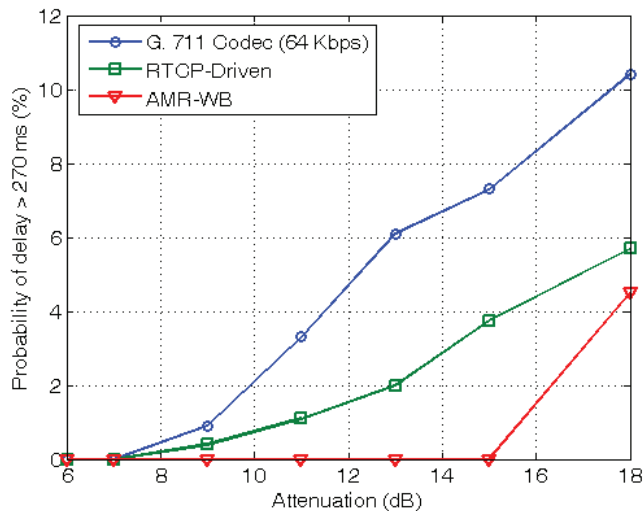


Figure 6.11: Probability of exceeding delay requirements (270 ms) for EF depending on rain attenuation. Comparison between reference scenario and cross-layer approaches.

high. For the AMR-WB scenario, we get better results; probability is zero for attenuation lower than 13 dB, only when attenuation is 18 dB the codecs cannot avoid exceeding delay threshold. Notice that 18 dB is considered deep attenuation (the probability of this kind of rain events is very low). In that case, other mechanisms such as Connection Admission Control or weight adaptation (as in the previous chapter) would be necessary to avoid congestion effects.

Finally, Figure 6.12 shows the R factor (see Table 6.5) assuming two CoSs defined by two different delay guarantees, namely CoS1 guarantees a maximum delay of 270 ms and CoS2 guarantees a maximum delay of 290 ms. It is observed not only the effect of the number of hops in the quality regardless rate adaptation, but also the difference between class of service guarantees. The 1-hop scenario achieves better quality factor than 2-hops scenario. In terms of delay, for 1-hop scenario, delay is around 125 ms lower, which explains the difference of R-Factor.

6.2.2.5 Centralized approach: Load Control

A distributed approach, as the one detailed above, aims at optimizing individually each speech quality, and it obtains the best result only when all terminals are implementing the same VoIP rate adaptation. In the centralized scenario, where we also assume both mechanisms for rate adaptation (RTCP-driven and AMR-WB), the NCC monitors the network using the information reports and the Channel State Information (CSI) from all users in order to adapt the system capacity not only to channel variations but also to traffic increase. The NCC aims at maintaining the number of active connections by means of a cross-layer load control, and therefore, it is able to maximize the overall quality.

In order to maximize the active connections, we need to dimension both subnetworks (satellite and terrestrial).

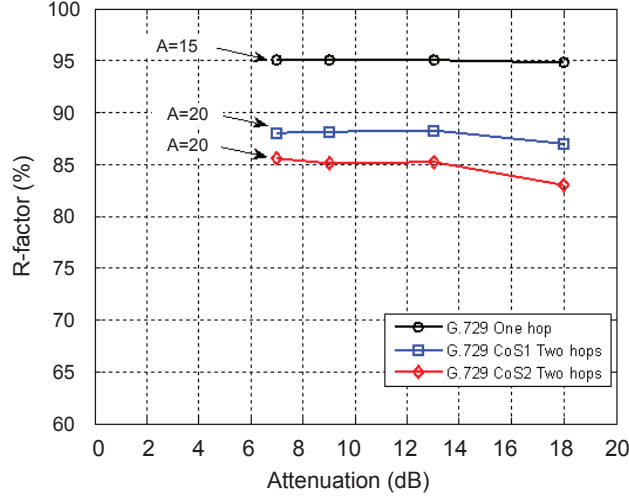


Figure 6.12: R-factor for G. 729 for different delay guarantees and number of satellite hops.

6.2.2.5.1 Dimensioning of Satellite subnetwork

For the satellite subnetwork, we consider a R_s bandwidth channel for the DVB-S2. The maximum capacity of connections is achieved using the highest VoIP speech rate as follows. Each speech payload packet has a length $n^{codec} = R^{codec} t_{pkt}$, where R^{codec} is the codec speech rate (in kbits/s), and t_{pkt} is the packet duration (20 ms for all codecs). After adding all headers (RTP/UDP/IP of 12/8/20 bytes respectively), the total number of VoIP bits is $n_{VoIP}^{codec} = n^{codec} + h_{RTP} + h_{UDP} + h_{IP}$. On the other hand, the number of bits per frame in the DVB-S2 standard depends on the spectral efficiency and thus the number of users that can be served is:

$$N^{S2}(m) = \left\lfloor \frac{T_f^{S2} \cdot \eta_m \cdot R_s}{n_{VoIP}^{codec}} \right\rfloor \quad (6.9)$$

being η_m the spectral efficiency of ACM mode m and T_f^{S2} the duration of the DVB-S2 frame designed here as 20 ms. This maximum number of users that can be served are assumed to be uniformly distributed among the satellite terminals within the beam and so to the WiMAX networks attached to them.

6.2.2.5.2 Dimensioning of Terrestrial subnetwork

In the Terrestrial WiMAX Subnetwork, the purpose of the VoIP adaptation is to keep the offered capacity by performing load control, i.e. to keep the number of active connections by matching the aggregated VoIP capacity to the satellite link.

In this case, an adaptive physical layer composed of 6 ModCods [16] is assumed in order to counteract channel fading. CSI of user is available in the ACM scheduler. Thus, both physical layer and VoIP are assumed to be able to adapt to medium-long term variations due to propagation losses (not to short-term fading). We assume the propagation model for the 3.5 GHz band detailed in Section 3.1.2.

In a 5MHz (with 512 FFT size) WiMAX model, the number of bits per sub-channel is computed as follows. From [16], only 360 of the total 512 sub-carriers are used for data, the rest are used for pilots (60) and nulls (92). Therefore the number of users per sub-channel is:

$$N_{b,sub-channel} = \frac{T_f^{WiMAX} \cdot 360 \cdot \eta_m}{T_s \cdot n_{VoIP}^{codec}} \quad (6.10)$$

where T_s is the symbol period, then:

$$N^{WiMAX}(m) = \left\lfloor \frac{T_f^{WiMAX} \cdot 360 \cdot \eta_m}{T_s \cdot n_{VoIP}^{codec}} \right\rfloor \quad (6.11)$$

We assume here that the capacity of the WiMAX network is not affected by the mobility of its users since only when all users move to the limits of cell coverage, the network experiences a relevant reduction in VoIP connections. Therefore, the main cause of capacity dropping is due to the satellite link capacity dynamics, which is affected by atmospheric conditions.

6.2.2.5.3 Optimization: Load Control

In a realistic scenario, the end users might use different codecs and their corresponding WiMAX BS might be attached to different Satellite Terminals, which at the same time might be affected by different channel conditions. Such requirements need to be contemplated in the load control. Following the definition of NUM used in Chapter 5, next, we propose a VoIP load control.

Let us assume that a VoIP packet of n_u (bits) is sent by user u in order to maintain a VoIP call during a frame. The satellite capacity to be shared among U users, in terms of number of bits during this frame, is $T_f^{S^2} \cdot \bar{\eta} \cdot R_s$, where $\bar{\eta}$ is the average spectral efficiency of the transmission (among all users). Note that the solution of n_u are limited to the available codec sizes, and so $n_u \in \{n^1, n^2, \dots, n^C\}$. The objective of the NCC is to maximize the capacity of the satellite link:

$$\begin{aligned} \max_{n_u} \quad & \sum_{u=1}^U \Omega_u n_u \\ \text{s.t.} \quad & \sum_{u=1}^U n_u \leq T_f^{S^2} \cdot \bar{\eta} \cdot R_s \\ & n_u \in \{n^1, n^2, \dots, n^C\} \end{aligned} \quad (6.12)$$

Let us assume different policies for the priority Ω_u weight included in the utility function. For example, by allocating the same weight to all users, the load control aims at maximizing/maintaining the number of users. However, we can also apply the tunable fairness that depends on channel conditions 5.2 (i.e. $\Omega_u = w_u$), or different service priorities (emergency calls, business calls or home calls).

Note that a load control such as the one presented below is compatible with the forward link solution presented in Chapter 5, and in particular, with the MAC/IP scheduler of the

forward link detailed in Section 5.2.4. Therefore, we assume this MAC/IP architecture for the simulation results.

6.2.2.5.4 Cross-layer and Primitives

The VoIP frames are considered as aggregated EF type of DVB-S2, which is controlled by the MAC/IP scheduler to guarantee that speech delay does not exceed the maximum allowed. The centralized approach take advantage of the cross-layer flows at the NCC, and no RTCP packet exchange is necessary to be aware of delay and jitter.

Let us consider the PHY/MAC/IP queues and scheduling scheme presented in Figure 6.2, which are designed for the previous cross-layer scheme. However, in this case, the optimization is performed at the APP layer of the NCC, which is able to adapt to the channel conditions. In particular, herein the entity responsible of the codec adaptation is called Speech Codec Manager. The cross-layer performs as follows:

1. The PHY layer of the NCC receives the CSI feedback from Satellite Terminals.
2. PHY layers write primitives to the APP layer specifying the ACM mode of each terminal.
3. The Speech Codec Manager at APP layer computes the available capacity and performs (6.12) in order to recompute the codecs of VoIP calls (either the bank of codecs or the AMR codecs).
4. The CMR is included in the header of RTP packets, and it is used by the NCC to indicate to end-users the desired codec to be used in order to perform adaptation.

Therefore, we can suppose direct communication between PHY and APP layer using primitives. However, since [10] primitives only take into account MAC-IP interaction, we can use them to upward an information flow from MAC to IP layer with the new ACM mode. Then IP layer should send the information to the Speech Codec Manager (Application layer), by means of some of the signaling methods detailed in Chapter 3, which will compute the new speech codec rate. We propose the following primitive, which is our contribution to the SI-SAP primitives defined in ETSI TS 102 463 for VoIP Cross-Layer Design (*Survey on the envisaged explicit cross-layer approaches for BSM*):

- Primitive name: *SI - C - PHY_INFO - **** (new primitive in ETSI TS 102 463)
- Function: To send from the STQRM to RSVP agent the new ACM mode (1-23) of each user.
- Parameters:
 - *ACM_mode*: the new ACM mode of a specific user.
 - *User_id*: a user id of the user that changed its ACM mode.

6.2.2.5.5 Simulation Results

In this section we obtain numerical results in terms of VoIP connections that can be maintained when a WiMAX station using DVB-S2 satellite link as its backhaul adapts VoIP rate controlled by the NCC. We assume that the delay added by the satellite hops is lower than the attenuation coherent time, which is a realistic assumption [11]. We also assume the channel model in Chapter 3, which allows us to generate channel attenuation time series with desired statistical properties. A comparison is carried out between the cross-layer approaches (both using RTCP-driven and AMR-WB) and a non cross-layer approach. The latter consists of a single codec (the highest of the RTCP-driven bank of codecs, i.e. G.711) which does not adapt to channel variations.

Regarding the system architecture, we assume the satellite serving 15 satellite terminals within the beam and only one WiMAX network attached to each terminal. Our simulations focus on the aggregated VoIP traffic of DVB-S2 traffic, which is controlled at the MAC layer to guarantee a maximum delay of 290 ms, i.e. 20 ms excess w.r.t. GEO propagation delay (250 ms) and WiMAX RTT/2 (20 ms). We assume a fully loaded system. As we mentioned before, we assume the MAC/IP architecture and the bandwidth allocation of the previous Chapter, including the weight allocation (5.2). Different possible policies can be adopted when allocating the satellite bandwidth resources among the WiMAX networks: e.g. equally allocated bit-rate (throughput) among all RCST terminals of the beam ($\alpha = 1$), or equally allocated bandwidth ($\alpha = -1$). We assume a fair policy in terms of bit-rate. In this case, the reduction of spectral efficiency due to rain experienced by some terminals affects also the rest of terminals within the beam coverage, since we have fixed a policy where all users have the same opportunity to transmit. Therefore, we take into account not only different rain events but also the percentage of users in rain conditions within the beam.

Figure 6.13 shows that the proposed cross-layer load adaptation allows the number of users for peak attenuations lower than 7 dB to be maintained even when all the users within the beam coverage experience bad weather conditions. However, using the non-cross-layer approach, capacity critically drops to 70 % for the same maximum attenuation when 75 percent of the users experience bad channel conditions. Similar behavior is obtained for a 10 dB peak attenuation, where AMR-WB and RTCP-driven mechanisms avoid reducing the VoIP connections (only 5 % of connections are lost for AMR), and the single codec can only keep 55 % of the initial connections. Finally, even for deep attenuations of 15 dB, cross-layer mechanisms are successful and only when 75 % of users in the beam are affected by rain does capacity suffer a minimal reduction. It should be stressed that a situation where 3/4 of users are under rain conditions is considered very rare since typically cloud size is comparable to 1/4 of the beam area (as stated in [11]). Moreover, attenuations of 15 dB are also a worst case scenario, since only 0.01 % of time does rain reach this value. Therefore, the cross-layer load control mechanism considerably outperforms the single codec scenario; even in hard conditions they perform successfully.

Note that the RTCP-driven approach seems to adapt better to load variations. However, this is not the case since AMR-WB actually allows a significantly larger number of simultaneous connections. This is due to the fact that the highest codec bit rate of AMR-WB is (two times) lower than the highest codec of the RTCP-driven bank of codecs

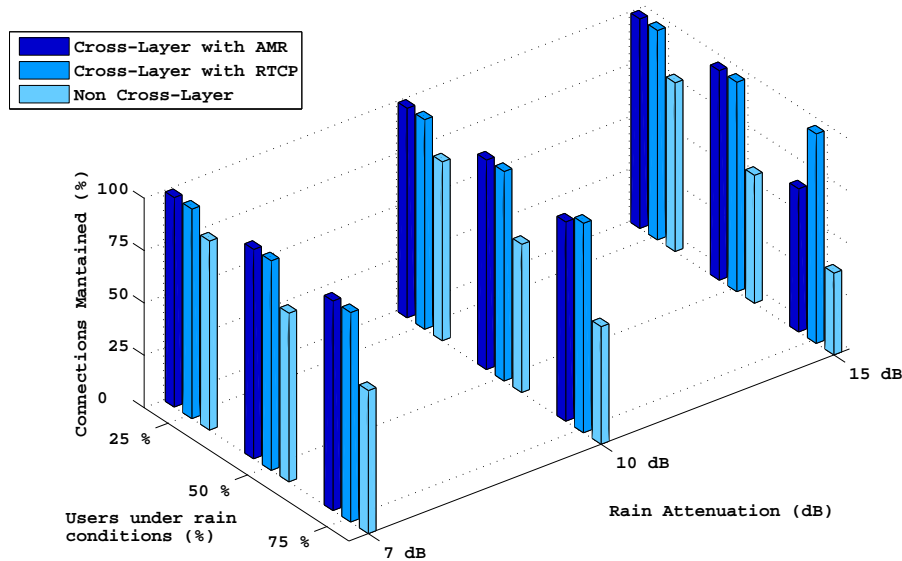


Figure 6.13: Percentage of VoIP Connections that can be maintained using cross-layer VoIP load control for different peak attenuations and percentage of users experiencing the attenuation.

and with better quality. It can therefore be concluded that even if AMR-WB suffers a higher percentage of capacity dropping, the AMR-WB system allows a larger number of simultaneous connections with and without load balancing.

Figure 6.14 and 6.15 show the performance comparison of the proposed cross-layer load adaptation when the rural area is located in temperate and tropical areas, when 50 % are affected by rain conditions. In this case time evolution of aggregated traffic adaptation is shown. It is interesting to see in Figure 6.14 the effect of the cross-layer load control for a rain event of temperate areas with a peak of 15 dB, which can be exceeded during the 0.01% of time. As expected, using both cross-layer approaches, capacity remains constant. In Section 3.1.1.1.1, we detailed that rain attenuations in Brazil can reach 22 dB 0.05 % of the time. We have tested our cross-layer mechanism for this worst case scenario, and the results are shown in Figure 6.15. It is important to note that the strong capacity drop is caused by the fact that the dynamic range of DVB-S2 ACM is not sufficient to adapt the transmission rate for the target PER of 10^{-7} ; thus, connections are dropped. It is evident that such a stringent target PER could be relaxed for VoIP, and no drops would happen. Such strong capacity drop is beyond the scope of this cross-layer design, however, we aim at solving it in the design of Section 6.3.1, by means of higher layer protection. We have shown that our cross-layer load adaptation across layers and networks behaves very well for any type of attenuation, allowing the number of active connections to be maintained independent of channel conditions even in very adverse conditions in remote areas.

