UNIVERSITÀ
DEGLI STUDI
FIRENZE

Submitted for the Degree of
Doctor of Philosophy in
Informatics, Systems and Telecommunications
XXVI Cycle

Resource Allocation Frameworks for Broadcast and Multicast Service Delivery

Academic Discipline Code ING-INF/03

Author .....
Andrea Tassi

Thesis Supervisor .....
Prof. Romano Fantacci

Ph.D. Program Director .....
Prof. Luigi Chisci

Years 2011-2013

This thesis has been co-supervised by

Dr. Francesco Chiti and Dr. Dania Marabissi
Department of Information Engineering,
University of Florence,
Firenze, IT

Prof. John S. Thompson
Institute for Digital Communications, School of Engineering,
University of Edinburgh,
Edinburgh, UK

Troppo accuratamente l'ho studiato
senza saperne nulla. Un imprevisto
è la sola speranza.

(Eugenio Montale)

Qualunque cose tu dica o faccia
c'è un grido dentro:
non è per questo, non è per questo!

E così tutto rimanda
a una segreta domanda:
l'atto è un pretesto.

(Clemente Rebora)

Acknowledgements

I would like to thank my supervisor, Romano Fantacci, who introduced me to the academic world. In addition, I am grateful to all my co-authors, without whom this work would have been impossible.

I owe my deepest and earnest gratitude to Fabio Schoen, John S. Thompson, Chadi Khirallah and Dejan Vukobratović for their precious advice.

I am truly thankful to the friends that I met in the Advanced Communication Laboratory (University of Florence, Italy) for their constant help and support.

I would like to thank Andrew and Joyce Rosie because they welcomed me like I was family. Furthermore, I am very grateful to my friends who helped and encouraged me throughout my Ph.D. studies.

I would like to extend my deepest gratitude to my family. Without their encouragement, support and love, I would not have a chance to conclude my Ph.D.

I would like to thank the person who has shaken my everyday life by a desirable and sweet chaos.

Abstract

Delivery of Point-to-Multipoint (PtM) services over 4G cellular networks is gaining momentum. This thesis focusses on two different broadcast/multicast service types: fully reliable and delay sensitive services. The first category imposes that each PtM communication is delivered in an acknowledged fashion. On the other hand, the delay sensitive category embraces all those services aiming at broadcasting and multicasting, in an unacknowledged way, multimedia traffic flows (such as layered video services belonging to the H.264/SVC family).

For what concerns fully reliable services, this thesis proposes a Modified HARQ scheme characterized by a minimum energy consumption and reduced delivery delivery. Furthermore, in a similar system model, we propose an optimized error control strategy based on the Network Coding (NC) principle. Also in that case, the proposed strategy aims at minimizing the overall transmission energy and significantly reducing the communication delay.

In addition, we propose multiple NC-based broadcast/multicast communication strategies suitable for delay sensitive services. We prove that they can efficiently minimize either the transmission energy or delivery delay. In particular, this thesis also refers to video service delivery over 3GPP's LTE and LTE-A networks as eMBMS flows. We address the problem of optimizing the radio resource allocation process of eMBMS video streams so that users, according to their propagation conditions, can receive services at the maximum achievable service level in a given cell (depending on their propagation conditions). Developed resource allocation models can minimize the overall radio resource footprint.

This thesis also proposes an efficient power allocation model for delay sensitive services, delivered by the NC approach over OFDMA systems. The developed allocation model can significantly reduce the overall energy footprint of the transmitting node.

Contents

List of Figures **xv**

Introduction **xvii**

1 State of the Art **1**

1.1 Rate Allocation Strategies 1
1.2 Power Allocation Strategies 5
1.3 Background of LTE and LTE-A Standards 8
1.3.1 EPS and eMBMS Data Delivery 10
1.3.2 eMBMS in LTE and LTE-A 10
1.3.3 RAN Protocol Stack and Radio Resource Modelling 12

2 Acknowledged Rate Allocation Strategies **15**

2.1 Efficient HARQ Scheme for Multicast Communications 16
2.1.1 MHARQ-SC Approach 17
2.1.1.1 Performance Evaluation 18
2.1.1.2 Optimization Procedure 22
2.1.2 Analytical Results 23
2.2 NC Error Control Strategies with Symbol Combining 27
2.2.1 Background and Related Works 29
2.2.1.1 RLNC Communication Strategy 29
2.2.1.2 Related Works 30
2.2.2 Symbol Combining NC Principle 30
2.2.2.1 Application to the Butterfly Network Topology 35
2.2.3 Preliminary Results 37
2.2.3.1 Broadcast Network Scenario 37
2.2.3.2 Butterfly Network Scenario 40
2.2.4 Short Message Transmission over Satellite Networks 42
2.2.4.1 System Model 44

2.2.4.2 MNC Principle 44
2.2.4.3 Numerical Results 48
2.2.5 Cross-Layer Design of Reliable Multicasting in LTE 50
2.2.5.1 System Model 51
2.2.5.2 Energy Efficient E-RLNC Scheme 53
2.2.5.3 Numerical Results 56

3 Unacknowledged Rate Allocation Strategies 61
3.1 Standard Agnostic Rate and Energy Optimization 62
3.1.1 Optimization Frameworks 63
3.1.1.1 CP-MNC Model 65
3.1.1.2 CR-MNC Model 67
3.1.2 Numerical Results 67
3.2 General Allocation Model for Layered Video Broadcasting 70
3.2.1 H.264/SVC Broadcasting over LTE-A 71
3.2.2 System Model 72
3.2.3 Rate-Optimized and Coverage-Aware Model 75
3.2.4 Numerical Results 78

4 Unacknowledged Power Allocation Strategies 85
4.1 Power Allocation for Heterogeneous Broadcasting 85
4.2 System Model 86
4.3 Optimization Model 88
4.4 Numerical Results 92

5 Conclusions 95

Appendix A PEP of SC Communications 97
A.1 AWGN Propagation Conditions 97
A.2 Slow Rayleigh Fading Conditions 98

Appendix B Introduction to the AMC Modeling 101
B.1 Resequencial Delay Model 101
B.2 Absorbing Markov Chain Analysis 103

Appendix C Proofs of Proposition 2.1 and 2.2 109
C.1 Proof of Proposition 2.1 109
C.1.1 AWGN Propagation Conditions 109
C.1.2 Slow Fading Propagation Conditions 110

C.2 Proof of Proposition 2.2 110
C.2.1 Rician Propagation Conditions 110

Bibliography **113**

List of Figures

1.1 Evolved Packet System. 10
1.2 Evolved Packet System which includes the main eMBMS entities. . . . 12
1.3 LTE communication stack. 13

2.1 Multicast (broadcast) network model. 17
2.2 MHARQ-SC scheme from the point of view of the AP. 18
2.3 AMC associated to the packet transmission process for a multicast network with two receiving nodes. 21
2.4 Optimal values of m vs. γ in the case of AWGN propagation conditions (scenario 1). 24
2.5 Optimal values of m vs. the minimum values of γ_i in the case of AWGN propagation conditions (scenario 2). 24
2.6 Normalized mean multicast delivery delay as a function of γ in the case of AWGN propagation conditions (scenario 1). 25
2.7 Normalized mean multicast delivery delay as a function of the minimum value of γ_i in the case of AWGN propagation conditions (scenario 2). . 25
2.8 Optimal m values as a function of $\bar{\gamma}$ in the case of faded propagation conditions (scenario 1). 26
2.9 Optimal m values as a function of the minimum value of $\bar{\gamma}_i$ in the case of faded propagation conditions (scenario 2). 26
2.10 Normalized mean multicast delivery delay for information packet as a function of $\bar{\gamma}$ for faded propagation conditions (scenario 1). 27
2.11 Normalized mean multicast delivery delay for information packet as a function of the minimum $\bar{\gamma}_i$ for faded propagation conditions (scenario 2). 27
2.12 Normalized mean multicast delivery delay as a function of M for $\bar{\gamma}_i = 3$ dB. 28
2.13 Butterfly network model. 36
2.14 Optimal SDI factors vs. the mean SNR among users (AWGN propagation conditions and scenario I). 39

2.15 Normalized mean delivery delay vs. the mean SNR among users (AWGN propagation conditions and scenario I). 39
2.16 Normalized mean delivery delay vs. the number of receiving nodes for AWGN propagation conditions (scenario II). 40
2.17 Optimal SDI factors vs. the mean SNR among users in faded propagation conditions and scenario I 41
2.18 Normalized mean delivery delay vs. the mean SNR among users (faded propagation conditions and scenario I). 41
2.19 Normalized mean delivery delay vs. the number of receiving nodes for faded propagation conditions (scenario II). 42
2.20 Mean normalized end-to-end delivery delay as a function of the SNR (or mean SNR) values at the C and D sides. 43
2.21 Considered GEO satellite system model. 44
2.22 Normalized mean broadcast delay of RLNC and MNC vs. the average SNR value (Rician propagation conditions). 49
2.23 Normalized mean broadcast delay of RLNC and MNC vs. the number of receiving nodes (Rician propagation conditions). 49
2.24 E-RLNC optimized architecture for eNB and UE. 51
2.25 Value of m as a function of $\overline{\gamma}_h$ value (scenario A). 57
2.26 Mean energy consumption as a function of $\overline{\gamma}_h$ value (scenario A). . . . 57
2.27 Mean energy consumption as a function of $\tilde{\gamma}$ value (scenario B, $M = 30$). 58
2.28 Mean energy consumption as a function of number of UEs belonging to the MG (scenario B). 58