Extended results of our cross-layer design are published in [104, 105], as part of a collaboration within the European Commission BRASIL project (<http://www.dvb-brasil.org/>). The work focuses on an OPNET Software implementation used to model the cross-layer

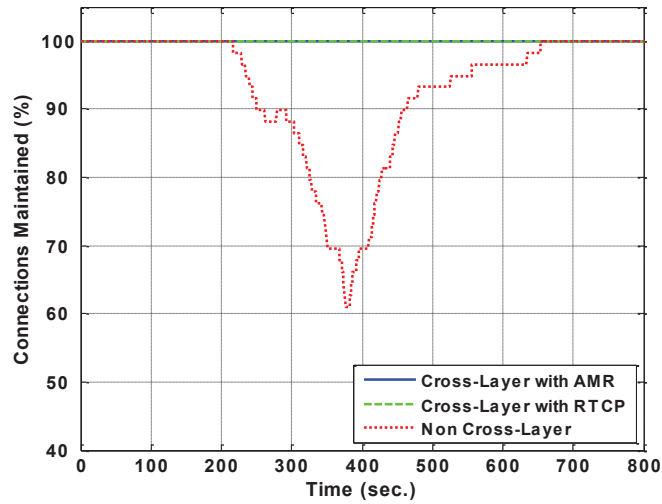


Figure 6.14: Time evolution of percentage of VoIP connections that can be maintained using cross-layer VoIP load control for an attenuation of 15 dB in a temperate area. Comparison to the non-cross-layer scenario.

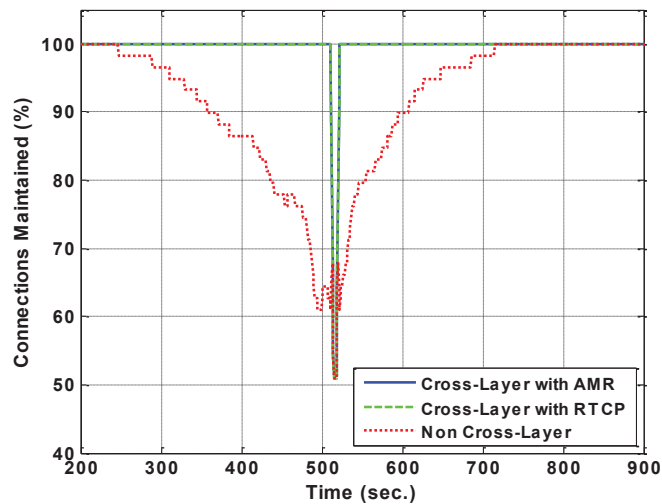


Figure 6.15: Time evolution of percentage of VoIP connections that can be maintained using cross-layer VoIP load control for an attenuation of 22 dB in a tropical area. Comparison to the non-cross-layer scenario.

design defined herein across the hybrid network.

6.3 Methodology: Adaptive Error Protection

6.3.1 XL Design 3: High layer-FEC protection and PHY Layer optimization

The demand for broadband connectivity in low-densely populated areas has given rise to a big interest in hybrid networks, especially those composed of a satellite backhaul and a wireless terrestrial network, as we have shown in the previous cross-layer design and in the hybrid architecture of Section 3.3.

In the previous design, we have shown that, for some rain events, the DVB-S2 ACM modes cannot cope with the deepest attenuations, exemplified in Figure 3.1, where the E_b/N_0 of the tropical rain is below the threshold of the minimum ACM mode, which is -1.5 dB for a 10^{-7} QEF FEC (Quasi Error Free Packet Error Rate). Herein, we propose to use (for the forward link) the Link Layer FEC (LL-FEC) that has been recently standardized in DVB-RCS+M [21] following the DVB policies of re-using already standard solutions.

However, since the maximum size of FEC frames is limited, and deep fading duration can be in the order of seconds, we additionally propose an Application Layer FEC solution (AL-FEC) like the one standardized for DVB-H (Handheld)[106] and DVB-SH (DVB Satellite to Handheld) [31], where FEC frames are allowed to be larger. Since the error protection we propose causes a noticeable delay, it cannot be used for real-time applications, which is why we focus on video streaming applications. We also demonstrate that, during the rain event, the delay in the ACM reaction to fade changes can affect the QoS of a multicast video streaming. Specifically, there is a noticeable delay (RTT of about $2 \times 250ms$ plus DVB super-frame length) from the moment an RCS Terrestrial (RCST) detects an ongoing fade event to when it receives from the NCC the transmission with the modulation and coding rates appropriate to cope with the increased atmospheric attenuation. We verify that the LL-FEC, tuning of super-frame (SF) length and a shifted threshold are useful to compensate for the packet errors introduced by the ACM switching delay.

In the current design, we aim at improving the satellite subnetwork performance of the hybrid architecture in severe tropical conditions. In order to do so, we focus on a DVB-S2/RCS system with adaptive physical layer for the satellite link and a Wi-Fi network for the indoor terrestrial wireless link. For the Amazon scenario considered here, the satellite link is the bottleneck of the system in terms of capacity, because the users in the indoor Wi-Fi network are fixed; consequently, the terrestrial link condition has no impact on the overall conditions of the hybrid channel. Figure 6.16 depicts the considered scenario, which reproduces the hybrid network used to give service to small villages along the Amazon River in Section 3.3. The video stream is transmitted by the video stream server to the satellite transmitter, which broadcasts it to the satellite network. Each satellite receiver broadcasts the video stream to a terrestrial indoor wireless network.

The combination of these two technologies is useful to enable connectivity in non-urban areas. It is one of the solutions that are envisaged in the context of the BRASIL

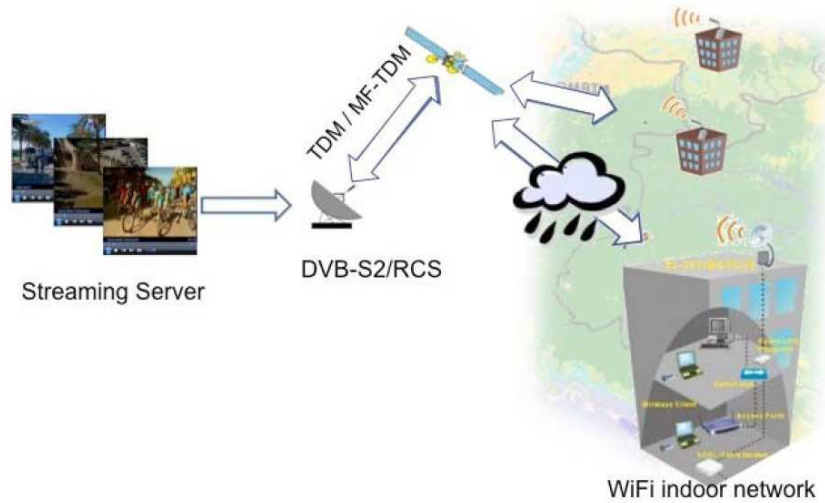


Figure 6.16: Hybrid Architecture

project.

6.3.1.1 Contributions

We can summarize the contributions of this cross-layer design as follows:

- Providing different cost-efficient solutions for the channel impairments in tropical areas.
- We focus on different Fading Mitigation Techniques (FMT) in order to avoid the QoS reduction of video transmission in the hybrid network.
- In order to improve accuracy in the results obtained, we will take into account the allocation delay and an extensive set of rain events (including time and space diversity).
- The design includes a multi-layer scheme protection, using LL-FEC and AL-FEC at higher layers, and optimizing the ACM thresholds and super-frame length at the PHY layer.
- The main parameters evaluated for this purpose are the packet error rate, the delay, and the bandwidth overhead due to the use of the proposed techniques. The received video will be evaluated by using the PSNR (Peak Signal to Noise Ratio) metric.

6.3.1.2 Fading Mitigation Techniques

To compensate for signal outages, the application of erasure-based FEC codes extending the time diversity is a well-known method. Generally, the larger the time-diversity, the higher the efficiency of the system, as signal outages can be averaged out more easily. Two codes have been proposed for high layer error correction: Reed-Solomon (RS) and Raptor codes. RS codes had commonly been used if only small dimension block codes are

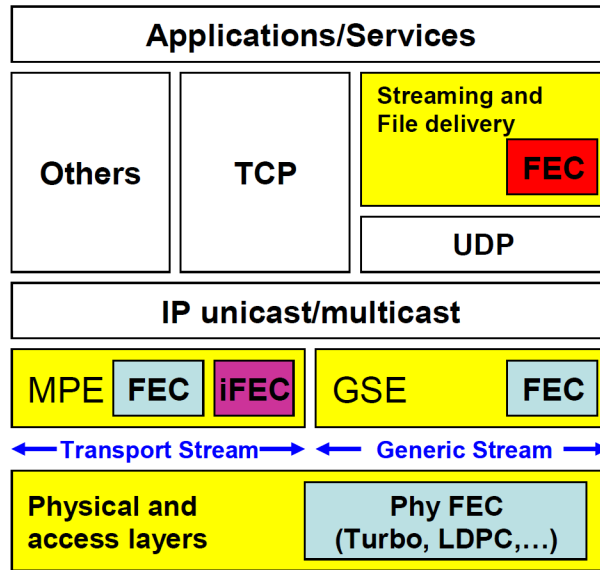


Figure 6.17: FEC location in the protocol stack

required, and are applied in the first generation of DVB family of standards, e.g. in DVB-C, DVB-S or DVB-H. On the other hand, Raptor Codes have been proposed lately and introduced into standards. In contrast to RS codes, they provide more flexibility, larger blocks of coded data, and lower decoding complexity. Raptor codes have therefore been adopted in the latest DVB standards, e.g. within DVB-H for file delivery and DVB-IPTV (IP Television).

Figure 6.17 shows a high-level protocol stack highlighting the usage of FEC codes for DVB services over IP-based networks [107]. In the physical layer we can find the well known Turbo codes and LDPC codes used in the ACM scheme. At the data link layer, DVB has adopted a LL-FEC (MPE-FEC) in DVB-H, which has been later standardized for DVB-RCS+M. However, during the DVB-SH standardization activities, it was recognized that for satellite-to-handheld services, the MPE-FEC is not sufficient. Therefore, it was decided to specify a multi-burst link layer FEC framework referred to as Inter-Burst FEC (IFEC). Finally, an AL-FEC might be applied at the Application or Transport layers as in 3GPP's Multimedia Broadcasting/Multicast Services (MBMS) or Internet Protocol Datacasting (IPDC) file delivery over DVB-H based on the Raptor codes.

As an alternative or a complement to the LL-FEC, we also consider a threshold-based algorithm, where the NCC changes the ACM mode when the standardized threshold of DVB-S2 is reached, with a given margin (in dB). Both alternatives will be compared in terms of PER and available bandwidth.

6.3.1.2.1 LL-FEC

DVB has adopted a LL-FEC in DVB-H at the data link layer (MPE Layer) referred to as MPE-FEC. At the time when DVB-H was specified, only RS codes were available, and therefore, the MPE-FEC is based on RS codes. FEC operations are performed in the DVB-H link layer as illustrated in Figure 6.18 and the processes are fully defined in [20].

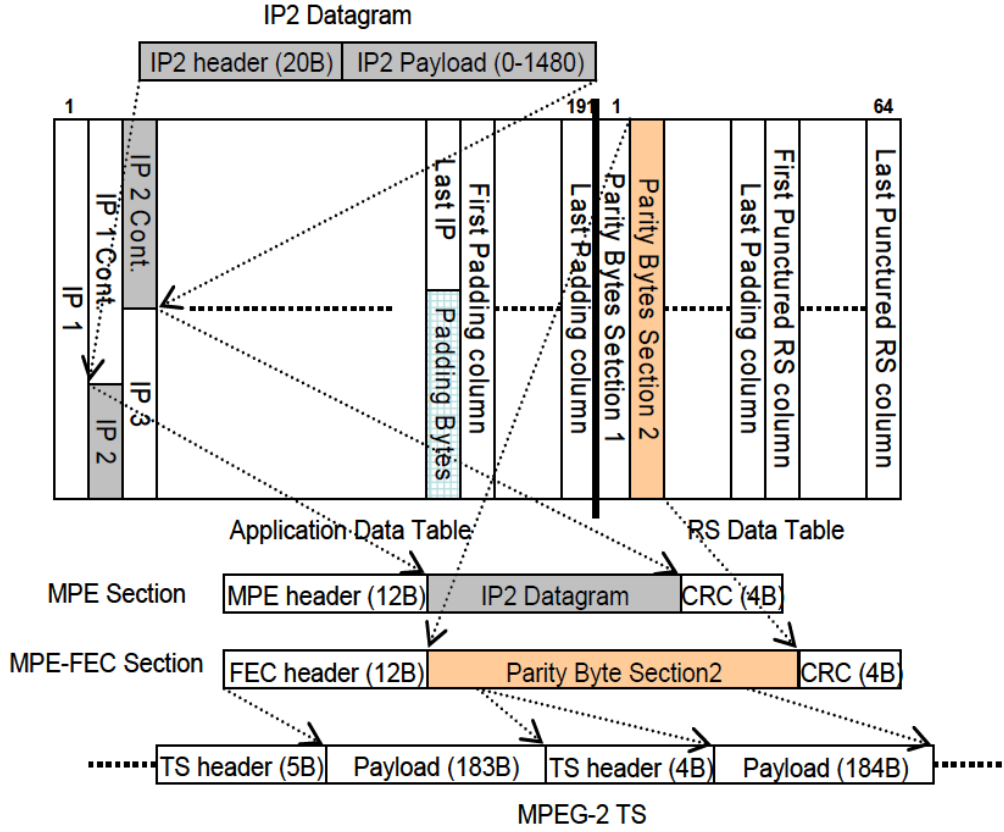


Figure 6.18: MPE-FEC Frame and the MPE encapsulation process

For an ideal memory-less erasure channel with symbol erasure probability ϵ , P_e is the residual PER of $RS(n, k)$ code and can be computed as:

$$P_e \cong \epsilon \left(1 - \sum_{i=1}^{n-k} \binom{n}{i} \epsilon^i (1 - \epsilon)^{n-i} \right) \quad (6.13)$$

For $RS(255, 191)$ in DVB-H, $n = 255$ and $k = 191$. However, the code, which can also be punctured and shortened such that any k can be used with $0 < k \leq 191$ and $n = k + 255 - 191$.

Our purpose is not only to avoid the loss of packets due to the ACM delay reaction when the attenuation of the rain event is changing, but also when the ACM modes cannot cope with deepest attenuations. However, one of the main drawbacks when using this framework for DVB-S2 is the size of the MPE-FEC frame, which is not big enough to protect against long burst errors since the number of address signaling bits for the Application Data Table (ADT) and RS data table is only 18 bits [20].

6.3.1.2.2 AL-FEC Framework

In order to protect from the longest error bursts (deep attenuations can be in the order of seconds), which LL-FEC cannot recover, we propose to use an AL-FEC. It can be

activated before the rain attenuation exceeds the threshold of the lowest ACM mode, and thus data can be protected during heaviest conditions, otherwise all data will be lost. Note that AL-FEC provides the same error correction capability as LL-FEC for a given coding rate, but the key is that it can provide protection across several error bursts, rather than across a single burst as with MPE-FEC.

6.3.1.2.3 Shifted Threshold

The physical layer selector of the RCST (see Annex E in [18]) estimates the optimal physical layer configuration (i.e. the ACM transmission mode) to be used in each channel condition. The Channel State Information (CSI) is sent to the NCC, which switches to the adequate ACM mode in order to protect data efficiently according to the channel variations. For fixed RCST, as is the case of our scenario, weather conditions are the main channel impairment.