3.1 Normalized overall transmission energy cost and delivery delay vs. average SNR of the UE experiencing the worst propagation conditions. . . . 69
3.2 Normalized overall transmission energy cost vs. number of UEs. 69
3.3 LTE/LTE-A protocol stack and (a part of) radio frame (for $L = 2$). . . 73
3.4 Resource load index as a function of $\hat{N}_{\mathrm{TH}}$. 80
3.5 Video delivery probabilities and maximum PSNR of the stream A. . . . 81
3.6 Video delivery probabilities and maximum PSNR of the stream B. . . . 82

4.1 Example of the considered framed communication delivery for $S = 3$. . 87
4.2 Normalized overall transmission energy vs. minimum value of $\overline{\gamma}_{o,h}$. . . . 92
4.3 Maximum value of m_s among S PtM services vs. minimum value of $\overline{\gamma}_{o,h}$. 93
4.4 Normalized overall transmission energy vs. number of UEs M. 93

B.1 State transition diagram. 102

Introduction

According to recent forecasts [1], global IP traffic has increased more than four-fold in the past five years, and will increase three-fold over the next five. In addition, half of all IP traffic will be originated by "non-PC" devices. Finally, traffic conveyed over wireless networks will exceed that delivered through wired systems by 2016.

Among the plethora of services delivered by wireless telecommunication networks, it is worth mentioning those ones which target multiple users at the same time. Generally speaking, this thesis deals with services relying on the Point-to-Multipoint (PtM) communication pattern. In particular, we refer to the *fully reliable* and *delay sensitive* service type. Consider the first one, we defined it refers to services (delivered in a broadcast or multicast fashion) which have to be successfully received by a predefined group of users. Hence, if a user recovers the delivered information message, he has to transmit an acknowledgement message to the transmitting note. It is straightforward to note that the transmission of that kind of services is feasible only if the number of receiving users is small. In spite of the aforementioned scalability issues, we would like to note that fully reliable services are of paramount importance whenever users have to receive cryptographic session keys, mission critical communication alarms, etc.

For what concerns delay sensitive services, this thesis focusses on video content delivery. In particular we refer to video services encoded by using the H.264/SVC [2] codec. An H.264/SVC service consists of multiple video layers, namely, the *base layer* and several *enhancement layers*. The base layer provides basic reconstruction quality which is gradually improved by decoding subsequent layers [3]. Unlike fully reliable services, the transmission of a delay sensitive flow should not be acknowledged by users. As a consequence, one user usually could not have any chance to ask for for the retransmission of any part of the service stream. Finally, due to the fact that delay sensitive services are delivered in an unacknowledged fashion, the number of receiving nodes does not impact on the complexity of the transmission process.

In this thesis, we mainly refer to fourth generation (4G) mobile cellular networks, namely Long Term Evolution (LTE) and LTE-Advanced (LTE-A) networks, because they are able to deliver PtM services by means of a dedicated framework, called evolved

Multimedia Broadcast Multicast Service (eMBMS) [4]. Finally, even though we referred to specific communication standards, we would like to note that all the proposed resource allocation frameworks can be easily implemented in any modern Orthogonal Frequency-Division Multiple Access (OFDMA) communication system.

This thesis consists of four chapters in which we give: (i) the necessary theoretical background on the resource allocation issue, (ii) a detailed description of the resource allocation frameworks suitable for fully reliable and delay sensitive applications, and (iii) a general energy allocation model for delay sensitive services (delivered over a OFDMA-based communication system).

Chapter 1 provides a general state of the art about the resource allocation issues in telecommunication networks. Furthermore, it gives a complete taxonomy of the two main research branches presented in this thesis, namely the rate and power resource allocation topics. Finally, the chapter gives a quick overview of the LTE/LTE-A standard and the eMBMS framework.

Chapter 2 proposes multiple rate allocation strategies suitable for fully reliable services. All the developed allocation models address the reliability issue of PtM data delivery over modern OFDMA-based communication network. In particular, the chapter proposes an efficient Hybrid Automatic Repeat reQuest (HARQ) scheme (suitable for multicast communications) which has been optimized in order to reduce the mean packet delivery delay and energy consumption. Moreover, the chapter deals with a the use of the Network Coding (NC) principle as an error control strategy. In particular, we propose a modified Random Linear NC (RLNC) communication scheme where the transmission rate of each packet is properly optimized. The goal of the proposed optimization is to minimize the average number of transmissions and energy consumption per-transmitted packet.

Chapter 3 focusses on delay sensitive services. In particular, it describes a couple of broadcast/multicast communication strategies which can improve system performance both in terms of transmission energy and delivery delay. These goals are achieved in two ways: (i) by optimally selecting the transmission data rate while the power associated to transmission of each packet is kept constant, or (ii) by optimizing the transmission power cost and keeping constant the transmission rate. In addition, we present a novel radio resource allocation process suitable for delivery H.264 Scalable Video Coding (H.264/SVC) video services over eMBMS networks by using the RLNC principle. To this end, we propose a model that can jointly optimize Modulation and Coding Scheme (MCS), transmission rate and NC scheme used to deliver each H.264/SVC video layer (to heterogeneous set of user groups). A key aspect of the described system model is that the allocation scheme we propose does not rely on any user feedback.

In Chapter 4, we propose a standard-agnostic energy efficient resource allocation model suitable for delay sensitive services. Also in this case, data flows are delivered by means of the RLNC approach. The optimization model we propose aims at minimizing the overall transmission energy by jointly optimizing the transmission power and RLNC scheme. We proved that the complexity of the general optimization model can be efficiently overcome by resorting to an heuristic strategy which provides, in a finite number of steps, good quality feasible resource allocations.

Finally, Chapter 5 gives the concluding remarks of this thesis.

Chapter 1

State of the Art

This chapter gives a detailed insight into both the rate and power allocation problem. In particular, the chapter provides useful taxonomies to characterize the proposed allocation strategies from both the operational research and communication engineering point of view. Finally, the chapter presents a general description of the telecommunication network standard which we mainly referred to in numerical results reported in the thesis.

1.1 Rate Allocation Strategies

In a telecommunication system, the matter of rate allocation represents a critical point in a network. Rate adaptation strategies have a tremendous impact both on carriers and end-users. Obviously, the fist one aims at delivering as many services as he can differentiate his business offer to that of the others. On the other hand, users demand reliability in communications.

If we look at the reliability of a communications with a wider perspective we cannot separate that concept to the Quality of Service (QoS) problem. With other words, an information flow directed to one or more users, not only has to be successfully received with a certain probability but also it has to be delivered on time (in order to be fully enjoyable or even, intelligible by the communication ends). That is a common problem in Voice over IP (VoIP) communications or during the delivery of video contents. In these cases, the end-user QoS is negatively affected by the communication jitter of by fluctuations in the communication throughput.

Han *et al.* [5] classify the rate allocation strategies into three classes:

Source rate adaptation - As for an information source point, it is in charge of handle, encode and deliver to the other elements of the communication chain, the raw

information stream. In the case of a raw voice communication, the source has to efficiently select a source coding strategy in order to reduce the communication footprint, for e.g., by encoding (or even suppressing) the silence periods, etc. On the other hand, in the case of video content delivery, the source point has to manage the bursty nature of information flows. In this case, the encoded video stream has to potentially meet Quality of Experience (QoE) constraints of a plethora of users who access to the network by using devices characterized by different capabilities. To this aim, modern communication networks usually deliver video services in a multi-resolution fashion such that one user can access to the video service which meets the capabilities of the device in use. That goal is achieved, for e.g., by transmitting the same content at different quality levels or by delivering a basic information stream (characterized by a low quality level) whose perceived quality can eventually be improved by (using one or more) enhancement streams (which are independently transmitted). In these cases, it is straightforward to note that, any rate adaptation strategy has to jointly optimize the source transmission rate of each video stream.

Network and Media Access Control (MAC) layer optimization - The network later, as well as the MAC layer, acts as a kind of logic interface between the information source, which aims at delivering the (encoded) information stream, and the lower protocol layers, which have to reliably deliver a stream of bits representing data coming from several application services. All the modern communication standards usually assume that the network layer, or more likely the MAC layer, is in charge of scheduling and reliably delivering information streams. For these reasons, a rate adaption strategy, which takes place in one of these layers, has to mediate between three factors: (i) the QoE that is experienced by the end-users, (ii) the actual transmission rate of each communication links, and (iii) the retransmission delay which is eventually caused by ARQ or HARQ protocols.