In order to guarantee reliable and efficient system operation, the ACM selector shall account for: a) SNIR estimation errors, b) propagation delay and possible channel variations occurring between the user request and the application of the updated ACM mode.

According to [108], we assume that the RCST can send the CSI only once per super-frame, and moreover, we consider that measurements are averaged during whole super-frame in order to filter out the highest frequencies of the attenuation process. Specifically, the terminal takes measurement samples during super-frame i , their average is sent at the beginning of super-frame $i + 1$, and NCC can use the correct modcod in super-frame $i + 3$. The delay can thus be computed as $RTT + 2 \cdot SF_{dur}$, where SF_{dur} is the super-frame length duration and RTT , the Round Trip Time (about 500 ms). SNIR measures are exponentially averaged at each time t_i using a decay constant of 0.5. The number of measurements averaged depends on SF_{dur} and the CSI sampling time t_s , e.g. if $SF_{dur} = 800\text{ms}$ and $t_s = 250\text{ms}$, the RCST can compute and average 3 attenuation measurements.

Our goal is to develop an algorithm for computing the shifted threshold that produces the best quality in the video transmission. The algorithm consists of shifting by a given margin (in dB) the threshold of the standardized ACM modes of DVB-S2. Two are the main purposes of shifting the threshold, as specified in the standard: a) to avoid loss of data when attenuation is increasing, since the data is not correctly protected, b) to avoid extra payload and loss of efficiency when attenuation is decreasing, since the data is overprotected. Although the second point will be also studied in terms of the bandwidth efficiency, it is out of the scope of this design, and our algorithm will focus on minimizing the lost data. The shifted threshold (thr_s) is offset against the original threshold (T_{SNIR}^m) by a quantity that is generally dependent on the estimation error and the slope of the rain event. In particular, the shape of the new threshold of the ACM mode m can be stated as follows:

$$\tilde{T}_{SNIR}^m(t) = T_{SNIR}^m(t) + thr_s(t, \sigma_e, r_{slope}) \quad (6.14)$$

Although in (6.14), thr_s might depend on the time (t), the estimation error (σ_e) and the slope of the rain event (r_{slope}), in this design, we will focus on a fixed shifted threshold

for different types of tropical rain events. The results we obtain will guide us in the design and the implementation of an adaptive algorithm that we will consider for future works.

6.3.1.3 Results

6.3.1.3.1 Simulation Platform

A Multilayer Protection simulation platform has been developed in order to quickly assess the performance of different parameter configurations without repeating the time-consuming physical layer simulations. Given that this performance assessment entails many layers, specifically from the physical to the network/transport/application layers of the protocol stack, a modular approach has been considered. The ACM switch Physical-Layer module depicted in the upper-right corner of Figure 6.19 generates the satellite channel attenuation time series and interfaces with the Link Layer simulator. The PHY settings module chooses the shifted threshold and the super-frame length. The simulator takes streams of IP packets as input and applies the MPE-FEC encoding technique, generating an MPEG-2 TS by encapsulating Multi-Protocol Encapsulation (MPE) sections and MPE-FEC sections. Then, MPE-sections are encapsulated in different BBFrames as specified in [18]. The output of the physical-layer simulator is used to mark the BBFrames as correctly received or being erroneous. After BBFrame decoding, the MPE-FEC decoding process is applied by reconstructing columns of the FEC matrix applying the correction capabilities of different FEC codes (RS or Raptor). Next, the sequence of IP packets affected by the unreliable columns is obtained and the PER at the IP level is computed. Finally, PSNR is computed by comparing both videos; original and corrupted; when the decoder is incapable of decoding a video frame, the previous decoded frame is frozen until the decoder can recover from the stream corruption – normally at the next I frame starting a new GOP (Group Of Pictures).

The introduced LL-FEC framework allows a significant variability in terms of parameter settings. The amount of data (bits), that can be protected with target delay τ_{ll} , can be computed as $S_{protect} = \tau_{ll} B_s M r_{phy} r_{ll}$. Where B_s is the symbol rate, M is the modulation order, r_{phy} is the physical layer coding rate, and r_{ll} is the LL-FEC coding rate. Note that M and r_{phy} depend on the selected ACM.

6.3.1.3.2 Parameters and Results

In order to get a realistic estimation of the system’s efficiency, we consider a significant amount of measured fading events at 12 GHz K_u Band, which have been selected from 1 month measurements at two different locations in the Amazon area (Tabatinga and Caxias, favourable and unfavourable case respectively). See Table 3.1 for satellite and area information.

A data rate of 530 kb/s has been chosen to study the performance of the video streaming application. First of all, in Figure 6.20, we evaluate performance in terms of PER versus the threshold offset, for two different locations in rain conditions: the Tabatinga area with maximum attenuation att_{max} of 16 dB and the Caxias area with att_{max} 9 dB, plus a quasi-clear sky case with att_{max} around 2-3 dB. No MPE-FEC is used for this simulation. We studied the performance for two SF_{dur} values. As expected, the PER is

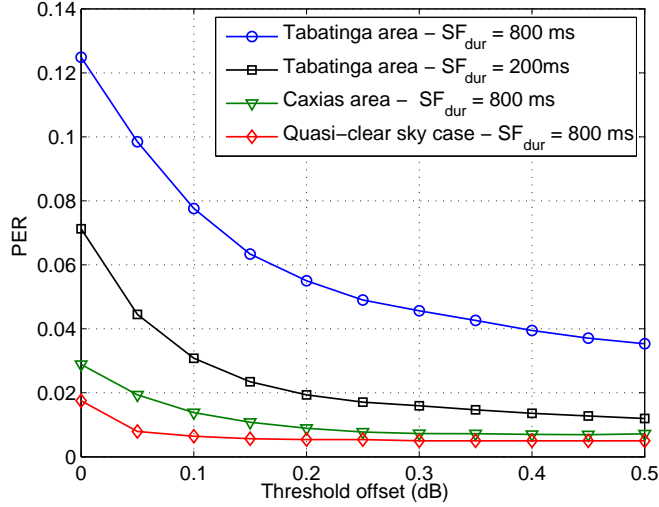


Figure 6.20: PER as a function of thr_s (dB) for several Amazon areas (rain and clear sky conditions), and different super-frame length durations

strongly affected by both the SF_{dur} and the location. In particular, decreasing the SF_{dur} , the ACM delay decreases and consequently also the PER, without significantly increasing the bandwidth on the forward link. By increasing the threshold offset, the PER is reduced: we observe more than 50% reduction for 0.2 dB. Note that when

Let us take the Tabatinga conditions with $SF_{dur} = 200ms$ from Figure 6.20 as a reference scenario in order to study the higher layer FEC protection. In that case, it can be stressed that the estimated PER is bigger than 7% when the proposed techniques are not applied, i.e. intervention threshold $thr_s = 0$ and no LL/AL-FEC. The Tabatinga reference scenario is compared to the multi-layer protection design (high layer FEC and threshold offset) in Figure 6.21. We show that a FEC coding rate of $3/4$ and a target delay of 2s are not enough to significantly reduce the packet loss for low values of thr_s , and we need lower code rates. For the lowest-rate code ($r_u = 1/4$), we verify that LL-FEC can significantly reduce the corrupted bits due to the ACM switched delay. However, all results seem to tend to a residual packet loss (1.5%), which cannot be recovered. In fact, this packet loss value is due to the deepest part of the fading event, in particular, when SNIR reception conditions experimented by satellite terminals are too low to be protected by standardized ACM modes. For bigger protection delay values of τ_u , the PER is reduced, however, the maximum delay allowed for the LL-FEC settings is 4 sec. due to the address signaling bits (limited to 18), as explained previously. Thus, we propose an AL-FEC, by using a modified LL-FEC platform. We have increased the τ_u delay, and as Figure 6.21 shows, the PER has been reduced to 0.1% ($thr_s = 0.5$). Note that the use of bigger protection delays leads to better results at AL-FEC. We conclude that AL-FEC should be introduced to mitigate this type of tropical channel conditions since it can provide protection for time spans of several seconds.

In order to make efficient use of bandwidth, we also investigate the trade-off between threshold offset and bandwidth efficiency in Figure 6.22. A high threshold offset reduces the efficiency when the ACM mode is switched to a lower one (more protection) because

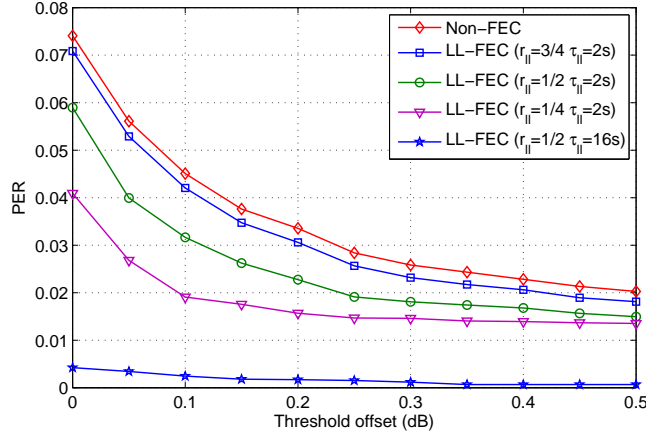


Figure 6.21: PER as a function of thr_s (dB) and for different FEC code rates using RS ($SF_{dur} = 200$ ms)

the switch happens earlier than due, i.e. we protect data more than the channel state requires when attenuation is increasing. On the other hand, as mentioned in Section 6.3.1.2.3, when attenuation is decreasing, the delay in changing to a more bandwidth efficient ACM mode introduces an overhead and thus reduces efficiency.

As expected, the threshold offset influences the performance of our technique. The type of rain event – smooth slope or with fluctuations – impacts on PER and efficiency, and a higher att_{max} does not necessary imply worse performance. For example, in Figure 6.22, the case of the rain event with $att_{max} = 15$ dB, which has big fluctuations and thus more ACM mode changes, performs worse than the rain event with $att_{max} = 17$ dB, which fluctuates less.

Our results generally indicate that a fixed threshold offset is not efficient in terms of bandwidth, e.g. for clear-sky conditions, there is a remarkable loss of efficiency. In other words, the threshold offset that is good for deep fading events, or events with big fluctuations, cannot be adequate for low fading or clear sky.

6.3.1.3.3 Video Quality Evaluation

In this section the performance of our recovery technique is shown in terms of video quality estimation, by using the traces produced by our simulation to corrupt an MPEG-2 TS (Transport Stream) and comparing it with an uncorrupted TS. We consider the well-known City video samples [109], by evaluating the video quality at the end-user versus the threshold offset. We consider H.264 [98] with a data rate of 530 kbits/s to study the performance of the video streaming application. To this end, we compare the video samples received by the end-user with the transmitted original video by using the standard PSNR metric [110]. The PSNR of a picture P is expressed as:

$$PSNR = 10 \log \frac{(2^n - 1)^2}{MSE(P)} \quad (6.15)$$

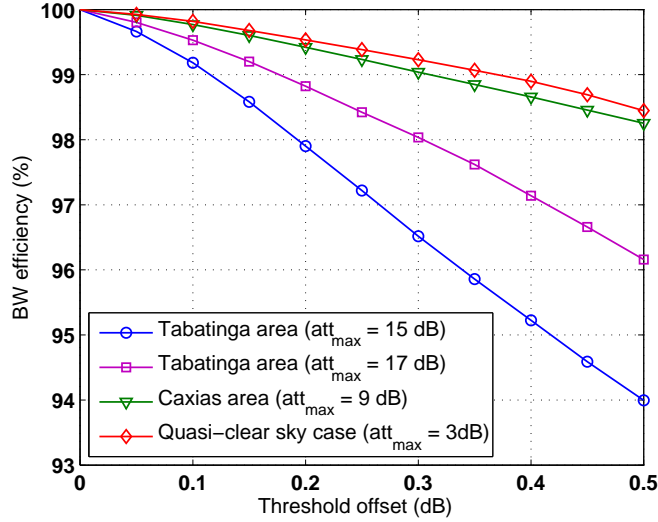


Figure 6.22: BW efficiency as a function of thr_s (dB) for several Amazon areas (rain and clear sky conditions)

where n is the number of bits used to represent the luminance of a pixel, and $MSE(P)$ is the Mean Squared Error:

$$MSE(P) = \frac{1}{N} \sum_{i=1}^N (P_i^o - P_i^c)^2 \quad (6.16)$$

where N is the number of pixels in the picture, while P^o and P^c are the original and the corrupted video pictures, respectively.

The PSNR can be a good indicator of the variation of the video quality when the content and codec are fixed across the test conditions [111]. Note that the reference value (initial value without lost data) of PSNR is 35 dB. Figure 6.23 shows the cumulative distribution of the PSNR of the corrupted video for different SF_{dur} values. Figure 6.24 shows a second-order statistics of PSNR, that is, the average length of time PSNR staying under a given threshold once it crosses it downwards. The figure shows that by fixing a $PSNR_{th}$ and decreasing the SF_{dur} length (this means increase the bandwidth overhead) the probability that the video quality is under $PSNR_{th}$ decreases. Both figures show that the super-frame length has a significant influence on the video quality.

Figure 6.25 depicts the received video quality versus the threshold offset for two different locations in rain conditions and a quasi-clearsky case. The error traces are the same used for Figure 6.20; the performance is evaluated for two SF_{dur} values. An interesting observation can be made by looking at the quasi-clearsky trace, whose PER is lower than other traces; however its PSNR is similar to the other traces' when SF_{dur} is 800 ms. One reason is that the severity of corruption events depend mainly on the number of error bursts, and less on the burst duration.

Figure 6.26 concentrates on the duration of corruption events, where a corruption event is defined as a sequence of video frames where some video corruption exists. This statistic is insensitive to the threshold offset: there is little difference between setting it

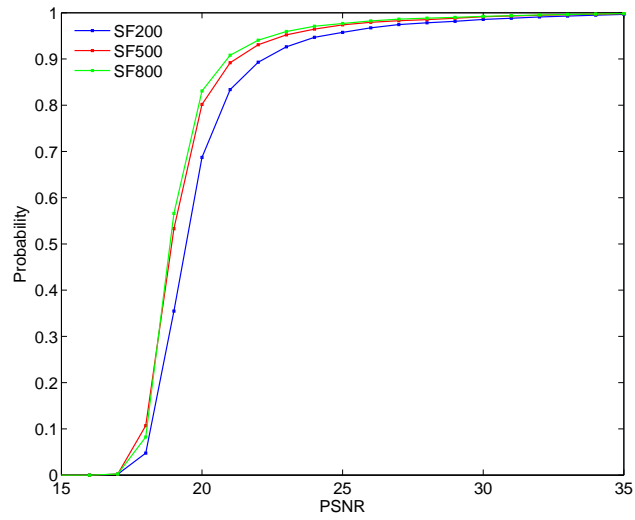


Figure 6.23: Cumulative distribution of PSNR.

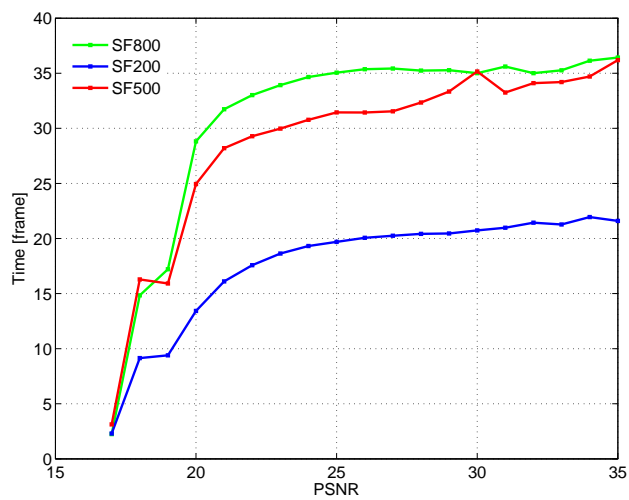


Figure 6.24: Average time length of PSNR persistence under a given threshold.