Physical layer optimization - That can be considered as the last chance for a communication system to implement a form of rate adaptation. All the modern wireless communication standards adapt the MCS used to deliver data to one user. In the case of a Point-to-Point (PtP) communication, each User Equipment (UE) notify to the base station the MCS which meets its needs. Hence, it appears as a kind of deterministic optimization strategy where one UE choses, independently from the others, the transmission rate of communications directed to itself. However, it is straightforward to note that things significantly change as soon as the base station tracts the UE feedbacks as inputs of a rate adaptation strategy

which optimize a global utility function that, for instance, represents the quality perceived by all the users in the network.

Chapters 2 and 3 propose several rate allocation strategies suitable for fully reliable and delay sensitive services, respectively. The proposed allocation strategies focus on the rate optimization issue in broadcast and multicast network scenarios. Moreover, according to the aforementioned taxonomy, it is worth noting that all the rate allocation solutions we propose belong to either the MAC or physical layer-related optimization classes. This choose originates from the following practical observations:

- A fully reliable communication could be used to deliver short mission critical alarm messages, cryptographic session keys, etc. Hence, they are messages with a predefined format that a network operator should not alter.

- An hypothetic network operator cannot usually access to the raw multimedia service which aims at delivering. Multimedia content providers (such as YouTube, the BBC's iPlayer TV, etc.) encode, for instance, all the TV or radio programs independently to the needs of the actual network conditions experienced in any network. In addition, even though core infrastructure of 3G and 4G communication networks can (virtually) re-encode a service before delivering it, all these operations may not be possible because of their computational load or legal terms imposed by the content provider itself.

- Assuming that a network operator can re-encode a stream, that operation can improve the QoE of each network user only if it takes into account: the scheduling strategy implemented at the MAC layer, the actual propagation conditions experienced by the network clients, the specific ARQ and HARQ schemes in use, etc. To the best of our knowledge, we are not aware of any resource allocation strategy such that: (i) impacts on both raw information sources and all the MAC/PHY layer-related issues, and (ii) is characterized by a reduced computational load.

Due to the fact that this thesis also refers to 3GPP's LTE and LTE-Advanced network scenarios[1], the rest of this chapter reviews the main rate allocation strategies which meet the tight constraints imposed by LTE standard. Among several survey papers, Afolabi *et al.* [6] delivered one of the most detailed review of the main rate allocation-based resource allocation techniques. In particular, the authors note that modern 3G and 4G systems can optimize the transmission rate of PtM services according to the following approaches:

[1]With the only exception of Section 2.2.4.

Single-Rate (SR) - Each broadcast or multicast service is delivered to all UEs using the same transmission rate. One example is the Pre-defined Fixed Rate (FR) strategy [7], where each service is always delivered at the same rate regardless of propagation conditions experienced or quality level perceived by UEs. However, Section 1.3.2 shows that the 4G core network can control the transmission rate of delivered multicast/broadcast services. Hence, in this case the rate can be (virtually) optimized according to the Least Channel Gain (LCG) strategy [8], i.e., the rate can be optimized such that a broadcast/multicast flow is transmitted with a data rate that the UEs in the worst propagation conditions (not necessarily the cell-edge UEs) can experience the desired QoS level. Finally, the idea underlying the LCG strategy can be extended as proposed in the Average Group Throughput (AGT) approach [9] where the transmission rate is optimized such that, on average, UEs will experience a target quality level.

Multi-Rate (MR) - It is straightforward to note that a SR strategy cannot exploit the full potential of a 4G communication network. UEs that could benefit from an higher transmission rate possibly stuck at low QoS levels. The Information Decomposition Techniques (IDT) [10] split each multicast flow into several subflows so that UEs in the worst channel conditions would be able to decode the lowest channel rate substreams (i.e., lowest QoS), while UEs in a better channel conditions can decode, beside the basic substream, the higher data rate substreams, and eventually combine those substreams to yield high QoS results. With other words, in contrast to the FR allocation techniques, UEs experiencing ideal propagation conditions are no longer penalised, in terms of achieved rates/QoS, by the UEs in the worst channel conditions. It is worth noting that a pure IDT strategy mainly deals with the problem of the source coding optimization. That potential limitation can be overcome by Multicast Subgroup Formation (MSF) strategies which aim at optimizing the transmission rate of each subflow over a subset of the communication end points. In this case, we assume that the multimedia content provided encodes the service in a layered fashion (see Section 3.2.1), independently form the network provider. A similar multi-rate approach for video delivery over broadband cellular networks has been investigated in [11, 12].

Chapter 2 proposes both SR and MR strategies which deal with the following issues:

Rate allocation as a reliability control problem - The problem of delivery broadcast/multicast services such that all the communication ends (on average) meet tight QoS constraints can be summarised in this way: Is there a multicast protocol suitable for reliable communications where reliability of communications can

be practically optimized by changing the transmission rate? Chapters 2 and 3 aims at both defining and optimising multiple of reliable communication scheme for PtM transmissions.

Definition of an LTE/LTE-A-based MR allocation strategy - There are a plethora of rate allocation strategies that promise to be suitable for PtM service delivery in 3G/4G contexts. However, in most cases they just refer to a generic OFDMA-based communication protocol. To this end, in Chapters 2 and 3 we propose rate allocation strategies fully integrated into the LTE/LTE-A communication stack. We will show that the aforementioned allocation strategies are suitable for service multicasting and broadcasting over eMBMS networks.

1.2 Power Allocation Strategies

In a wireless communication network, power allocation issues plays are of paramount importance. In particular, the matter of power adaptation impacts on several aspects, such as: the cochannel interference reduction, maximisation of the network capacity, improvement of the service quality, reduction of a ecological footprint of the telecommunication infrastructure, etc.

Unlike rate allocation strategies, a power allocation approach primarily aims at improving the quality of communications link between transmitter and receiving ends. Of course, a somewhat optimal power allocation can be constrained to meet certain QoS requirements (as well as usually happens with a rate adaptation strategy). Moreover, in this case the focus of the optimization model should be on the quality of the access links. Hence, the Signal to Interference plus Noise Ratio (SINR) is one of the most important performance index in an optimization framework.

It is worth noting that, the SINR value can greatly fluctuate because of the combination of multiple factors, such as: the path loss, shadowing and small-scale fading. A power allocation scheme should regularise those fluctuations and keep the SINR level above the *minimum protection ratio*. The minimum protection ratio is a threshold value which depends on the target QoS level associated to the delivered service, on the access priority level of the communication end, etc. Han *et al.* [5] identify several issues that a power allocation model may take into account:

- Optimising in uplink vs. downlink - There is a plethora of power allocation strategy which focus on the uplink phase [13] but they are out of the scope of this thesis. In fact, this thesis focuses on the power allocation strategies that take place at the base station side.

- Increasing the SINR of an user vs. minimizing the cochannel interference - if the SINR of an user increases, it will necessarily increase the interference level which impacts on all the other network users.

- Centralising vs distributing the allocation strategy - A resource allocation algorithm can be either centralised or distributed. In this thesis we focus on the centralized approaches (see Chapter 4).

- To pursuit of the optimum solution vs. To settle of a suboptimal one - Usually, a resource allocation problem is nonlinear and characterized by a nonconvex feasible set. In addition, the objective function and/or constraints are non-differentiable. Hence, we have to resort to derivative free optimization strategies. In most cases, it is hard (or even impossible) to derive the optimum allocation but there are chances to develop heuristic strategies which can find good suboptimal solutions.

- To consider service related QoS constraints or not - From a theoretical point of view, a power allocation strategy should primarily aim at keeping the user SINR above the minimum protection ratio. However, that threshold value can vary not only on a per-user basis but also on a per-service basis (i.e., the minimum protection ratio can be a function of the user priority and QoS level associated to the delivered service). Hence, in most cases the service-related constraints cannot be considered as minor ornamentals of the optimization problem, as shown in Chapter 4.

there are several power allocation strategies which can be trivially classified as: uplink/downlink or centralized/distributed approaches. However, it is worth defining a couple of different taxonomies [5]:

Acknowledged vs. unacknowledged taxonomy - In this case the focus is on the protocol used to practically optimize the transmission power.

Closed-loop strategies - As for a downlink allocation strategy, it belongs to the closed-loop family if the power allocation is periodically updated by the base station on the basis of feedback provided by UEs. Such kind of allocation strategies cause a communication overhead that could not be negligible.

Open-loop strategies - These strategies take place when the user feedback cannot be managed by the base station or when propagation conditions vary too rapidly to be addressed by a closed-loop strategy. In this case, each client estimates the reception quality of the downlink channel and that measurement entirely characterizes the uplink communication link. The aforementioned

measurements are transmitted to the base station with a periodicity that is significantly smaller than that characterising a closed-loop scheme.