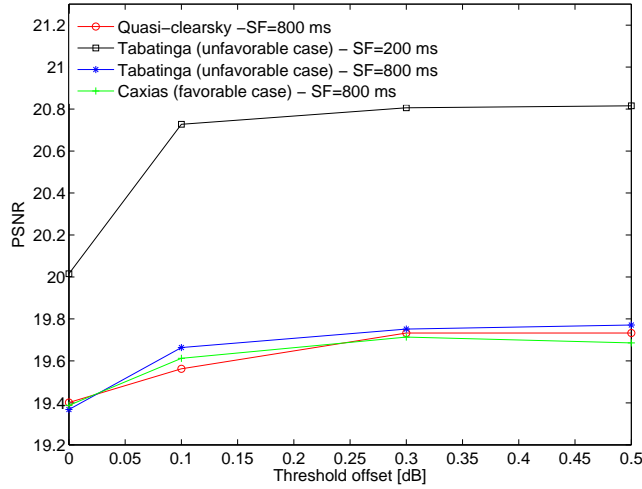


Figure 6.25: Mean PSNR versus threshold offset for different Amazon areas and rain types.

to 0.1 dB or 0.5 dB.

6.3.1.3.4 Cross-layer design

According to the results, we demonstrate that a shifted threshold and LL-FEC are useful to compensate for the packet errors introduced by the ACM switching delay, thus allowing the transmission of good video quality. In particular, an appropriate choice of the *shifted threshold* can achieve almost the same PER reduction as using the strongest high layer codes, but with a higher bandwidth efficiency gain. We also show that AL-FEC allow to protect data during deep fading events of several tens of seconds.

Results also show that a static threshold offset is suboptimal, which means that choosing it dynamically has the potential for significant improvement. The DVB-S2 technique of hysteresis is in principle superior to the simpler shifted offset and would merit a dedicated study. The superframe length also has a considerable influence on the video quality, and it should be jointly optimized with the threshold offset. Note that one of the drawbacks of reducing the SF_{dur} is the fact that the signaling overhead increases.

Therefore, given the observations of the multi-layer protection scheme study, we are able to propose an overview of a cross-layer design suitable for our hybrid scenario.

The three main parameters involved in the design are: PSNR, SF_{dur} and thr_s . In particular, according to the PSNR video quality measured at the Transport/APP layers, the cross-layer manager reduces or increases the SF_{dur} and thr_s in order to achieve the minimum requirements of video quality while maximizing the bandwidth efficiency.

6.3.2 XL Design 4: Unequal Error Protection (UEP) optimization

The previous cross-layer design has shown that for the particular case analyzed, when satellite transmission tries to adapt to channel conditions, the optimization of physical

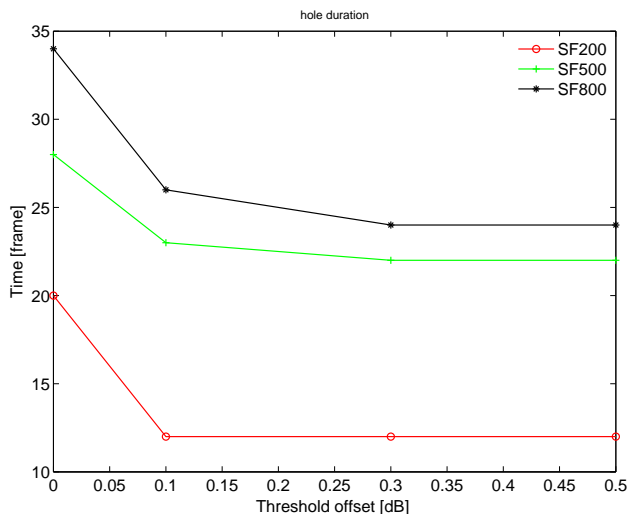


Figure 6.26: Mean duration of PSNR corruption depending on the thr_s and for different SF_{dur} .

layer parameters (depending on measured quality at higher layers) is not only more efficient but happens to have a similar impact on the quality reception than higher layer protection mechanisms. Although this is one of the typical ways to proceed for cross-layer designs based on data protection (i.e. to optimize the parameters of PHY/MAC layers according to higher layer measured quality), in some case, a common strategy of protection mechanisms from different layers (by means of cross-layer interaction) is more suitable, it is able to improve the performance of the protection scheme. Such joint strategy is shown next.

Mobile multimedia services are one of the most promising key applications for next wireless systems. However, the challenging dynamic nature of the mobile wireless channel has a great impact on the system design. Moreover, the intrinsic problem of multi-user channel diversity among a group of wireless receivers makes it difficult to determine an effective transmission strategy.

Big efforts using cross-layer designs have been carried out in order to comply with the stringent QoS requirements, as referenced in the literature of the previous designs addressed herein. However, cross-layer protection schemes has not yet been exploited in depth, and in particular those that follow a common strategy among different layer mechanisms.

6.3.2.1 Multicasting and Protection strategies

In the following design, we aim at implementing a multi-layer unequal protection scheme, which is based on a Physical-Transport-Application cross-layer design. The conclusions extracted from previous designs allow us to differentiate two contribution lines:

- A multicast strategy that takes advantage of:
 - A Scalable Video Coding (at APP layer) addressed in Section 6.2.1.
 - The Hierarchical Modulation (HM) at the PHY layer developed in Chapter 4.

- A protection strategy which includes:
 - Hierarchical Modulation at the PHY layer.
 - A new Unequal Error Protection (UEP) at the Transport layer

The multicasting properties of HM are shown in Chapter 4, where we demonstrate that a transmission scheme based on HM increases the overall user capacity of a wireless communication. On the other hand, SVC (see Chapter 5 and Section 6.2.1) is also understood as a Layered Video Multicasting thanks to its superposition properties. We will analyze how both solutions can perfectly match to follow a multicast strategy. Moreover, as previously argued in Chapter 4, DVB-S2 [18] includes guidelines about the compatibility of HM and the adaptive physical layer (i.e. ACM). Therefore, the implementation of the protection design presented next is highly compatible to most of the cross-layer designs presented in this dissertation.

Note that herein, multicasting is understood at system level instead of at networking level, meaning that receivers are organized in groups according to the channel conditions. Moreover, the transmission strategy can be optimized for every group.

In the second contribution line, the two techniques are jointly optimized in order to enable recovering lost data in case the protection is performed separately. On the one hand, the decoding Bit Error Rate (BER) of the different HM levels can be modeled depending on the distance between symbols, as we will explain latter. On the other hand, unequal erasure protection codes at the transport layer turned out to be an efficient method to protect video data (generated at the APP layer) by exploiting their intrinsic properties. Both techniques are jointly optimized in order to enable recovering lost data in case the protection is performed separately. We show that the cross-layer design proposed herein outperforms the performance of hierarchical modulation and unequal erasure codes taken independently.

6.3.2.1.1 HM, UEP and SVC Introduction

Included in several standards, such as DVB-H, and the recently standardized DVB-SH and WiMAX (IEEE 802.16), HM has demonstrated to be an efficient transmission mode achieving better system performance and customer experience than other solutions. Note that HM is known as the practical implementation of the Superposition Coding scheme. One major limitation of hierarchical modulation is that it has been standardized only for a two layer scheme, which reduces the diversity of the transmission (only two quality levels). In [112], it is shown that the 2-level SC achieves part of the throughput gain in a quasi-static Rayleigh Channel, however, it might vary depending on channel conditions. Thus, in order to compensate this low granularity at the physical layer, we take advantage of the scalable video codes at higher layers.

At the Application layer, we consider SVC with Fine Granularity Scalability (FGS) for video streaming services, where it increases the flexibility [113]. With FGS coding, the video is encoded into a base layer and one enhancement layer. Similar to conventional SVC, the base layer must be received completely in order to decode and display basic quality video. However, in contrast to conventional SVC, which requires the reception of complete enhancement layers to improve the basic video quality, with FGS coding the

enhancement layer stream can be cut anywhere at the granularity of bits before transmission. The received part of the FGS enhancement layer stream can be successfully decoded and improves the basic video quality. With the fine granularity property, FGS encoded videos can flexibly adapt to changes in the available bandwidth of wireless networks.

In addition, we consider erasure codes at the Transport layer, which allows recovering lost packets thanks to redundant information. These codes can be adapted to the data properties by allocating more protection to specific parts of the data. Several works have addressed this issue [114]-[115], based on Priority Encoding Transmission (PET), which allows the sender to decompose the data into classes of given importance. However we focus on a data protection scheme that integrates the dependencies at the packet-level by keeping the data dependencies produced by the source. In particular, we focus on Dependency-Aware Unequal Erasure Protection (DA-UEP) codes studied in [116], where it is presented a different approach for protecting multi-classes data by generating specific redundancy according to existing dependencies in MPEG4 streams. This solution integrates the data dependencies at data-level, in particular within the construction of the Cauchy generator matrix, which is used to encode the defined erasure code.

Thus, the cross-layer protection scheme proposed herein will take profit not only of the granularity available at APP level with FGS, but also at PHY layer.

6.3.2.2 Hierarchical Modulation

More than thirty years ago, [32] showed that one strategy to guarantee basic communications in all conditions is to divide the transmitted information into two or more classes, and to give every class a different degree of protection, which is the principle of SC. The goal is that the most important information (basic) can be recovered by all receivers, while the less important (refinement) can only be recovered by best users.

One of the practical ways investigated to perform the principle of SC is based on Hierarchical Modulation, where two separate data streams are modulated onto a single stream. One stream, called high priority stream is embedded within a low priority stream. Receivers in good reception conditions can receive both streams, while those with poorer reception conditions may only receive the high priority stream. E.g., in DVB-SH standard, the hierarchical system maps the data onto the 16QAM in such a way that there is effectively a QPSK stream (high priority) buried within the 16QAM stream (low priority). The QPSK/64QAM is another common hierarchical scheme used in DVB-T.

Although the concept of this design might be applied to all types of Hierarchical Modulation, we will focus on a typical QPSK/16QAM transmission scheme (as standardized in DVB-SH [31]) as depicted in Figure 6.27. We will focus on the BER to evaluate the performance of this scheme and also to detect the parameters to be optimized in the physical layer. Several approximate BER expression are available, such as in [117], but they underestimates BER at low SNR, and for fading channels, the performance is severely degraded. Therefore, we will use the exact expressions from [118].

From the distances d_1 and d_2 defined in Figure 6.27, we can obtain the average energy per symbol (E_s), given by:

$$E_s = 2d_1^2 + 2d_2^2 \quad (6.17)$$

where the first term represents the average energy per symbol of the QPSK modulation

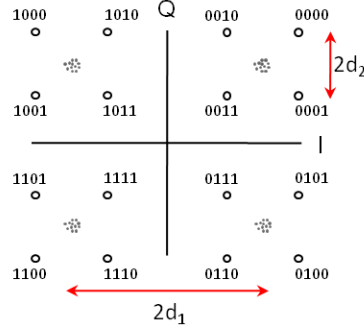


Figure 6.27: QPSK/16QAM Hierarchical Modulation Scheme, where noise emulates the symbols received by QPSK users

with symbols separated by $2d_1$. The bit error probability for users decoding QPSK and 16QAM modulations in a hierarchical system are defined as (6.18) and (6.19) respectively.

$$BER_{QPSK} = \frac{1}{4} (U(1, -1) + U(1, 1)) \quad (6.18)$$

$$BER_{16QAM} = \frac{1}{4} (2U(0, 1) + U(2, -1) - U(2, 1)) \quad (6.19)$$

where

$$U(a, b) = \text{erfc} \left(\frac{ad_1 + bd_2}{\sqrt{N_o}} \right) \quad (6.20)$$

Several conclusions can be extracted from the equations above. First of all, the first term of (6.18) and (6.19) are known as the BER approximations of each type of receiver. Considering only the first term of (6.18) (i.e. the approximation), we can stress that the performance of QPSK users in a hierarchical system is degraded in terms of BER by d_2 , if we compare it with the BER when they are placed in a non-hierarchical system (6.21).

$$BER_{QPSK'} = \frac{1}{2} \left(\text{erfc} \sqrt{\frac{E_s}{2N_o}} \right) = \frac{1}{2} \left(\text{erfc} \frac{d_1}{\sqrt{N_o}} \right) \quad (6.21)$$

d_1 and d_2 are the parameters to be considered when designing a hierarchical modulation scheme. More interesting is the ratio $\lambda = d_2/d_1$, which allows us to characterize the system for a given E_s , which is hierarchical only when $0 < \lambda \leq 1/2$. For high values of λ , 16QAM receivers experience better BER performance, opposite to QPSK users, which are clearly affected by the hierarchical distortion. Note that if $\lambda = 1/2$, it is the uniform hierarchical modulation, and if $\lambda = 0$, it is the fixed modulation QPSK system.

Therefore, although the hierarchical scheme can increase the system capacity, it degrades QPSK users in terms of bit error probability, and thus it is necessary to find a trade-off between PER and overall throughput capacity.

Let us highlight that using the BER expressions in [118], we can easily extend 6.18 and 6.19 equations to Rician channels.

6.3.2.3 DA-UEP for Transport Layer

Video data have specific properties and constraints that must be taken into account in the design of reliability systems. One example of these properties is that video decoder can support a low packet erasure rate (up to 5%) by implementing error concealment mechanisms. On the other hand, these streams often require constraints in terms of delay, *e. g.* for video conferencing or video streaming.

The internal structure of the stream generated by a video encoder has particular properties leading to unequal importance of packets carrying the video frames and to dependencies between these packets.

A classical type of such dependency is generated by FGS encoder [113] (see Section 6.3.2.1.1) where the base layer must be first received and decoded in order to use enhancement layers. A second type of dependencies occurs in "natural" video streams (*e. g.* MPEG/H.264) composed of I-frames (intra-coded pictures) and P and B inter-frames. The two last types of frames are encoded from the previous P or I frame, and from the previous and the next frame respectively.

Dependency-Aware unequal erasure protection (DA-UEP) introduced in [116] can be applied to any type of data containing dependencies between the data units to be protected. For the sake of simplicity, we present this system only for streams of intra and inter-frames.

DA-UEP codes are block codes that aim at protecting a set of K data units by generating $N - K$ redundancy packets. Its originality is to integrate the intrinsic dependency relationships between the data units in the construction of the redundancy packets. This integration is done by applying a simple set of rules to each generated redundancy packet. These rules can be expressed as follows. A redundancy packet protecting a packet belonging to the frame t must:

1. protect all the packets belonging to this frame
2. protect all the frames (i.e. all the packets belonging to those frames) on which this frame depends.

Considering the dependency relationships that hold within a Group of Pictures (GOP) containing one I-frame and several P and B inter-frames, and by applying the above defined rules to these data, it is possible to differentiate several kinds of redundancy packets:

- r_I : Number of packets protecting the I frame.
- r_P : Number of packets protecting a P frame and all the precedent P frames until the first I frame.
- r_B : Number of packets protecting a B frame, the corresponding pair of reference frames and all frames on which they depend. With r_B equals to $N - K - r_I - r_P$.

A simple way of implementation is to encode data on a GOP basis, i.e. it is assigned a certain amount of redundant data to each GOP. Moreover, the shape of the generator

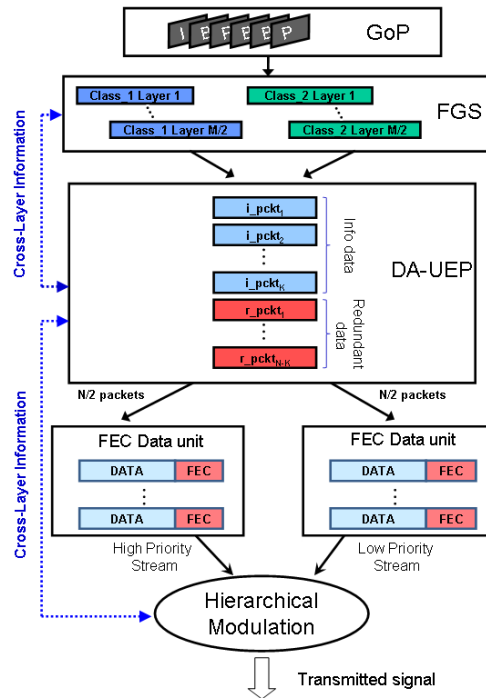


Figure 6.28: Cross-layer Unequal Protection Scheme

matrix is closely related to the size of each frame of the GOP. Hence, by knowing that the size of a frame varies from GOP to GOP, the use of this code with real video data requires to dynamically build the generator matrix for each GOP. The first step of this process consists in determining $N - K$, r_I and r_P .