Combined Open- and Closed-loop strategies - Such kind of strategies provide a power allocation by using both an open- and closed-loop strategy. Usually, an open-loop phase is followed by a closed-loop one.

Performance index driven taxonomy - Objective functions and constraints of a power allocation model are expressed as a function of a performance index which can be one of the following ones:

Power-strength - By measuring the base station signal strength at the receiving end. Even though, that performance index can be measured, modern communication standards usually express quality level measurements in terms of SINR or information error rate (that can be measured by different levels of the communication stack).

SINR based - This is the performance index which is commonly used in 3G and 4G communication systems to describe both the quality and capacity of a communication link.

Error rate based - Each protocol layer can measure its own error probability, such as: the Bit Error Rate (BER) at the physical layer, the Transport Block Error Rate (TBLER) at the MAC layer, etc. Hence, the power allocation model in use can be restated as a function of these high level performance indexes.

Chapter 4 proposes a power allocation strategy which is suitable for broadcast and multicast communications. In particular, we propose an open-loop allocation model (suitable for delay sensitive services) which aims at: (i) minimizing the overall transmission energy of a base station, (ii) providing a feasible power allocation for PtM service delivering, and (iii) efficiently tacking into account all the QoS constraints.

In spite of the power optimization topic has a very long tradition, a little attention has been paid on the downlink transmission power reduction for networks delivering PtM services. To the best of our knowledge there are a couple of allocation strategies which are suitable for PtM service delivery [14]:

Margin Adaptive (MA) strategy - The model aims at minimizing the total transmission power over a feasible set defined by the desired users' data rate.

Rate Adaptive (RA) strategy - Unlike the MA approach, the model maximises the communication throughput and puts constraints on the overall transmission

power.

Unfortunately, due to the fact that we aim at minimizing the overall transmission energy (under some QoS constraints), the power allocation strategy we propose does not belong to neither the MA class nor RA one.

In addition, Yuan *et al.* [15] inspected the problem of deriving the optimal solution of the power allocation problem and pointed out two important issues:

- Most of the power allocation strategies just focus on heuristic strategies which comprise of a two-step procedure (such as [16]). Such kind of procedure firstly aims at matching the users' rate constraints and then at optimising the transmission power.

- The user's rate (r) is expressed in terms of a logarithmic function $r(P) = \log(1 + gP)$, where g is the channel gain and P is the transmission power. In addition, authors of [15] prove that, for $g > 0$, each MA strategy is NP-hard.

To this end, Chapter 4 overcomes the aforementioned issues by proposing a power allocation strategy which minimize the overall energy footprint of a system delivering multiple delay sensitive services. A key aspect of the developed framework is that delay sensitive services are delivered by means of the NC principle. Finally, we show that the proposed model can be solved by en efficient one-step heuristic strategy.

1.3 Background of LTE and LTE-A Standards

This thesis mainly refers to the 3GPP's LTE and LTE-A standards. In particular, we focus on the eMBMS framework (see Section 1.3.2). To this end, this section provides a quick insight into the main features of LTE and LTE-A.

In the 1980s, the International Telecommunication Union (ITU) started its work on the definition and implementation of the third-generation mobile communication (3G) which after a couple of decades became known as International Mobile Telecommunications-2000 (IMT-2000). In addition, in the 1990s, several organisation (such as ETSI, ARIB, etc.) all around the world defined and implemented several CDMA-based communication systems, among them there was the Universal Mobile Telecommunications System (UMTS). The independent evolution of Wide CDMA (WCDMA) systems continued until the end of 1998 when Third Generation Partnership Project (3GPP) was founded. 3GPP aimed at delivering a common standardisation framework for the future evolutions of the WCDMA standards proposed until that

point. It is worth noting that at the very beginning, the 3GPP's work benefited from the ITU-R's standardisation activity of 3G systems.

Within ITU-R, the Working Party 5D (WP5D) is in charge of listing a set of satellite and terrestrial radio interfaces (such as DECT, WiMAX, UTRA, E-UTRA) suitable for both IMT-2000 and IMT-Advance. In addition, the WP5D provides a set of recommendations that should reflect the current and future directions of all the overall standardisation body. In any case, the WP5D does not provide any technical requirement or specification. In addition, WP5D is providing a set of recommendations also for those standards which are beyond IMT-2000, namely the communication systems belonging to the IMT-Advanced (which is the gateway to the 4G communication era). The LTE (until 3GPP's Release 9) belongs to the IMT-2000 set. On the other hand, from 3GPP's Release 10, LTE, called LTE-A, takes part in the IMT-advanced body.

It is worth noting that LTE, as well as LTE-A, are standards for the Radio Access Network (RAN). In particular, from an architectural point of view, a LTE-based network (including the core network) has a "flat radio-access" [4]. The radio access part of the system consists of just one entity, which is the eNodeB (eNB). However, unlike WiMAX, 3GPP started the standardisation process of the core network (namely, the Evolved Packet Core, EPC). ECP and RAN form the Evolved Packet System (EPS).

From a low level perspective, LTE-based systems rely on the well-known OFDM modulation format for both the downlink and uplink phase. Hence, all the available radio resources can be modelled, from a logic point of view, as a frequency/time matrix whose elements hold a variable amount of downlink/uplink data.

Namely, each UE measures and reports to the eNB the quality of the downlink channel by checking a set of reference signals placed along the communication frame. Hence, each eNB periodically collects Channel Quality Indicators (CQIs) from all the served UEs.

LTE can be operated both in Time Division Duplexing (TDD) as Frequency Division Duplexing (FDD). In addition, it can be used to deploy a multi-cellular network characterized by a frequency reuse factor equal to one. That means that the same frequency/time resource can be virtually used for uplink or downlink communications in two (or more) neighbouring cells.

LTE-based systems can manage multi-antenna transmission/reception schemes. In particular, both eNBs and UEs have to be equipped with at least two difference receiving antennas. Hence, they can exploit the channel diversity and adopt any beam-forming based communication strategy.

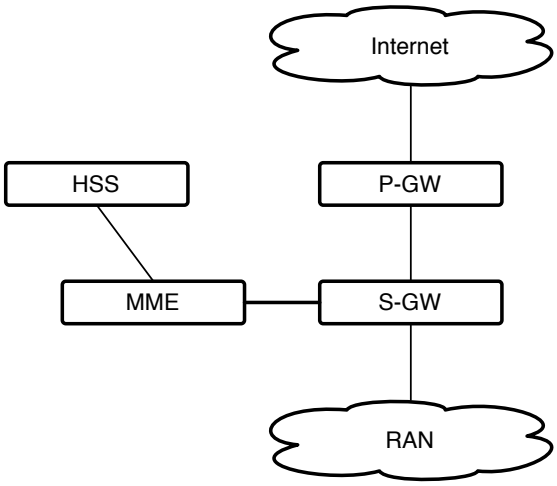


Figure 1.1: Evolved Packet System.

1.3.1 EPS and eMBMS Data Delivery

The EPS comprises the entities below (sketched in Figure 1.1):

eNodeB - The RAN consists of just is entity. It is worth noting that it is a logical block. Hence, it can be modelled as a classic Base Station (BS) or as a central entity managing several (independent) radio heads.

MME - The Mobility Management Entity manages all the communication flows between UEs and the EPC. Hence, it manages users' bearers and plays a key role in the LTE security framework [4].

S-GW - The Serving Gateway acts as a gateway between the RAN and the EPC. In addition, it tracks the position of UEs and collects the billing-related information.

P-GW - The Packet Data Network Gateway is the Internet gateway of the EPC. It gives an IP address to each UE.

For the sake of clarity, the EPC also comprises other entities. All those entities have not been reported because are related to aspects that are out of the scope of this thesis.

1.3.2 eMBMS in LTE and LTE-A

Starting form 3GPP's Release 6, the MBMS framework has been proposed to efficiently deliver multimedia services in a broadcast and multicast fashion [4]. 3GPP's Release 9 and 10 afterwards foresee an enhanced version of the MBMS framework (namely, the eMBMS) which specifies two designs:

- The Single Cell-eMBMS (SC-eMBMS) - Each eNB of the network delivers multicast (or broadcast) services independently from the others. The SC design allows each eNB to independently select a suitable MCS that ensures reliable transmission to all users in the cell.