Once these variables are set or calculated, the encoder first builds a classical systematic Maximum Distance Separable (MDS) generator matrix. For each column corresponding to a redundancy packet, it computes the information data packet that must not be protected by this redundancy repair packet and put to *zero* the coefficient of the corresponding row. At the receiver side, like for most of erasure codes, the decoding simply performs an inversion of the sub-matrix of the generator matrix corresponding to the received packets and multiplies the obtained matrix by the received packets.

The analytical evaluation of the performance of this code can be done by considering this code as several nested MDS codes. This analysis is beyond the scope of this dissertation.

6.3.2.4 Cross-layer Architecture Design

Our objective is to analyze the interest of jointly optimizing unequal protection schemes implemented at different layers. We then integrate all the main reliability mechanisms used at the different layers: hierarchical modulation and error correcting code at the physical layer, erasure code at the transport layer and video FGS coding at the application layer. In order to understand and to evaluate the interactions between these mechanisms, we intentionally choose generic instances of these mechanisms whose performance can be easily modeled and integrated in an optimization system. Furthermore, we choose to

not consider additional layering mechanisms (segmentation/re-assembly, protocol header, etc...), which are specific of each protocol stack and that can modify the evaluation of the set of reliability mechanisms.

Figure 6.28 depicts the considered cross-layer architecture. A Group of Pictures (GOP) containing the three frame types (I, P and B) is encoded into different scalable video layers following FGS coding, and considering both SNR and temporal scalability. Two types of layers are differentiated; Class 1 and Class 2. Class 1 layers contain the base layers that must be received for decoding basic video quality, however not necessary meaning that all Class 1 layers are base layers. Then, Class 2 layers are used to improve the video quality. Each type of layer are jointly treated in the DA-UEP coding block, and decomposed in K packets of the same size. Next, the unequal protection scheme is applied in order to create the $N - K$ redundant packets according to the dependencies between packets, their importance and the code rate specified.

Encapsulation protocols are not considered in the Network/Link layers. In order to simplify, the link between DA-UEP and the Modulator is seen as a data unit, which includes an erasure code representing the physical layer forward error correction. The $PER = 1 - P_c$ of the data unit is computed considering randomly error bits, therefore, P_c (i.e. packets correctly received) can be computed using the binomial (6.22), and simplified using the normal approximation in (6.23):

$$P_c = \sum_{m=0}^t C_m^n p^m (1-p)^{n-m} \quad (6.22)$$

$$P_c \approx \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{t + 0.5 - np}{\sqrt{2np(1-p)}} \right) \right) \quad (6.23)$$

where n is the total number of bits, t is the number of errors that the code is capable of correcting ($t = (n - k)/2$ in case of MDS codes), $n - k$ is the number of redundant bits and k is the size of info bits. Note that $n \geq 30$ and $np(1-p) \geq 5$ conditions must be accomplished in order to allow the approximation.

Finally, $N/2$ packets go to each stream, each one containing the original Class 1 or Class 2 source files, together with redundant packets built in the DA-UEP (using dependencies between both classes). The streams are modulated according to the HM QPSK/16QAM scheme, one as High Priority (HP) stream (or basic information) and the other as Low Priority (LP) stream (or enhancement information). Note that since we optimize the transmission for different channel distribution features (multipath, line of sight, shadowing, etc), it can be seen not only as a receiver affected by different channel conditions, but also as a group of multicast users, each of one served with a different transmission strategy.

In the Figure 6.28, it is also highlighted the cross-layer information used in order to proceed with the joint optimization. In our approach, a global performance parameter (video distortion) is selected and the intervening protocols in PHY/Transport/APP layers are reviewed in order to identify how their behavior can be improved so as the distortion is minimized. In particular, this approach would design an efficient information flow among layers from the APP level down to the PHY level. It can be essentially seen as an application-centric approach, where the APP layer optimizes the lower layer parameters.

d_1 and d_2 are sent to the PHY layer, the redundancy to be allocated to each video layer is used in the DA-UEP block, and the optimal video layer length distribution is computed at the FGS block of the APP layer.

6.3.2.5 Optimization: Maximize Distortion

6.3.2.5.1 Formulation Framework

The analytical evaluation of the different reliability mechanisms is presented in the last part of this section. To evaluate their interactions, we have expressed the set of these analytical expressions as an optimization problem which was implemented in Matlab. The objective of this implementation is to evaluate the optimal configuration of the parameters of the unequal protection schemes in order to minimize the video distortion (D_t) for a given E_s/N_0 and different channel distributions (Rayleigh, Rice, Loo, etc.) at the PHY layer. The distortion depends on the probability of decoding up to the layer i by DA-UEP (f_i), and on the video distortion achieved ($D(R_i)$) if we decode up to this layer. Thus, D_t can be defined as:

$$D_t = \sum_{i=1}^M f_i(\vec{k}, \vec{r}, P_c^{HP}(\lambda), P_c^{LP}(\lambda)) \times D(R_i) \quad (6.24)$$

where f_i depends on the number of information packets \vec{k} and the redundancy added to each one \vec{r} . It is also affected by the probability of correct packets received of HP and LP streams (P_c^{HP} and P_c^{LP} respectively), which are obtained from (6.23). At the same time, P_c depends on the BER of each receiver type; (6.18) and (6.19), which depend on λ and the channel conditions.

6.3.2.5.2 NUM: non-convex problem

Therefore, considering the variable parameters \vec{r} and λ , together with their defined constraints in HM and DA-UEP sections, the optimization problem can be formulated as follows:

$$\begin{aligned} & \arg \min_{\vec{r}, \lambda} D_t(\vec{r}, \lambda) \\ & s.t. \quad \sum_{i=1}^M r_i = r_{max}, \quad 0 \leq \lambda \leq 0.5 \end{aligned} \quad (6.25)$$

Since it is a complex problem due to the non-convexity, and thus the solution is difficult to find, we propose an iterative solution based on Matlab optimization functions in order to optimally allocate the needed protection in the HM scheme and at DA-UEP. Moreover, in order to reduce complexity, we fix the FGS coding (video layers distribution) as explained next.

6.3.2.5.3 Parameters Definition

Following classical approximations of video distortion, we consider that the video rate distortion curve can be modeled by a decreasing exponential function of bitrate R whose

form is $D(R) = S_v \times \exp(-\alpha R)$, where S_v is the source variance set to 100 and α is a constant. We consider that the FGS video encoder produces a stream of 1 base layer and 3 enhancement layers with a maximum rate of $R_{max} = 384kbps$.

Considering the packets size equal to 500 bytes, the DA-UEP code then protects $K = 96$ data packets containing the video frames by generating $N - K = 32$ redundancy packets. Among the K data packets, $K/4$ packets comes from each layer. The first two layers and their associated redundancy packets are considered as the high priority stream and the two other layers are the low priority stream.

These two streams are protected (independently) at the physical layer by the same $[n = 2256, k = 1504]$ error-correcting code. The codeword length corresponds to the length of MPEG2-TS packets and allows recovering from 376 errors at most.

The high priority stream is then mapped onto the two most protected bits of the hierarchical modulation and the low priority stream is mapped onto the two least protected bits. Under these assumptions, the variable parameters of the system are the value of $\lambda = d_2/d_1$ characterizing the hierarchical modulation, and the number of redundancy packets allocated to each "layer" of DA-UEP.

An interesting point is that the set of possible parameters includes the cases "HM only" and "DA-UEP only". It follows that the resolution of this optimization problem allows evaluating the interest of jointly optimizing hierarchical modulation and unequal erasure protection compared to non cross-layer solutions.

6.3.2.6 Simulation Results

In this section, we analyze the performance of the cross-layer optimized design and the non-cross-layer designs (only DA-UEP or HM) for different parameters configurations and depending on channel conditions.

Figure 6.29 shows our goal by comparing the cross-layer joint optimization with the non-cross-layer optimizations, i.e. HM or DA-UEP independently optimized. In the non-cross-layer scheme, when HM is optimized, DA-UEP is set to equal protection code, and inversely, when DA-UEP is optimized, HM is set ($\lambda = 0.5$) to uniform hierarchical modulation. It can be observed that our cross-layer design obtains better results than the other designs in terms of video distortion, specially for lower values of E_s/N_0 , i.e. for worst channel conditions. In this case, the cross-layer solution outperforms DA-UEP up to 50% and HM up to 20%. As expected, when channel conditions are favourable, all optimizations obtain similar values. Note also that HM optimization performs clearly better than the DA-UEP solution, which obtains big distortion for lower E_s/N_0 values.

Figure 6.30 will be very useful in order to know the way the optimization is performed in the cross-layer case and how the parameters evolve depending on channel conditions. In particular, we focus on the redundancy allocated to each video layer as DA-UEP parameter (left axis), and the λ which defines the HM scheme (right axis). We can observe that, when users are affected by bad channel conditions, $\lambda = 0$, which means that only QPSK modulation is transmitted in order to avoid increasing the PER. At higher layers, the redundancy allocated to the corresponding 16QAM data is simply set to zeros ($r_3 = r_4 = 0$) and the whole redundancy is allocated to the first two layers of the DA-UEP codes (r_1 and r_2). In this state, the DA-UEP favors the first layer over the second one

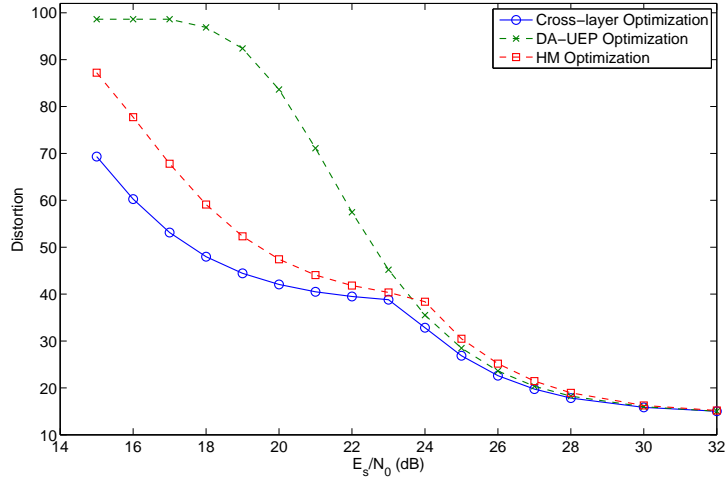


Figure 6.29: Comparison in terms of Video Distortion between the cross-layer optimization and the independent non-cross-layer optimizations.

($r_1 > r_2$) when the channel is too bad. This state holds until $E_s/N_0 = 22$ dB, then at $E_s/N_0 = 23$ dB the channel is better enough for DA-UEP to protect both first layers equivalently ($r_2 > r_1$).

On the other hand, for values greater than 23 dB, the behavior changes, and higher values of λ are optimally allocated, which makes the overall throughput increase due to the hierarchical transmission with QPSK and 16QAM users. At DA-UEP, the code provides unequal protection for both layers of HM schemes ($r_1 > r_2$ and $r_3 > r_4$). As the channel gets better, both HM schemes and DA-UEP tends to equally protect data $\lambda = 0.5$ and $r_1 \approx r_2 \ll r_3 \ll r_4$. Therefore, all video layers are protected in the optimal way, i.e. in order to obtain the best video quality distortion.

6.3.2.6.1 LMS Analysis

We extend the analysis of the cross-layer design for different channel environments, in particular, LMS at S-band [15], depicted in Figure 3.2 (see Chapter 3). The results, shown in Figure 6.31, let us verify that the cross-layer design outperforms the non-cross-layer design in all scenarios. As it was expected, users in an open area environment are receiving better video quality than intermediate and heavy tree shadowing for the same average E_s/N_0 .

6.4 Conclusions

6.4.1 Conclusions on the methodology

In this chapter, we have extended the cross-layer interaction of previous designs to higher layers (Transport and APP) in order to address not only the way upper-layer mechanisms

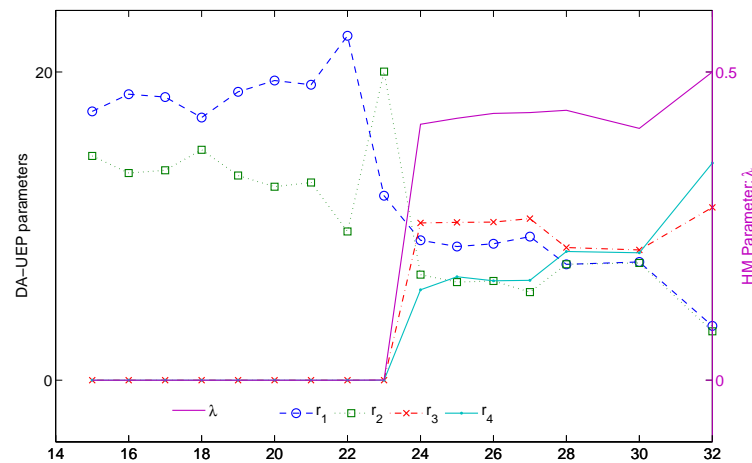


Figure 6.30: Evolution of DA-UEP (redundancy allocated to each video layer, left axis) and HM (λ , right axis) parameters for the cross-layer optimization.

are able to adapt to channel conditions, but also to measure the quality received with subjective parameters (for both video and voice).

Next, we proceed with the evaluation of the methodologies performed in this chapter. We identify two methodologies according to the nature of the implementation; adaptive bit rate and adaptive error protection.

6.4.1.1 Adaptive bit rate

Regarding the first designs based on bit rate adaptation, the *Layered Video Multicasting* and *Adaptive VoIP Codecs-Centralized* are clearly following a hybrid approach, as the designs of the previous chapter. In that case, the QoS requirements (delay, jitter, minimum video quality) can come from either IP layer or higher layers.

The *Adaptive VoIP Codecs-Distributed* design also follows a hybrid approach; information flow about available codecs from APP to Transport layers and information flow from Transport to APP with the desired VoIP codec to be used. However, the design is not performed according to the available resources as in the previous cases, but it depends on the measured delay of RTCP reports. This a clear example of the practical point of view of cross-layer designs, since no explicit optimization is performed; the VoIP codec is reduced or increased if delay/jitter requirements are no fulfilled.

6.4.1.1.1 Layered Video Multicasting

1. Objective: Based on the forward link design of the previous chapter, we aim at optimally distributing the video layers among the ACM modes. Scalable video coding is considered at the APP layer.
2. Layers involved: PHY, MAC, IP and APP are considered (see Figure 6.32). The new cross-layer interaction is a MAC to APP layer optimal video layers distribution.

6. Cross-Layer Design at PHY/APP Level

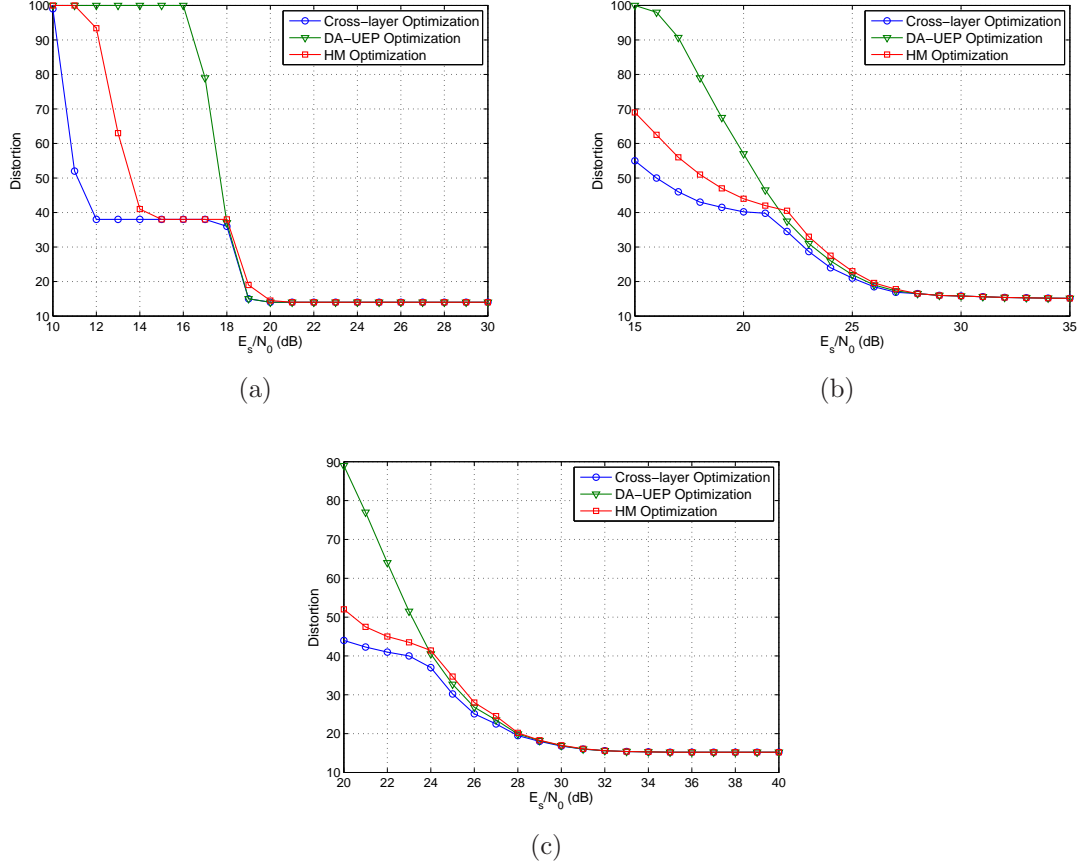


Figure 6.31: Performance of the UEP Cross-Layer for different LMS scenarios: (a) Open area; (b) Intermediate tree shadowing (ITS); (c) Heavy tree shadowing (HTS).

3. Mathematical tools: The NUM allow us to define a new utility functions that takes advantage of the multicast nature of ACM, and to introduce a new priority weight that distributes layers according to channel conditions. The solution can be interpreted as a water filling layered multicasting.
4. Formulation: Same constraints of the previous design, little modification regarding the SVC features is included.

6.4.1.1.2 Adaptive VoIP Codecs-Centralized

1. Objective: To maintain the VoIP connections and maximize the overall quality of the network according to the available capacity of a satellite/hybrid system implementing ACM. The optimization is performed by means of adaptive VoIP codecs (AMR-WB or RTCP codecs)
2. Layers involved: PHY and APP layers are the base of this design (see Figure 6.33). PHY layer flow to APP with the ACM mode of each ST, where the cross-layer man-

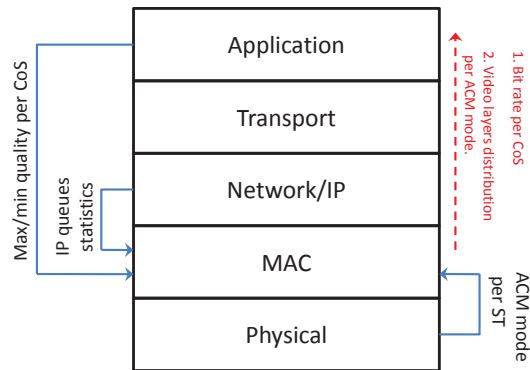


Figure 6.32: Cross-layer interaction and methodology conclusions: Layered Video Multicasting

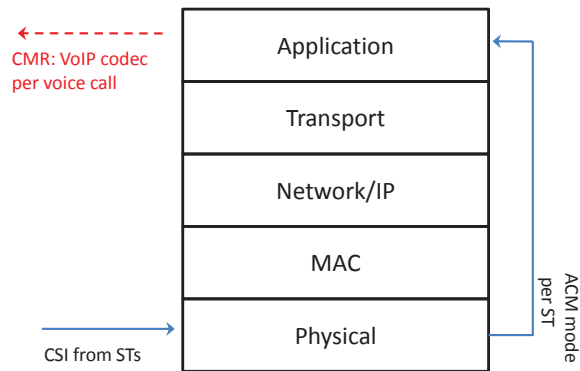


Figure 6.33: Cross-layer interaction and methodology conclusions: Adaptive VoIP Codecs-Centralized

ager computes the VoIP codecs per voice call depending on the available capacity. The manager includes the optimal codecs inside the CMR headers. Note that information of headers and encapsulation protocols from Transport and IP layers are also needed to perform the network dimensioning.

3. Mathematical tools: We assume NUM to formulate the problem due to the already mentioned features. However, since the possible solutions are a set of values defined by the available codecs, the problem becomes non-convex and we must perform an iterative solution to obtain the optimal allocation.
4. Formulation: The priority weight can depend on channel conditions or services priorities. The constraints are the set of available codecs and available capacity considering IP/Transport headers.

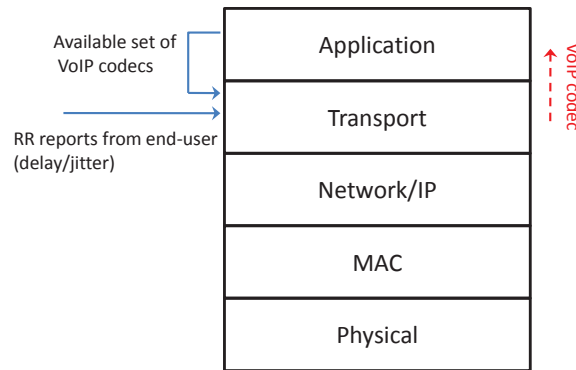


Figure 6.34: Cross-layer interaction and methodology conclusions: Adaptive VoIP Codecs-Distributed

6.4.1.1.3 Adaptive VoIP Codecs-Distributed

1. Objective: Each user aims at optimizing the quality voice call by means of adaptive VoIP codecs and depending on measured delay/jitter statistics.
2. Layers involved: An interaction between Transport and APP layers is assumed (see Figure 6.34). The quality of the transmission is measured thanks to RTCP control packets. Transport layer decides to reduce/increase VoIP codec depending on the received report. Then, it inform APP layer about the new desired codec to be used.
3. Mathematical tools: In this case, it is not possible to directly match the delay statistics with the optimal codec to be used by means of NUM since the end-user is not aware of the available capacity of the link. In such case, a pseudo-code algorithm is the optimal way to perform the speech rate adaptation.
4. Formulation: Delay and jitter constraints.

6.4.1.2 Adaptive error protection

On the other hand, the error protection designs do not seem to be easily classified according to the direction of the information flow. Firstly because the effect of the adaptive physical layer (increase/decrease capacity) is not directly appreciated in the protection mechanisms (assuming that modulations and coding rates at PHY layer are already defined). And second, the protection mechanisms are totally independent between layers, and they are not interconnected by means of the data plane (queues, buffers, schedulers), such as in the case of load/rate control mechanisms, which might contribute with status parameters. In that case, we argue that a classification according to the optimization criteria is more suitable.

The study performed shows that our two Adaptive Error Protection designs follow a global performance optimization oriented approach. It is demonstrated that the impact on system design of this type of cross-layer optimization is higher than other solutions. Unfortunately, exhaustively trying all the possible strategies and their parameters during

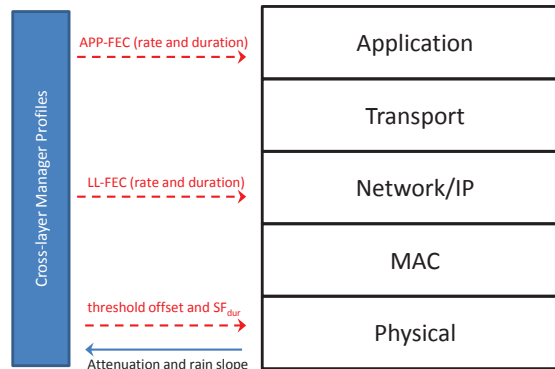


Figure 6.35: Cross-layer interaction and methodology conclusions: High layer-FEC protection and PHY layer optimization

the transmission in order to choose the composite strategy leading to the best quality performance is impractical due to the associated complexity. A solution proposed herein is to perform an *offline* cross-layer, meaning that all the protection strategies are predefined and associated to different possible scenarios. In that case, the cross-layer entity in charge of performing the adaptation only needs to check simple system/channel features and choose the best strategy. Next, a summary of both designs and scenarios:

6.4.1.2.1 High layer-FEC protection and PHY layer optimization

1. Objective: Maximize the PSNR of a video transmission in hybrid and satellite scenarios that implement adaptive physical layer. In particular, we focus on errors due to ACM reaction delay and deep tropical rain attenuations. Trading video quality and bandwidth efficiency is part of the objective of the cross-layer design.
2. Layers involved: PHY, IP and APP layers are the main layers considered for the design (see Figure 6.35). Threshold offset and SF_{dur} at the PHY layer, LL-FEC at the IP layer and APP-layer at the APP layer are the proposed mechanisms to be jointly optimized.
3. Mathematical tools: An iterative solution that checks all the possible parameters combinations leads to the optimal cross-layer solution. However, since the complexity of verifying all possible results, we propose an *offline* approach with the following scenarios.
4. Possible scenarios: Depending on the rain slope and on type of area (Tropical and Temperate).

6.4.1.2.2 UEP optimization

1. Objective: Minimize video distortion

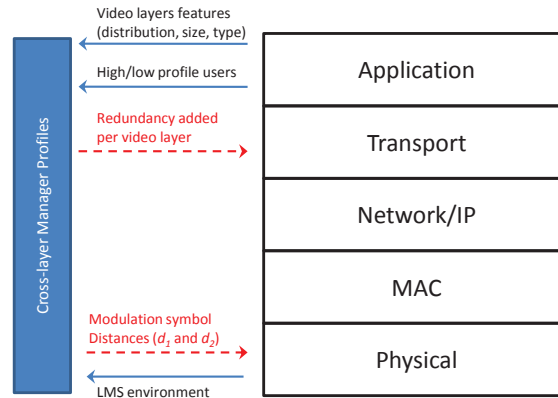


Figure 6.36: Cross-layer interaction and methodology conclusions: UEP optimization

2. Layers involved: PHY, Transport and APP layer (see Figure 6.36). Hierarchical modulation, DA-UEP and scalable video coding are the assumed mechanisms at PHY, Transport and APP layers respectively.
3. Mathematical tools: NUM is proposed to represent the problem. However, since the problem is non-convex, we propose an interactive solution as the one of the previous design.
4. Possible scenarios: Depending on different LMS scenarios (urban, tree shadowing, etc.) and on the number of users requesting for higher/lower quality.

6.4.2 Conclusions on the performance gain

A summary of the main achievements gains per cross-layer design can be found.

In the *Layered Video Multicasting*, our optimal way to distribute video layers among the ACM modes (taking advantage of the multicast nature of both) outperforms non-cross-layer broadcast and fair queuing designs, up to 20% and 40% of rate gain respectively. Delay and MOS results also show the strength of the design.

The *Adaptive VoIP Codecs-Distributed* design shows excellent results in terms of probability of exceeding the VoIP requirement since for attenuations lower than 15 dB the probability is zero, while the fixed codec and the RTCP-driven scenarios reach 7% and 4% of exceeding probability respectively.

In the centralized approach of the *Adaptive VoIP Codecs* design, we can observe that for heavy situation where 75% of users are under a rain attenuation of 10 dB, the single codec losses 45% of the VoIP connections, while AMR only 5%.

The main conclusion of the *High layer-FEC protection and PHY layer* design is that an appropriate choice of the shifted threshold can achieve almost the same PER reduction as using the strongest high layer codes, but with a higher bandwidth efficiency gain.

Finally, in the *UEP optimization* design, the results show that the joint protection cross-layer design outperforms the non-cross-layer solutions in up to 50% for DA-UEP, and up to 20% in HM, for the range of channel values studied herein.

Chapter 7

Overall Conclusions and Future Work

7.1 Conclusions

First of all, let us say that the work presented in this dissertation clearly shows that the goals of creating a cross-layer framework and providing QoS guarantees for broadband adaptive satellite systems has been successfully achieved. Moreover, the methodology for reaching this goal is probably as important as the results. Such methodology includes the design of cost-efficient satellite and hybrid architectures, the combination of different theoretical tools such as NUM and game theory, the choice of the available protocols/mechanisms at different layers and the definition of novel performance policies in terms of fairness, efficiency, etc.

For this dissertation, we have selected challenging problems of the satellite communications domain, where the typical layered optimization leads to inefficient performance. We have demonstrated that for these cases, only cross-layer can lead to an optimal design. In particular, the ACM-based satellite systems and how the adaptivity is followed at upper layers has been the main subject of this thesis.

The diversity of designs has been chosen on purpose in order to deal with different problem designs, implementations, formulation and solutions. Next, we summarize the main aspects of each design: objectives, methodologies and the main gains obtained.

A first approach of the cross-layer designs was to propose a multicast optimization of the air interface of a hybrid satellite system using SC, not only as an alternative for fast/slow channel dynamics but also as a possible cooperation framework with ACM. The results showed that the hierarchical allocation maximizes the capacity when non CSI is available at the transmitter.

In Chapter 5, the ACM has been studied in depth together with the bandwidth allocation in adaptive satellite systems. We focus on a cross-layer optimization design between PHY, MAC and IP layers by means of NUM and game theory tools. Both mathematical frameworks proved a very efficient and straightforward methodology for addressing resource allocation problems in adaptive physical layer systems. Moreover, they provided a

fair distribution of the satellite resources among users under different weather conditions, and for different QoS constraints.

For the forward link, the challenge has been to efficiently deliver delay-constrained applications, and we have demonstrated that the design works perfectly as a trade-off between the delay requirements of premium classes and the total delay. For the return link, we aimed at proposing a distributed capacity request game implemented in the bandwidth allocation of the multibeam broadband satellite networks. Nash equilibrium and Pareto optimality were used to explain the performance of the game efficiency. We proved that it not only performs efficient and fair capacity requests, but it is also able to reduce the bandwidth request signaling (increasing the bandwidth gain up to 8%).

The first design (*Layered Video Multicasting*) of Chapter 6 is a natural extension of the PHY /MAC /IP theoretical framework of Chapter 5 to the APP layer. We solved a new NUM that includes cross-layer interactivity with an adaptive physical layer and SVC. In particular, the solution (fairly and efficiently) distributed the video layers to be sent through each ACM mode in order to serve a multicast scenario. We found out that our solution was able to trade between rate, delay and layered multicasting in order to counteract not only with channel conditions but also with traffic requirements. Our optimal way to distribute video layers among the ACM modes (taking advantage of the multicast nature of both) outperforms non-cross-layer broadcast and fair queuing designs, up to 20% and 40% of rate gain respectively.

In the following design, we analyzed the physical adaptivity for a high delay constrained application, the VoIP, and we proposed to use adaptive codecs. A distributed solution, which aims at individually maximizing the user perceived quality, was designed. A side effect of the results observed was that the rate control on aggregated traffic provided a powerful tool that could be used for load control. Given that, we proposed a centralized approach, where the key idea was to maintain as many as possible VoIP connections and maximize the overall quality of the network.

One of the interests of the cross-layer scheme based on Adaptive VoIP is that the design was evaluated for a practical case; the Brasil Project, specifically focused on rural and remote areas, where VoIP connections are served by the hybrid network. Thus we tested our technique in tropical areas, where rain attenuation can be very deep and new design techniques must be identified. In particular, with our proposed scheme, we showed that a cross-layer design for the proposed reference hybrid system scenario allows maintaining the number of connections for rain attenuations lower than 15 dB with 50% of the beam affected by rain.

However, we observed that some effects of the tropical satellite channels, which reduced received QoS, could not be solved by means of adaptive rate mechanisms at higher layers. Such effects are; the delay in the ACM reaction to fade changes, and the fact that ACM modes cannot cope with the deepest attenuations. Different FMT have been evaluated in order to avoid the QoS reduction. LL-FEC and AL-FEC have been considered as high layer error protection, and threshold offset and super-frame as the design parameters to be optimized at the PHY layer. We showed that the proposed FMT are useful to avoid reduced video quality. However, we demonstrated that an appropriate choice of the *shifted threshold* can achieve almost the same PER reduction as using the strongest high layer codes, but with a higher bandwidth efficiency gain.

The last design has been built up as a solution to improve the error protection performance in multicast scenarios, such as the one studied for the *Layered Video Coding* design. The chosen Transport layer technique, named DA-UEP, achieves high efficiency by optimizing the video layers length and the redundancy added to each one, which is allocated depending on the dependencies between frames at data level. The design is completed with the presence of Hierarchical Modulation (HM) at the PHY layer, which is also seen as a solution for multicast services. HM achieves highest overall throughput than other solutions (as shown in Chapter 4) and it is modeled according to the distances between constellation symbols (d_1 and d_2). Therefore, we presented a multi-layer protection scheme based on a APP/Transport/PHY joint layer design, allowing users diversity depending on channel conditions. The results showed that the cross-layer design outperforms the non-cross-layer solutions (DA-UEP and HM not jointly optimized) in up to 50% when only DA-UEP is optimized, and up to 20% when only HM is optimized, for the range of channel values studied herein.