- The Multicast-Broadcast Single Frequency Network (MBSFN) eMBMS - Two or more eNBs are grouped into MBSFNs. All the eNBs of the same MBSFN are synchronised and transmit data with the same physical signals. The MBSFN-eMBMS mode improves the reliability of data delivery using coordinated transmissions from multiple cells. This mode is very beneficial for users positioned close to the cell-edge. Usually, MCSs of delivered services are chosen by the core network and are same for all eNBs [17, 4]. For the sake of clarity, each eNB provides a same resource allocation and broadcasts exactly the same content (at the same time). eMBMS transmissions received by a UE in a MBSFN area appear as it were transmitted by one eNB. Hence, the UE collects multiple copies of a same transmission, affected by different communication delays[1]. Inter-Symbol Interference (ISI) effects are mitigated by OFDM cyclic prefixes which are longer than those used in non-eMBMS communications. For these reasons, the UE can manage all the copies of the same OFDM symbol as the where the result of a multipath communication.

Let us consider a group of eMBMS-capable eNBs. They define a "MBMS Service Area". A MBMS Service Area consists of one or more "MBSFN Synchronisation Areas". A Synchronisation Areas is composed by a set of eNBs which are mutually synchronised. Hence, a MBMS Synchronisation Area can hold one or more "MBSFN Areas". Each MBSFN Area consists of one or more eNBs transmitting the same set of MBMS flows. It is worth noting that, an eNB can take part in multiple MBSFN Areas (but not at the same time).

The EPS can manage a broadcast or multicast service by the following entities (sketched in Figure 1.2):

BMSC - The Broadcast-Multicast Service Centre manages communication flows coming from/to a content provider that is external to the EPC.

MBMS-GW - The MBMS Gateway forwards real time services, coming from the BMSC, to the eNBs.

[1]The communication delay of each replica is both caused by the propagation delay and by the "physiological" drifts of eNB clocks.

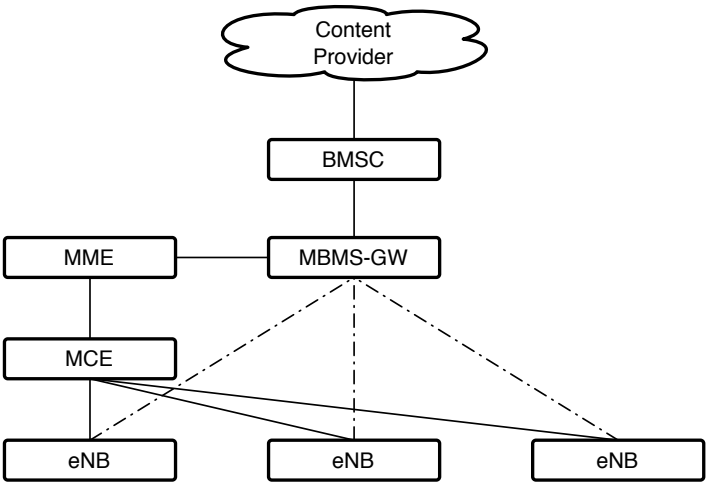


Figure 1.2: Evolved Packet System which includes the main eMBMS entities.

MCE - The Multicell Coordination Entity provides the radio resource allocation of each MBMS flow that is transmitted over any cell (in the case of SC-eMBMS) or MBSFN area (for MBSFN-eMBMS flows).

For what concerns physical aspect of eMBMS communications, it is worth noting that MBMS flows cannot be mixed to the uplink/downlink ones because eMBMS communications needs [4]: (i) a longer ciclic prefix, and (ii) a more dense reference signal pattern. Hence, only a specific group of subframes can be used to deliver eMBMS services.

1.3.3 RAN Protocol Stack and Radio Resource Modelling

The radio protocol stack of eNBs and UEs can be summarised as reported in Figure 1.3. In particular, it is composed by the following layers:

PDCP - The Packet Data Converge Protocol processes IP packets (namely, the PDCP Service Data Unit, PDCP SDU) by compressing their headers and cyphering their payloads. It also is in charge of duplicate-removal and in-order delivery processes on an end-to-end basis. A PDCP SDU is mapped on one PDCP Packet Data Unit (PDCP PDU).

RLC - The Radio Link Control segments and/or concatenates multiple RLC-SDUs (possibly belonging to different radio bearers). It implements ARQ strategies for PtP communications on an end-to-end basis.

MAC - RLC-PDUs are mapped on a set logical channels provided by the MAC layer. In addition, the MAC layer manages multiple data scheduling operations aiming

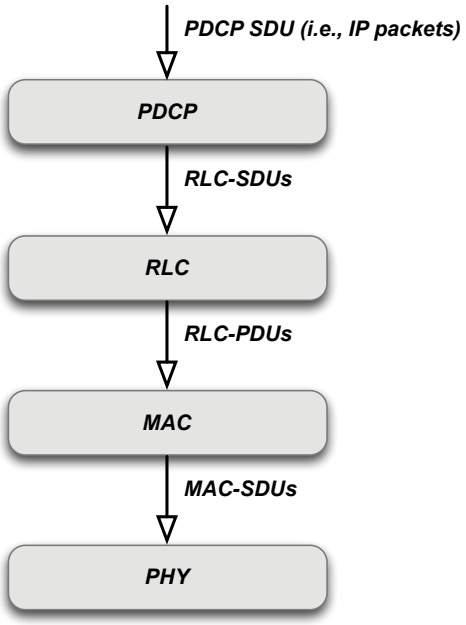


Figure 1.3: LTE communication stack.

at mapping all the logical channels on transport channels provided by the PHY layer.

PHY - The Physical layer segments/concatenates one or more MAC-SDUs in a Transport Block (TB). Each TB is mapped on physical radio resources as (indirectly) stated by the MAC scheduler.

Radio resources can be modelled as a matrix composed by Physical Resource Blocks (PRBs)[1]. A PRB is a fixed-size frequency/time structure composed by 12 contiguous subcarriers $\times$ 6 (or 7) OFDM symbols. In addition, regardless of the considered duplexing scheme, transmissions are organized in frames. A frame consists of subframes (1 ms long). The transmission time duration of a frame is 10 ms. Finally, in the case of TDD, each subframe can be assigned to the uplink, downlink phases or, it can be used to convey signalling information [4].

[1]For the sake of clarity, we will equivalently refer to PRBs also in terms of Resource Blocks.

Chapter 2

Acknowledged Rate Allocation Strategies

This chapter deals with several rate allocation strategy which address the issue of fully reliable service delivery. In particular, this chapter is organized as follows. Section 2.1 proposes a efficient HARQ scheme for multicast communications. In particular, we propose a Modified HARQ scheme based on the Symbol Combining principle (MHARQ-SC) where multiple copies of the same packet are consecutively transmitted. The proposed solution has been optimized by considering as performance metrics: the mean packet delivery delay and energy consumption (per-information packet). For the sake of comparisons, we inspected the performance of different HARQ schemes optimized for multicast communications.

In addition, the chapter deals with a particular application of the NC principle: the NC as error control strategy. Section 2.2 investigates the performance achieved by combining the Symbol Combining (SC) approach with the RLNC principle in PtM communications. In this case the transmission rate of each packet is properly optimized in order to meet different objectives. In addition to the optimal resource allocation model, we propose an efficient heuristic approach which is characterized by a reduced computational load. The performance of the proposed solutions has been validated in a broadcast network model and in classical butterfly topology network[1]. Analytical results clearly show that the proposed solutions outperform the basic RLNC alternative.

In Sections 2.2.4 and 2.2.5, the performance of the aforementioned allocation model is validated in a satellite communication system and SC-eMBMS network deployment.

[1]See Section 2.2 for a detailed definition of both the network topologies.

2.1 Efficient HARQ Scheme for Multicast Communications

Multicast communications are gaining momentum in wireless communications since they make possible delivering services to multiple users located within the coverage area of the same access node. Unlike unicast communications, multicast flows deliver the same message to several users at the same time. Multicast services include video applications, group text messaging and specific alerting messaging services (used to manage, for e.g., emergency situations, etc.). On the other hand, it is straightforward to note that the communication delay and transmission energy associated to reliable multicast communications decrease as the number of users increases [18, 19]. In particular, the mean time required to deliver an information packet to all the members of a Multicast Group (MG) increases with the MG size. Hence, the overall energy associated to the transmission of an information packet increases as well.

It is worth noting that the bottleneck of the overall system performance usually is the node experiencing the worst propagation conditions. HARQ schemes have been initially used to increase the reliability of unicast communications [20, 21]. In particular, Lin *et al.* [20] proposed a HARQ technique which aim at iteratively broadcasting fixed-length packets composed by a couple of subpackets (characterized by the same dimensions). In this way, each subpacket can be used to deliver new information elements or to increase the amount of redundancy information related to the packet which has not been successfully received. In addition, Kim *et al.* [22] proposed a HARQ error control strategy for multicast communication based on the classic SC approach, originally proposed by Chase [23]. The proposed approach allows a receiving end to recover one information packet by soft-combining (according to the maximum ratio combining) a fixed number of copies (of the information packet itself), even those containing errors.

Differently from [22], this section deals with a Modified HARQ-SC scheme, hereafter named MHARQ-SC, relying on the continuous transmission of a given number of copies (denoted as m) of the same packet, during each transmission opportunity. Afterwards, each receiving node combines all the received copies of the same packet as stated by the SC principle. It is worth noting that the implementation complexity of the proposed MHARQ-SC is equivalent to that associated to the classical HARQ-SC scheme.

In order to derive the performance bounds for the MHARQ-SC, the rest of the section relies on a theoretical framework based on the Absorbing Markov Chains (AMC) theory [24]. The reported analysis considers both the AWGN and (frequency-non-selective) slow Rayleigh faded propagation conditions. Moreover, we present an optimization framework for the proposed MHARQ-SC scheme. The accuracy of the pro-

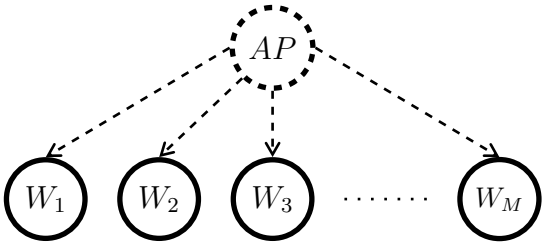


Figure 2.1: Multicast (broadcast) network model.

posed theoretical derivation will be validated by resorting to computer simulations. Finally, numerical results show that the optimized MHARQ-SC scheme outperforms the rate optimized HARQ-SC alternative proposed in [22].

2.1.1 MHARQ-SC Approach

As previously mentioned, the MHARQ-SC scheme relies on the SC principle to increase the delivery performance, in particular under poor propagation conditions. We prove that the transmission of an optimized number (m) of copies of the same packet leads the MHARQ-SC scheme to outperform the HARQ-SC alternative.

The system under investigation is sketched in Figure 2.1. In this case, a single source node, namely the Access Point (AP), transmits the same communication flow to the set $\mathcal{W} = \{W_1, W_2, \ldots, W_M\}$ of M nodes forming the MG. Each communication link is modelled as an independent lossy channel. We can imagine that one of the simplest retransmission scheme consists in retransmitting a new set of m copies of the same packet till all the nodes of the MG receive it[1] [25]. However, this solution is clearly not efficient, as a consequence, we refer to a scheme where each client acknowledges the reception of a information packet and the AP counts the number of clients which have successfully provided acknowledgements. The AP transmits the same set of m copies until all the clients successfully recover the information packet. In particular, one node of the MG recovers the information packet by using the procedure below:

1. m copies of the same packet are combined symbol-by-symbol according to the SC principle;

2. an SC detection is performed symbol-by-symbol as outlined in the Appendix A;

3. if the client recovers the information packet, it sends a positive acknowledgement message (ACK).

[1]Without any loss of generality we assumed that the acknowledgement messages are transmitted on a fully reliable channel.

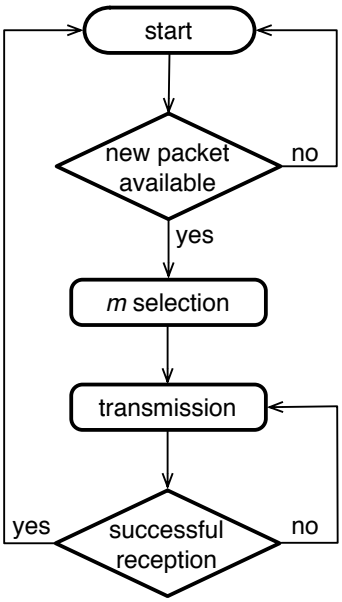


Figure 2.2: MHARQ-SC scheme from the point of view of the AP.

The algorithmic description of the MHARQ-SC scheme, from the point of view of the AP, is sketched in Figure 2.2. It is worth noting that, for $m = 1$, Figure 2.2 also defines an HARQ scheme (based on the SC principle) which we call classical HARQ-SC scheme. We remark that in the case of the MHARQ-SC approach, the receiving node performs the SC detection on the set of m copies of the same information packet. If the detection process fails, a new process start as soon as a new burst of m copies is received.

The performance of the MHARQ-SC scheme mainly dependents on m. Hence, we propose an analytical approach to derive optimum m values. In particular, we mainly focus on the minimization of the mean multicast delivery delay defined as: the time interval (normalized to the transmission time duration τ of one copy) which starts on the first transmission attempt of an information packet and finishes as soon the MG successfully receives it.

We assume that each information flow uses the Binary Phase Shift Keying (BPSK) modulation[1]. Finally, we assume that propagation conditions are statically independent during the transmission of each information packet copy.

2.1.1.1 Performance Evaluation

The performance behaviour of the MHARQ-SC scheme can be efficiently inspected by the AMC theory [24] (Chapter 3, page 43). For more information about the AMC, please refer to Appendix B which reports an instant prime on the theoretical framework

[1]Note that the derived results are quite general and they can be easily extended to different modulation schemes.

used in the rest of this section.

Let us consider the process modelling the successful reception of an information packet by the MG. Considering a scenario where all the communication channels have the same propagation conditions. Let s_i (for $i = 0, \ldots, M$) be the state of the process, defined as the number of nodes of the MG that have successful received the information packet. Hence, if the number of nodes which have successfully recovered the information packet passes from i to j (for $j \geq i$) then the process moves from the state s_i to s_j. For these reasons, s_M is the absorbing state of the process[1] (see Appendix B).

It is straightforward to note that the considered process always begins from the state s_0 and finishes as soon as it enters into the absorbing state s_M. According to the considered AMC framework, all the states $\{s_0, \ldots, s_{M-1}\}$ are transient (see Appendix B), which means that once the process leaves one of them, the process can no longer reenter it.

Let $p_{i,j}$ be the probability that a transition occurs from the state s_i to s_j (for $j \geq i$)

$$p_{i,j} = \binom{M-i}{M-j} P_B(m)^{M-j} \left[1 - P_B(m)\right]^{j-i} \tag{2.1}$$

where $P_B(m)$ is the packet error probability associated to a generic receiving end. In the Appendix A we will prove that $P_B(m)$ depends on m. Let $\mathbf{P}$ be the state transition probabilities matrix where the (i,j)-th entry, $p_{i,j}$, is given by (2.1). In particular, $\mathbf{P}$ can be reported as follows

$$\mathbf{P} \doteq \begin{bmatrix} P_B(m)^M & \binom{M}{M-1} P_B(m)^{M-1}\left[1 - P_B(m)\right] & \cdots & \left[1 - P_B(m)\right]^M \\ 0 & P_B(m)^{M-1} & \cdots & \left[1 - P_B(m)\right]^{M-1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 - P_B(m) \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{2.2}$$

Moreover, we have that the transition matrix $\mathbf{P}$ can be expressed in the canonical form as follows [24]:

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{Q} & \mathbf{R} \\ \hline \mathbf{0} & 1 \end{array}\right] \tag{2.3}$$

where

1. $\mathbf{Q}$ is the $M \times M$ transition matrix modelling the process as long as it involves

[1] If the process enters the absorbing state, it means that all the nodes of the MG have successfully recovered the information packet. So that, once the process reaches the absorbing state, it cannot be left.

only transient states

$$\mathbf{Q} \doteq \begin{bmatrix} P_B(m)^M & \cdots & \binom{M}{1} P_B(m)\Big[1-P_B(m)\Big]^{M-1} \\ 0 & \cdots & \binom{M-1}{1} P_B(m)\Big[1-P_B(m)\Big]^{M-2} \\ \vdots & \vdots & \vdots \\ 0 & \cdots & P_B(m) \end{bmatrix} \tag{2.4}$$

2. $\mathbf{R}$ is a M-dimensional column vector listing the transition probabilities of the process when it starts from a transient state and enters in the absorbing one

$$\mathbf{R} \doteq \begin{bmatrix} \Big[1-P_B(m)\Big]^M \\ \Big[1-P_B(m)\Big]^{M-1} \\ \vdots \\ 1-P_B(m) \end{bmatrix} \tag{2.5}$$

3. $\mathbf{0}$ is a M-dimensional row vector with entries all equal to zero.

The aforementioned analysis can be generalised to the case of unequal propagation conditions, i.e., it can be extended to the case where each communication channel is affected by differed propagation conditions. In this case, the i-th state of the packet reception process can be defined as a M-ary column vector $\mathbf{s}_i$ (with $i = 0, \ldots, 2^M - 1$), where the t-th entry $\mathbf{s}_i[t]$ (for $i = 0, \ldots, 2^M - 1$ and $t = 1, \ldots, M$) is equal to 1 if the node W_t has correctly received the information packet or 0, otherwise. As a consequence, the absorbing state is represented by the vector $\mathbf{s}_{2^M-1}$ which has all the components equal to one. Also in this case, the packet transmission process starts from the state $\mathbf{s}_0$ (whose entries are equal to 0) and finishes as soon as it enters into the absorbing state $\mathbf{s}_{2^M-1}$. Finally, it is straightforward to note that all the states $\mathbf{s}_i$ (with $i = 0, \ldots, 2^M - 2$) are transient.