7.2 Conclusions on the methodology

Given the description of the cross-layer designs addressed in this dissertation and the detailed methodologies at the end of the last chapters, it is straightforward to define a cross-layer methodology framework for adaptive satellite systems. According to the designs studied, we assume two cross-layer types: based on bit rate adaptation (which also includes bandwidth allocation mechanisms) and adaptive error protection. Such classification does not mean that, e.g. error protection could not be optimized according to information from adaptive bit rate mechanisms, however, it means that we are not considering a joint optimization of both mechanisms.

7.2.1 Bit rate adaptation

In this case, the adaptive physical layer strongly affects the higher layer performance due to satellite link capacity variations. The MAC layer plays a vital role since it is not only the first layer to realize that the capacity decrease/increase but it is also the responsible of the resource allocation (scheduler). Therefore, we propose the following designs regarding the direction of cross-layer interaction:

- A bottom-up approach, where the adaptive PHY layer must be followed at higher layers.
- An hybrid approach, where an intermediate layer such as MAC/IP is in charge of receiving the new available capacity from PHY layer and the mechanisms strategies/properties from higher layers.

The main drawback of the first approach is that each layer is only optimized according to information from lower layer, and it is no aware of higher layer strategies.

We aim also at presenting a cross-layer framework for both centralized and distributed allocation. Moreover, we match the chosen approaches with the most suitable mathematical tool to be implemented with:

1. Centralized mechanisms happen to provide straightforward implementations of resource allocation optimization problems. NUM has been proved to efficiently represent all priorities of centralized allocation mechanisms, constraints and policies. According to traffic properties, two ways of proceeding are differentiated:
 - (a) Inelastic traffic (delay constrained): it makes utility functions non-concave
 - i. Approximation of the original utility functions to deal with non-convexity: suboptimal solution
 - ii. Customize priorities (e.g. delay-dependants) and constraints to make the problem convex: optimal solutions with rate/delay balancing.
 - iii. Non-convex problem: Iterative algorithm to reach the optimal strategy
 - A. High complexity: *offline* profiles.
 - B. Low complexity: *online* computation of the optimal solution.
 - (b) Elastic traffic: concave utility functions which leads to optimal solutions.
2. Distributed mechanisms require more complex implementations, due to the fact that satellite systems are designed to be centralized. However, advantages such as faster adaptation of end-user or fairness make decentralization a future path for new designs. In this case, game theory is suitable to model non-cooperative scenarios/games, where the distributed nature and the big satellite delay do not allow direct cooperation between end-users. The ways to obtain the solution are classified as in the NUM case, however, according to the solutions obtained (either if traffic is elastic or if we achieve a convex problem), three different solutions can be obtained:
 - (a) Nash Equilibrium: Inefficient allocation with respect to a centralized solution, normally due to lack of network state information.
 - (b) Pareto Optimal: Optimal network operating point, i.e. maximize overall resource allocation.
 - (c) Nash Bargaining Solution: It is not only pareto optimal but also satisfies the axioms of fairness, i.e. fair allocation among all users.

7.2.2 Adaptive error protection

As we previously mentioned, the capacity drop due to ACM does not impact the error protection mechanisms as it does to the rate adaption designs. Therefore, we propose a different classification:

1. Single mechanism optimization: a particular error protection mechanism is adapted according to parameters of lower (e.g. depending on channel conditions) and upper layers (e.g. depending on type and features of video source). Two types of error protection are available:
 - (a) Equal error protection: implemented with linear and concave functions, leading to optimal solutions.
 - (b) Unequal error protection: non-convex problem
 - i. Lagrangian relaxation and dynamic programming might lead to an optimal solution.
 - ii. Iterative algorithm
 - A. High complexity: *offline* profiles.
 - B. Low complexity: *online* computation of the optimal solution.
2. Global performance optimization oriented approach: two or more error protection mechanisms are jointly optimized. The sole solution in this case is to iterate considering all the possible strategies and parameters in order to obtain a composite strategy. Therefore, we only need to determinate the complexity of the problem:
 - (a) High complexity: *offline* profiles.
 - (b) Low complexity: *online* computation of the optimal solution.

7.2.3 Final comments

Therefore, this dissertation has explored the universe of cross-layer designs and methodologies. We have demonstrated that the proposed designs are feasible thanks to detailed architectures and signaling parameters, which can be implemented in any ACM-based standard. In particular, for the signaling, we have proposed specific and detailed SI-SAP primitives for the cross-layer interaction between MAC and IP layers. Moreover, we defined different methodologies for adaptive physical layer designs according to the nature of the designs included herein (protection and allocation frameworks).

To conclude, let us say that the work has benefited from participating in several international projects, which has resulted in theoretical as well as practical understanding of the proposed optimizations and corresponding designs.

Moreover, this dissertation is backed by a good set of publications with four journal papers (two submitted), one chapter of a book, seven international conference papers, two contributions to international standards (ITU and ETSI) and one patent (currently being processed).

7.3 Future works

As final note, it is worth mentioning some future research areas directly related to the work presented in this document. We classify the future extensions of each cross-layer design.

In Chapter 4, we consider:

- To explore the compatibility between ACM and HA-SC in mobile scenarios affected by weather conditions. In other words, the terminal must be able to efficiently adapt to fast/slow/very slow channel conditions.

In Chapter 5, future works could be:

- The evaluation of the forward link approach in terms of PSNR to measure user satisfaction instead of optimizing the bit rate allocation.
- To verify the partially distributed design of the forward link in a protocol-based simulator.

And finally, in Chapter 6, we propose some possible extensions:

- To study the performance of UEP (at Transport layer) in the *Layered Video Multicasting* design in order to protect video layers under mobile conditions and within a multicast scenario.
- To design a dynamic threshold offset depending on the slope and the error estimation. Future work should additionally consider the trade offs between video quality and channel occupancy.
- To improve the cross-layer evaluation of the *UEP optimization* design with real video coding and in terms of PSNR/MOS.
- To propose specific scenarios for the Adaptive Error Correction solutions in order to create an offline cross-layer framework.

References

- [1] L. Baldantoni, H. Lundqvist, and G. Karlsson. Adaptive end-to-end FEC for improving TCP performance over wireless links. *Communications (ICC), IEEE International Conference on*, 7:4023–4027, 2004. 1
- [2] S. Hong, Y. Won, and D. I. Kim. Significance-aware channel power allocation for wireless multimedia streaming. *Vehicular Technology, IEEE Transactions on*, 59(6):2861–2873, 2010. 1
- [3] M.A.V. Castro and D. Pradas-Fernandez. VoIP cross-layer load control for hybrid satellite-wimax networks. *Wireless Communications, IEEE*, 15(3):32–39, 2008. 2
- [4] V. Kawadia and P.R. Kumar. A cautionary perspective on cross-layer design. *Wireless Communications, IEEE*, 12(1):3–11, feb. 2005. 9
- [5] G. Giambene (Ed.). *Resource Management in Satellite Networks: Optimization and Cross-Layer Design*. Springer, 2007. 10
- [6] M. van der Schaar and S. Shankar. Cross-layer wireless multimedia transmission: Challenges, principles and new paradigms. *Wireless Communications Magazine, IEEE*, 12(4):50–58, 2005. 10
- [7] Qi Wang and M.A. Abu-Rgheff. Cross-layer signalling for next-generation wireless systems. In *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, volume 2, pages 1084–1089 vol.2, march 2003. 13, 14, 15
- [8] ETSI: TR 101 985. *Satellite Earth Stations and Systems (SES), Broad-band Satellite Multimedia, IP over Satellite, ETSI Technical Report*. 2002. 15
- [9] ETSI TR 101 984. *Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM); Services and Architectures*. December 2007. 16, 58, 74
- [10] ETSI TR 102 353. *Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM); Guidelines for the Satellite Independent Service Access Point (SI-SAP)*. November 2004. 16, 58, 59, 112
- [11] M. A. Vazquez Castro and G. Seco-Granados. Cross-layer packet scheduler design of a multibeam broadband satellite system with adaptive coding and modulation. *Wireless Communications, IEEE Transactions on*, 6(1):248–258, 2006. 17, 102, 113

-
- [12] J. Goldhirsh. Two-dimensional visualization of rain cell structures. *Radio Science*, 35(3):713–729, Jun. 2002. 18
- [13] R. J. Acosta. *Presentation of RF propagation research, NASA Glenn Research Center (Ohio, 2006)*. Available at sunset.usc.edu/gsaw/gsaw2006/s10d/acosta.pdf. 19
- [14] M.S. Pontes, L. da Silva Mello, R.S.L. de Souza, and E.C.B. Miranda. Review of rain attenuation studies in tropical and equatorial regions in Brazil. In *Information, Communications and Signal Processing, Fifth International Conference on*, pages 1097–1101, 0-0 2005. 19
- [15] F. Perez-Fontan, M.A. Vazquez-Castro, S. Buonomo, J.P. Poiaraes-Baptista, and B. Arbesser-Rastburg. S-band LMS propagation channel behaviour for different environments, degrees of shadowing and elevation angles. *Broadcasting, IEEE Transactions on*, 44(1):40–76, March 1998. 19, 40, 41, 137
- [16] WiMAX Forum. *Mobile WiMAX-Part I: A Technical Overview and Performance Evaluation*. August 2006. 19, 28, 110, 111
- [17] V.S. Abhayawardhana, I.J. Wassell, D. Crosby, M.P. Sellars, and M.G. Brown. Comparison of empirical propagation path loss models for fixed wireless access systems. In *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, volume 1, pages 73–77 Vol. 1, May–1 June 2005. 19
- [18] ETSI: EN 302 307. *Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2)*. 2009. 19, 28, 40, 57, 72, 120, 121, 129
- [19] ETSI: TR 102 376. *Digital Video Broadcasting (DVB) User guidelines for the second generation system for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2)*. 2005. 19, 28
- [20] ETSI EN 301 192 v1.4.1. *Digital Video Broadcasting (DVB); DVB Specification for Data Broadcasting*. 2004. 22, 118, 119
- [21] ETSI: EN 301 790 v1.5.1. *Digital Video Broadcasting (DVB); Interaction Channels for Satellite Distribution Systems*. May 2009. 22, 23, 28, 74, 116
- [22] ETSI: TR 101 790. *Digital Video Broadcasting (DVB); Interaction channel for Satellite Distribution Systems; Guidelines for the use of EN 301 790*. July 2009. 22, 28, 73
- [23] IEEE Std 802.16e 2005. *Digital Video Broadcasting (DVB); Amendment to IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*. February 2006. 24

REFERENCES

- [24] H. Skinnemoen. *Presentation on Brasil Project*. Available at www.wisecom-fp6.eu/istsummit/11_skinemoen.pdf. 27
- [25] R. Aroso. *Presentation on e-Brasil Project*. Available at www.fp6-wisecom.org. 27
- [26] A. Centonza and S. McCann. Architectural and protocol structure for composite DVB-RCS/IEEE 802.16 platforms. *Digital Video Broadcasting Over Satellite: Present and Future. The Institution of Engineering and Technology Seminar on*, pages 35 – 40, November 2006. 29
- [27] S. Khan, S. Ali Mahmud, and H. Al-Raweshidy. Supplementary interworking architecture for hybrid data networks (UMTS-WiMAX). *Computing in the Global Information Technology*, August 2006. 29
- [28] WMF-T32-003-R010v05. *DWiMAX Forum Network Architecture - Architecture Tenets, Reference Model and Reference Points Part 2 - Release 1.0*. May 2009. 30
- [29] C.E. Perkins. IP mobility support for IPv4. *Internet Engineering Task Force (IETF), RFC 5944*, November 2010. 31
- [30] H. H. Lee, T. Kwon, D. Cho, G. Lim, and Y. Chang. Performance analysis of scheduling algorithms for VoIP services in IEEE 802.16e Systems. *Vehicular Technology Conference*, 3:1231–1225, 2006. 31
- [31] ETSI TS 102 584. *Digital Video Broadcasting (DVB);DVB-SH Implementation Guidelines*. Dec. 2008. 34, 116, 130
- [32] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, New York, 1991. 34, 130
- [33] S. Shamai and A. Steiner. A broadcast approach for a single-user slowly fading MIMO channel. *Information Theory, IEEE Transactions on*, 49(10):2617 – 2635, 2003. 35, 36
- [34] A. Steiner and S. Shamai. Achievable rates with imperfect transmitter side information using a broadcast transmission strategy. *Wireless Communications, IEEE Transactions on*, 7(3):1043 – 1051, 2008. 35
- [35] T.T. Kim and M. Skoglund. On the expected rate of slowly fading channels with quantized side information. *Communications, IEEE Transactions on*, 55(4):820 – 829, 2007. 35
- [36] C. Ng, A. Gunduz, D. Goldsmith, and E. Erkip. Distortion minimization in gaussian layered broadcast coding with successive refinement. *Information Theory, IEEE Transactions on*, 55(11):5074 – 5086, 2009. 35
- [37] J. Luo, R. Yates, and P. Spasojevic. Service outage based power and rate allocation for parallel fading channels. *Information Theory, IEEE Transactions on*, 51(7):2594 – 2611, 2005. 35

-
- [38] N. Jindal and A. Goldsmith. Capacity and optimal power allocation for fading broadcast channels with minimum rates. In *IEEE Global Telecommunications Conference (GLOBECOM)*, volume 2, pages 1292 – 1296, 2001. 35
- [39] C. T. K. Ng and A. Goldsmith. Capacity of fading broadcast channels with rate constraints. In *Forty-Second Annual Allerton Conference on Communication, Control and Computing*, 2004. 35
- [40] Y. Liu, K. N. Lau, C. Y. Takeshita, and M. P. Fitz. Optimal rate allocation for superposition coding in quasistatic fading channels. In *Information Theory, IEEE International Symposium on*, 2002. 35
- [41] S. R. Mirghaderi, A. Bayesteh, and A. K. Khandani. On the maximum achievable rates in wireless multicast networks. In *Information Theory (ISIT), IEEE International Symposium on*, pages 2776 –2780, 2007. 36
- [42] G. S. G. Beveridge and R. S. Schechter. *Optimization: Theory and Practice*. McGraw-Hill Book Company, New York, 1970. 37
- [43] R. E. Kalaba and K. Spingarn. Optimization of functionals subject to integral constraints. *Journal of Optimization Theory and Applications*, 24:325–335, 1978. 10.1007/BF00933285. 37
- [44] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Prentice-Hall, Englewood Cliffs, N.J, 1964. 37
- [45] L. Li and F. B. Gross. A new polynomial approximation for J_v Bessel functions. In *Asia-Pacific Microwave Conference Proceedings*, volume 4, 2005. 38
- [46] P.K. Gopala and H. El Gamal. On the throughput-delay tradeoff in cellular multicast. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, volume 2, pages 1401 – 1406 vol.2, 2005. 38
- [47] U.C. Kozat. On the throughput capacity of opportunistic multicasting with erasure codes. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pages 520 –528, 2008. 39
- [48] Linghang Fan, Hongfei Du, and B.G. Evans. A cross-layer delay differentiation packet scheduling scheme for multimedia content delivery in 3G satellite multimedia systems. *Broadcasting, IEEE Transactions on*, 54(4):806–815, December 2008. 48
- [49] Jia Tang and Xi Zhang. Cross-layer-model based adaptive resource allocation for statistical qos guarantees in mobile wireless networks. *Wireless Communications, IEEE Transactions on*, 7(6):1536–1276, June 2008. 48
- [50] Qingwen Liu, Shengli Zhou, and G.B. Giannakis. Cross-layer scheduling with prescribed QoS guarantees in adaptive wireless networks. *Selected Areas in Communications, IEEE Journal on*, 23(5):1056, May 2005. 48