Let us define the operator $\succeq$: the relation $\mathbf{s}_j \succeq \mathbf{s}_i$ holds if $\mathbf{s}_j[t] \geq \mathbf{s}_i[t]$ ($\forall t = 1, \ldots, M$). Let $P_{B,i}(m)$ (for $i = 1, \ldots, M$) be the packet error probability associated to the i-th node. Hence, for $\mathbf{s}_i = \mathbf{s}_j = \mathbf{s}_{2^M-1}$, the (i,j)-th entry of the transition matrix $\mathbf{P}$ is equal to one. On the other hand, if $\mathbf{s}_j \succeq \mathbf{s}_i$, the state transition probability $p_{i,j}$ can be defined as:

$$p_{i,j} = \prod_{\substack{t=1,\ldots,M \text{ t.c.} \\ \mathbf{s}_i[t]=\mathbf{s}_j[t]=0}} P_{B,t}(m) \prod_{\substack{t=1,\ldots,M \text{ t.c.} \\ \mathbf{s}_j[t]>\mathbf{s}_i[t]}} \Big[1-P_{B,t}(m)\Big] \tag{2.6}$$

In the Appendix A we derive $P_{B,i}(m)$ in the case of AWGN and frequency-non selective slow Rayleigh faded propagation conditions. As an example, the state diagram of an

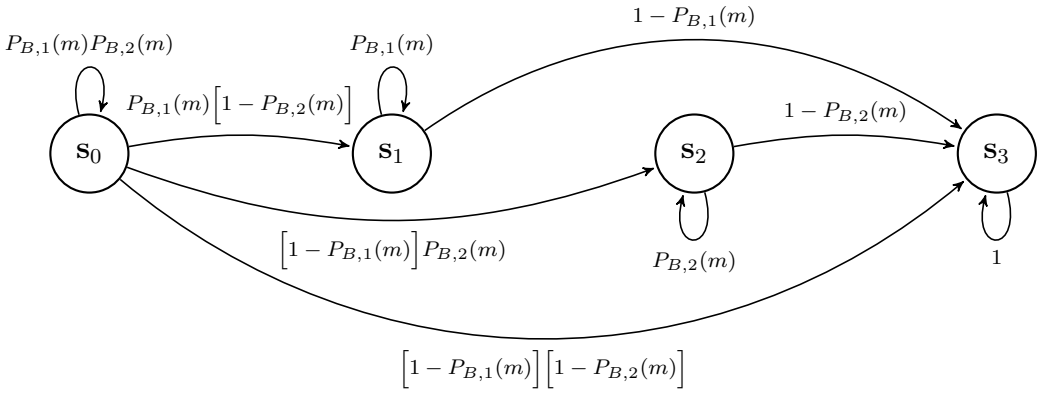


Figure 2.3: AMC associated to the packet transmission process for a multicast network with two receiving nodes.

AMC modelling the packet reception is shown in Figure 2.3.

Considering (2.6), the transition probability matrix $\mathbf{P}$ is in the canonical form (2.3). However, in this case we have:

1. $\mathbf{Q}$ is a $(2^M-1)\times(2^M-1)$ defined as

$$\mathbf{Q}=\begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,2^M-1} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,2^M-1} \\ \vdots & \vdots & \vdots & \vdots \\ p_{2^M-1,0} & p_{2^M-1,1} & \cdots & p_{2^M-1,2^M-1} \end{bmatrix} \tag{2.7}$$

2. $\mathbf{R}$ is a (2^M-1)-dimensional column vector defines as

$$\mathbf{R}=\begin{bmatrix} p_{0,2^M} \\ p_{1,2^M} \\ \vdots \\ p_{2^M-1,2^M} \end{bmatrix} \tag{2.8}$$

3. $\mathbf{0}$ is a 2^M-1 dimensional row vector with all entries equal to zero.

In oder to derive a closed form solution for the the mean multicast delivery delay, it is useful to define the fundamental matrix of the AMC (see Appendix B) as $\mathbf{N}=(\mathbf{I}-\mathbf{Q})^{-1}$, where $\mathbf{I}$ is the identity matrix having $M\times M$ dimensions in the case of equal propagation conditions or $(2^M-1)\times(2^M-1)$ components in the case of unequal propagation

conditions. Moreover, we will prove in in Section B.2 that the following relation holds:

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1} = \sum_{k=0}^{\infty} \mathbf{Q}^k \tag{2.9}$$

where the (i, j)-th component of the matrix $\mathbf{Q}^k$ is the probability that the process reaches the j-th transient state from the j-th one in k steps (i.e., after that the AP has transmitted k burst of m copies of the same information packet). In Section B.2 we will prove that the (i, j)-th element of the fundamental matrix N is the mean value of the total number of times that the process, started from the i-th state, the j-th one[1]. Hence on average, the AP has to transmits m copies of the same information packet a number of times which is

$$\zeta(m) = \sum_{l} \mathbf{N}[1, l] \tag{2.10}$$

where the term $\mathbf{N}[1, l]$ is the $(1, l)$-th element of the $\mathbf{N}$ matrix. Hence, the term $\zeta(m)$ is equal to the sum of all the elements forming the first row of the matrix $\mathbf{N}$.

2.1.1.2 Optimization Procedure

From (2.10), we have that the mean multicast delivery delay $\delta(m)$ of an information packet (normalized with respect to τ) is

$$\delta(m) = (m + s)\,\zeta(m) \tag{2.11}$$

where s denotes the round-trip delay normalized with respect to τ (i.e., the time interval required to possibly receive and acknowledge m copies of a packets). Let L and E_b be the packet length (expressed in bits) and energy per bit, respectively. The mean value (normalized with respect to LE_b) of the mean energy, $\epsilon(m)$, required to successful deliver an information packet, can be expresses as follows:

$$\epsilon(m) = m\,\zeta(m). \tag{2.12}$$

From (2.11) and (2.12), we note that the mean energy is minimized once $\epsilon(m)$ is minimized with respect to m. Hence, for the sake of simplicity, we focus our analysis on the minimization of $\delta(m)$ with respect to m. As a consequence, we define the optimal

[1]Note that both the i-th and j-th states are transient.

values of m as the solution of the following optimization problem:

$$\text{(P1)} \qquad \text{minimize} \quad \delta(m) \tag{2.13}$$

$$\text{subject to} \quad \hat{m} \in \mathbb{N}. \tag{2.14}$$

Unfortunately, P1 is an integer nonlinear optimization problem. Hence, it cannot be efficiently solved in real time. Hence, we resorted to the Mesh Adaptive Direct Search (MADS) algorithm [26]. In particular, in this case, we considered the MADS implementation provided by the NOMAD solver [27].

2.1.2 Analytical Results

This section presents numerical results concerning the performance analysis of the proposed MHARQ-SC scheme. In addition, the proposed MHARQ-SC strategy is compared with different HARQ alternatives. In particular, we considered the classical HARQ-SC scheme (Section 2.1.1) and the optimized HARQ approach proposed by Kim *et al.* [22]. In addition, the analytical predictions, of the proposed strategy also is validated by computer simulations. We considered two network scenarios:

- *scenario 1* - All the communication links have the same propagation conditions. This means that the SNR γ_i and the mean SNR $\bar{\gamma}_i$ of each receiving node (for $i = 1, \ldots, M$) is equal to γ and $\bar{\gamma}$, respectively. Both γ and $\bar{\gamma}$ can take a value in the interval $[3, 13]$ dB;

- *scenario 2* - Different propagation conditions with parameters γ_i and $\bar{\gamma}_i$ defined (for $i = 1, \ldots, M$) as

$$\gamma_i = \gamma_1 - \big(i-1\big)\eta, \quad i = 1, \ldots, M, \tag{2.15}$$

$$\bar{\gamma}_i = \bar{\gamma}_1 - \big(i-1\big)\bar{\eta}, \quad i = 1, \ldots, M \tag{2.16}$$

  where the values of η and $\bar{\eta}$ have been set equal to 0.45 dB. Hence, in this scenario the node W_1 is characterized by the best propagation conditions, while node W_M experiences the worst ones.

Moreover, we have assumed for both the considered scenarios:

- a multicast set composed by $M = 30$ nodes;

- the length L of each information packet equal to 512 and 1024 bits;

- the normalized round-trip time, s equal to 5 or 10.