REFERENCES

- [51] L.B. Le, E. Hossain, and A.S. Alfa. Service differentiation in multirate wireless networks with weighted round-robin scheduling and ARQ-based error control. *Communications, IEEE Transactions on*, 54(2):208, February 2006. 48
- [52] Hsien-Po Shiang and M. van der Schaar. Multi-user video streaming over multi-hop wireless networks: a distributed, cross-layer approach based on priority queuing. *Selected Areas in Communications, IEEE Journal on*, 25(4):770, May 2007. 48
- [53] J. She, Xiang Yu, Pin-Han Ho, and En-Hui Yang. A cross-layer design framework for robust IPTV services over IEEE 802.16 networks. *Selected Areas in Communications, IEEE Journal on*, 27(2):235, February 2009. 48
- [54] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *The Journal of the Operational Research Society*, 49(3):The Journal of the Operational Research Society, Vol. 49, No. 3. (1998), pp. 237–252. doi:10.2307/3010473 Key: citeulike:5175378, 1998. 48, 51, 54, 63, 68
- [55] S. H. Low. A duality model of TCP and queue management algorithms. *Networking, IEEE/ACM Transactions on*, 1(4):1063–6692, August 2003. 48, 63
- [56] D.P. Palomar and Mung Chiang. A tutorial on decomposition methods for network utility maximization. *Selected Areas in Communications, IEEE Journal on*, 24(8):1439, August 2006. 49
- [57] S.H. PMung Chiang Low, A.R. Calderbank, and J.C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1):255, January 2007. 49
- [58] P. Hande, Zhang Shengyu, and Chiang Mung. Distributed rate allocation for inelastic flows. *Networking, IEEE/ACM Transactions on*, 15(6):1063–6692, December 2007. 49
- [59] S. Shenker. Fundamental design issues for the future internet. *Selected Areas in Communications, IEEE Journal on*, 13(7):1176, September 1995. 49
- [60] Ying Li, Zhu Li, Mung Chiang, and A.R. Calderbank. Content-aware distortion-fair video streaming in networks. In *IEEE Global Telecommunications Conference (GLOBECOM)*, November 2008. 49
- [61] M.S. Talebi, A. Khonsari, M.H. Hajiesmaili, and S. Jafarpour. A suboptimal network utility maximization approach for scalable multimedia applications. In *IEEE Global Telecommunications Conference (GLOBECOM)*, November 2009. 49
- [62] F. Vieira, M. A. Vazquez Castro, and G. Seco. A tunable-fairness cross-layer scheduler for DVB-S2. *Satellite Communications and Networking, International Journal of*, 24(5):437450, September 2006. 53

-
- [63] funded by the European Space Agency (ESA). *IP-friendly cross-layer optimization of adaptive satellite systems*. available at <http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=28782>, 2008. 53
- [64] Haining Wang, Chia Shen, and K.G. Shin. Adaptive-weighted packet scheduling for premium service. In *Communications (ICC), IEEE International Conference on*, volume 6, page 1846, June 2001. 54
- [65] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *Networking, IEEE/ACM Transactions on*, 8(5):556, October 2000. 54
- [66] L. Boyd and S. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. 55, 56, 71
- [67] W. Yu and J. M. Cioffi. Constant-power waterfilling: Performance bound and low-complexity implementation. *Communications, IEEE Transactions on*, 54(1), January 2006. 56
- [68] D. D. Yu and J. M. Cioffi. Iterative water-filling for optimal resource allocation in OFDM multiple-access and broadcast channels. In *IEEE Global Telecommunications Conference (GLOBECOM)*, November 2009. 56
- [69] G. Seco-Granados, M. A. Vazquez Castro, A. Morell, and F. Vieira. Algorithm for fair bandwidth allocation with QoS constraints in DVB-S2/RCS. In *IEEE Global Telecommunications Conference (GLOBECOM)*, December 2006. 56
- [70] ETSI TR 102 463. *Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM); Interworking with IntServ QoS*. November 2004. 58
- [71] RFC. *Resource ReSerVation Protocol (RSVP)*. September 1997. 59
- [72] Illsoo Sohn, Kwang Lee, and Young Choi. Comparison of decentralized time slot allocation strategies for asymmetric traffic in TDD systems. *Wireless Communications, IEEE Transactions on*, June 2009. 61
- [73] M. Salem, A. Adinoyi, M. Rahman, and H. Yanikomeroglu. Fairness-aware radio resource management in downlink OFDMA cellular relay networks. *Wireless Communications, IEEE Transactions on*, May 2010. 61
- [74] R. Bolla, F. Davoli, and M. Marchese. Adaptive bandwidth allocation methods in the satellite environment. In *Communications (ICC), IEEE International Conference on*, volume 10, pages 3183 – 3190, June 2001. 61
- [75] A. Morell, G. Seco-Granados, and M.A. Vazquez-Castro. Fairness-aware radio resource management in downlink OFDMA cellular relay networks. *IEEE Systems Journal*, March 2008. 61

REFERENCES

- [76] N. Celandroni, E. Ferro, and F. Potorti. Comparison between distributed and centralised demand assignment TDMA satellite access schemes. *Satellite Communications and Networking, International Journal of*, March 2006. 61
- [77] N. Iuoras and Tho Le-Ngoc. Dynamic capacity allocation for quality-of-service support in IP-based satellite networks. *Wireless Communications, IEEE Transactions on*, October 2005. 62
- [78] M. J. Osborne and A. Rubenstein. *A course in game theory*. The MIT Press, Cambridge, Massachusetts, 1994. 62
- [79] H. Zhang, D. Towsley, and W. Gong. TCP connection game: a study on the selfish behavior of TCP users. In *Network Protocols, International Conference on*, November 2005. 62
- [80] A. Akella, S. Seshan, R. Karp, S. Shenker, and C. Papadimitriou. Selfish behavior and stability of the internet: A game-theoretic analysis of TCP. In *Proceedings ACM SIGCOMM Conference*, 2002. 62, 69
- [81] H. Yaiche, R.R. Mazumdar, and C. Rosenberg. A game theoretic framework for bandwidth allocation and pricing in broadband networks. *Networking, IEEE/ACM Transactions on*, 8(5):667–678, October 2000. 63, 70, 71
- [82] E. Suris, L. A. DaSilva, Han Zhu, and A. B. MacKenzie. Asymptotic optimality for distributed spectrum sharing using bargaining solutions. *Wireless Communications, IEEE Transactions on*, October 2009. 63
- [83] Z. Han, Z. Ji, , and K. Liu. Fair multiuser channel allocation for OFDMA networks using nash bargaining solutions and coalitions. *Communications, IEEE Transactions on*, August 2005. 63
- [84] E. Del Re, G. Gorni, L. Ronga, and R. Suffritti. A power allocation strategy using game theory in cognitive radio networks. In *International Conference on Game Theory for Networks (GameNets)*, pages 117–123, May 2009. 63
- [85] D. K. Petraki, M. P. Anastasopoulos, C. Hsiao-Hwa, and P. G. Cottis. Distributed resource allocation for delay-sensitive services in satellite networks using game theory. *Computational Intelligence and AI in Games, IEEE Transactions on*, June 2009. 63, 67
- [86] E. Del Re, G. Gorni, L. S. Ronga, and M. A. Vazquez Castro. A game theory approach for DVB-RCS resource allocation. In *Vehicular Technology Conference*, pages 2937–2941, May 2008. 63, 67
- [87] M. A. Vazquez Castro, Z. Han, A. Hjorungnes, and N. Marina. Rate allocation for satellite systems with correlated channels based on a Stackelberg game. In *International Conference on Game Theory for Networks (GameNets)*, pages 638–645, May 2009. 63, 67

-
- [88] A. Yun and J. Prat. *AmerHis meet mesh connectivity*, White paper by <http://www.satlabs.org>. 2007. 65
- [89] S. Vassaki, A. D. Panagopoulos, P. Constantinou, and M. A. Vazquez Castro. Market-based bandwidth allocation for broadband satellite communication networks. In *International Conference on Advances in Satellite and Space Communications (SPACOMM)*, pages 110–115, June 2010. 67
- [90] T. Basar and G. J. Olsder. *A Game-Theoretic Analysis of TCP Vegas*. Springer Berlin, 2004. 69
- [91] T. Alpcan and T. Basar. Distributed algorithms for Nash equilibria of flow control games. In *Advances in Dynamic Games: Applications to Economics, Finance, Optimization, and Stochastic Control*, Boston, MA. Birkhauser. 69, 70, 72
- [92] A. Muthoo. *argaining Theory with Applications*. Cambridge University Press, U. K. 71
- [93] J. Nash. The bargaining problem. *Econometria*, 18:155 – 162, 1950. 71
- [94] *Satlabs System Recommendations*. June 2008. 73
- [95] A. Viterbi and J. Omura. *Principles of Digital Communication and Coding*. McGraw-Hill Electrical Engineering, 1979. 91
- [96] ITU-T Recommendation G.107. *The E-model, a computational model for use in transmission planning*. 2001. 94, 105
- [97] ITU-T Recommendation P.800. *Methods for subjective determination of transmission quality*. 1996. 94, 104
- [98] ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10). *JVT, Advanced Video Coding (AVC)*. 2004. 94, 124
- [99] R.S. Wong, N. Sai Shankar, and M. van der Schaar. Integrated application MAC modeling for cross-layer optimized wireless video. In *Communications (ICC), IEEE International Conference on*, volume 2, pages 1271 – 1275 Vol. 2, may 2005. 98
- [100] M. van der Schaar, S. Krishnamachari, Sunghyun Choi, and Xiaofeng Xu. Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs. *Selected Areas in Communications, IEEE Journal on*, 21(10):1752 – 1763, dec. 2003. 98
- [101] S. Frederick R. Jacobson V. Schulzrinne, H. Casner. *RTP: A transport protocol for real-time applications*. 2003. 99
- [102] ITU-T Recommendation P.861. *Objective quality measurement of tel-ephone band (300-3400 Hz) speech codecs*. 1996. 104

REFERENCES

- [103] ITU-T Recommendations G.108. *Application of the E-Model - A planning guide*. 1999. 105
- [104] J. Radzik, O. Vidal, D. Pradas, M. Bousquet, and M.A. Vazquez Castro. Performance of hybrid satellite-WiMAX networks based on DVB-RCS satellite systems for rural and remote areas. In *International Communications Satellite Systems Conference (AIAA-ICSSC), Proceedings of*, jun. 2009. 114
- [105] J. Radzik, O. Vidal, D. Pradas, and M.A. Vazquez Castro. Cross-layer optimization of hybrid satellite network over the Amazonian region. In *Satellite and Space Communications (IWSSC). International Workshop on*, pages 398–402, sept. 2009. 114
- [106] ETSI EN 302 583 v1.1.2. *Digital Video Broadcasting (DVB); Framing Structure, channel coding and modulation for Satellite Services to Handheld devices (SH) below 3 GHz*. 2010. 116
- [107] ETSI TS 102 034 v1.4.1. *Digital Video Broadcasting (DVB); Transport of MPEG 2 Transport Stream (TS) Based DVB Services over IP Based Networks*. 2009. 118
- [108] R. Secchi, A. Sathiaselan, F. Potortì, A. Gotta, and G Fairhurst. Using quick-start to enhance TCP-friendly rate control performance in bidirectional satellite networks. *Satellite Communications and Networking, Int. Journal of*, 27(3):141 – 161, 2009. 120
- [109] B. Wang, X. Gu, and H. Zhang. An improvement to fine granularity scalability based on H.26L. In *Circuits and Systems (ISCAS). Proceedings of the International Symposium on*, volume 3, page 833, May. 2004. 124
- [110] B. G. Haskell, , and A. N. Netravali. *Digital Pictures: Representation, Compression, and Standards*. Perseus Publishing, 1997. 124
- [111] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 44(13), 2008. 125
- [112] Y. Yiu, K. N. Lau, C. Y. Takeshita, and M. P. Fitz. Optimal rate allocation for superposition coding in quasistatic fading channels. In *Information Theory (ISIT). IEEE International Symposium on*, Jun. 2002. 129
- [113] Seeling. P., P. de Cuetos, and M. Reisslein. Fine granularity scalable (fgs) video: Implications for streaming and a trace-based evaluation methodology. *IEEE Communications Magazine*, 43:138–142, 2005. 129, 132
- [114] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan. Priority encoding transmission. *IEEE Transactions on Information Theory*, 42:1737–1744, 1996. 130
- [115] A. E. Mohr, R. E. Ladner, and E. A. Riskin. Approximately optimal assignment for unequal loss protection. In *Image Processing. Proceedings of the International Conference on*, Aug. 2000. 130

- [116] A. Bouabdallah and J. Lacan. Dependency-aware erasure protection codes. *Journal of Zhejiang University (JZUS)*, 7:27–33, 2006. 130, 132
- [117] M. Morimoto, H. Harada, M. Okada, and S. Komaki. A study on power assignment of hierarchical modulation schemes for digital broadcasting. *IEICE Transactions on Communications*, E77-B, 1994. 130
- [118] P. K. Vitthaladevuni and M.-S. Alouini. BER computation of 4/M-QAM hierarchical constellations. *IEEE Transactions on Broadcasting*, 47(3):228–239, 2001. 130, 131

Cross-Layer Design and Methodology for Satellite Broadband Networking

Ph. D. Dissertation

DAVID PRADAS FERNÁNDEZ

Conception et méthodologie cross-layer pour les réseaux satellite à haut débit

Les satellites de communication apparaissent comme une solution particulièrement intéressante pour fournir une connectivité à large bande à un grand nombre d'utilisateurs. La propriété de diffusion naturelle des satellites (et ses bandes de fréquence Ku/Ka) rend leur utilisation évidente pour les services de multidiffusion, multimédia, ainsi que pour l'interconnexion de réseaux à haut débit. Toutefois, les signaux de cette bande de fréquence sont beaucoup plus sensibles aux interférences atmosphériques. La principale innovation permettant de résoudre ces problèmes a été l'adoption du codage et de la modulation adaptatifs. Ceci est le principal moteur de cette thèse car cette adaptabilité rend les systèmes par satellite traditionnels inefficaces.

Nous concentrons sur un paradigme différent pour traiter ces nouveaux défis. Celui-ci est basé sur l'optimisation conjointe des couches de la pile de protocoles. L'idée fondamentale est le fait que l'adaptabilité de la couche physique doit être suivie au niveau des couches supérieures afin de réaliser une gestion efficace des ressources, ceci afin de respecter les exigences strictes (QoS) des nouveaux services. Nous couvrons plusieurs aspects liés à "l'optimisation réseau"; réaliser une allocation efficace des ressources qui maximise le débit tout en assurant l'équité entre tous les utilisateurs, et ce, en fonction des conditions de canal.

Nous avons également mis l'accent sur le choix de la meilleure méthodologie permettant de choisir les outils mathématiques appropriés, l'architecture la plus efficace, des technologies adaptatives sur les couches supérieures, ainsi que la meilleure approche de conception "cross-layer".

Mots clés : Satellite – Télécommunications - Cross-layer - Bande Ka – Multidiffusion – DVB-S2 – Multimédia – Optimisation

Cross-Layer design and methodology for satellite broadband networking

Communication satellites emerge as an attractive solution in providing broadband connectivity to a variety of users thanks to its inherent global coverage. The broadcast nature of satellites (with higher frequency bands Ku/Ka) makes them the natural choice for multicasting services, for interconnecting geographically high-speed networks, and for providing multimedia services. However, this higher frequency band imposes challenging channel conditions. To avoid such problem, upcoming broadband satellite networks have adopted the adaptive coding and modulation at the physical layer. This is the main driver of this thesis due to the crucial fact that such adaptivity makes traditional satellite system design totally inefficient.

In this thesis, we focus on a different paradigm to address such new challenges, based on a joint optimization across layers of the protocol stack. The fundamental idea behind this concept is the fact that adaptivity at the physical layer should be followed at upper layers in order to achieve efficient management of the system resources, and in order to comply with the stringent QoS of new applications services. We cover several aspects related to the networking optimization; allocating resources efficiently, maximizing the throughput and assuring fairness among all the users, according to channel condition.

Our efforts have been also focused on choosing the best methodology in terms of selecting; efficient mathematical tools, most suitable architecture, novel adaptive technologies at higher layers, the best approach to the cross-layer design, and truly available standardized tools.

Keywords: Satellite – Telecommunications – Cross-layer – Ka-band – Multicasting – DVB-S2 – Multimedia - Optimization