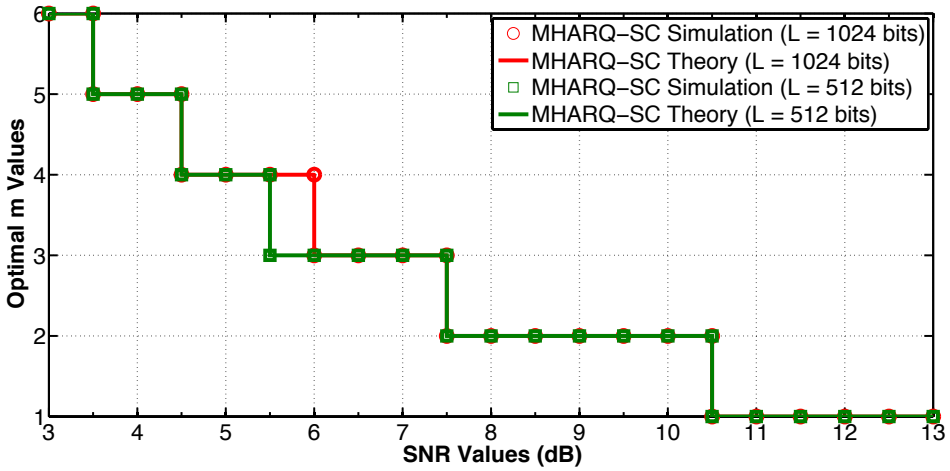


Figure 2.4: Optimal values of m vs. γ in the case of AWGN propagation conditions (scenario 1).

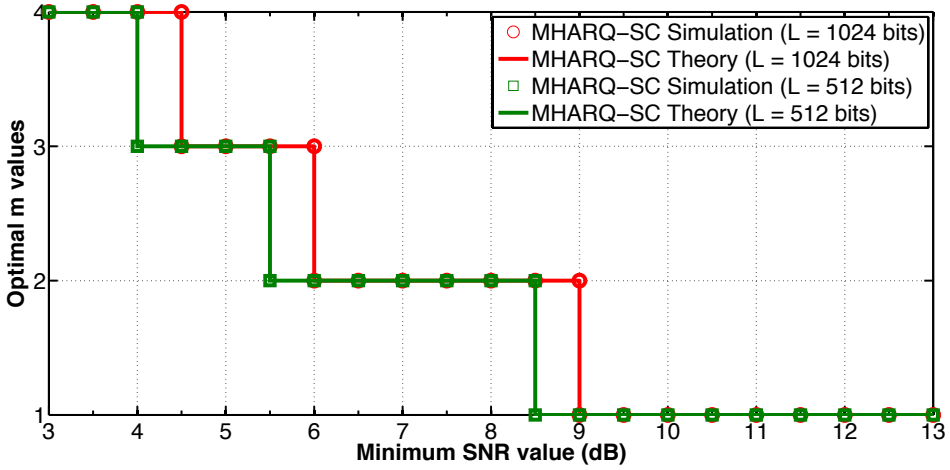


Figure 2.5: Optimal values of m vs. the minimum values of γ_i in the case of AWGN propagation conditions (scenario 2).

We start our analysis by providing numerical results concerning the optimization procedure of the proposed MHARQ-SC scheme (outlined in Section 2.1.1). The obtained results are shown in Figure 2.4 and 2.5 as a function of γ and $\bar{\gamma}$, respectively, under the assumption of AWGN propagation conditions. These figures report the optimal values of m derived by solving the problem P1. The aforementioned values have been compared to those obtained by numerical results. It is straightforward to note that the analytical predications match the simulation results[1].

Figures 2.6 and 2.7 show the performance in terms of the mean multicast delivery delay in the case of the optimized MHARQ-SC scheme (normalized with respect to

[1]For what concerns the latter aspect, the value of m has been optimized by simulating the transmission of the same information packet and testing different values of $m \in [1, 30]$. The value of m which minimizes (2.13) has been considered as the optimum one.

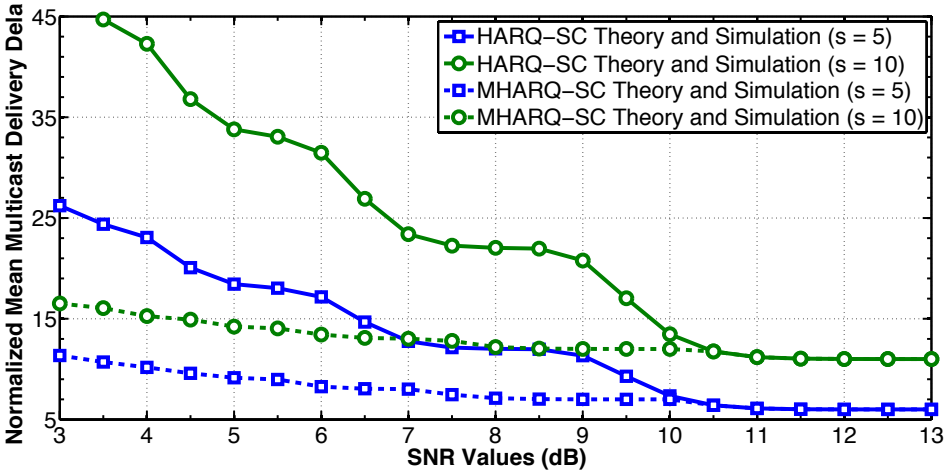


Figure 2.6: Normalized mean multicast delivery delay as a function of γ in the case of AWGN propagation conditions (scenario 1).

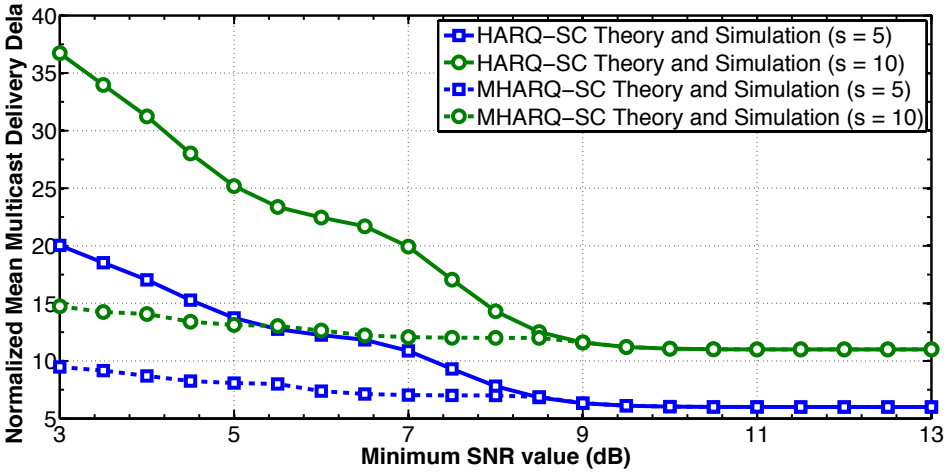


Figure 2.7: Normalized mean multicast delivery delay as a function of the minimum value of γ_i in the case of AWGN propagation conditions (scenario 2).

τ). The figures also compare the MHARQ-SC with the classical HARQ-SC scheme by assuming $L = 1024$ bits and $M = 64$ nodes. These figures clearly show that the proposed MHARQ-SC outperforms the classic HARQ-SC.

Let us consider frequency-non selective slow Rayleigh faded propagation conditions. According to this, Figures 2.4-2.5 and Figures 2.8-2.9 show the optimal values of m. The figures also compare both analytical and simulation results. Also in this case we can note that the analytical predictions match the simulation results.

Figures 2.10-2.11 show the (normalized) mean multicast delivery delay of the optimized MHARQ-SC scheme and classical HARQ-SC (for $L = 1024$ and M=64). These figures clearly show that the optimized MHARQ-SC scheme outperform the other alternative.

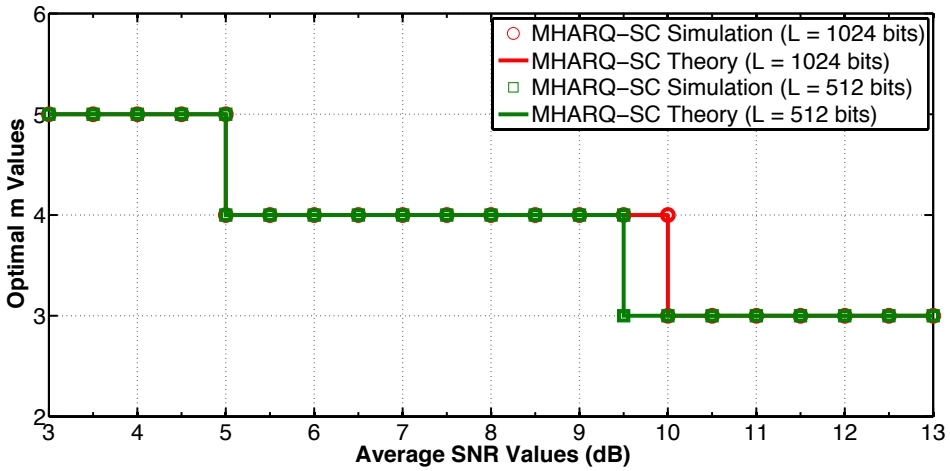


Figure 2.8: Optimal m values as a function of $\bar{\gamma}$ in the case of faded propagation conditions (scenario 1).

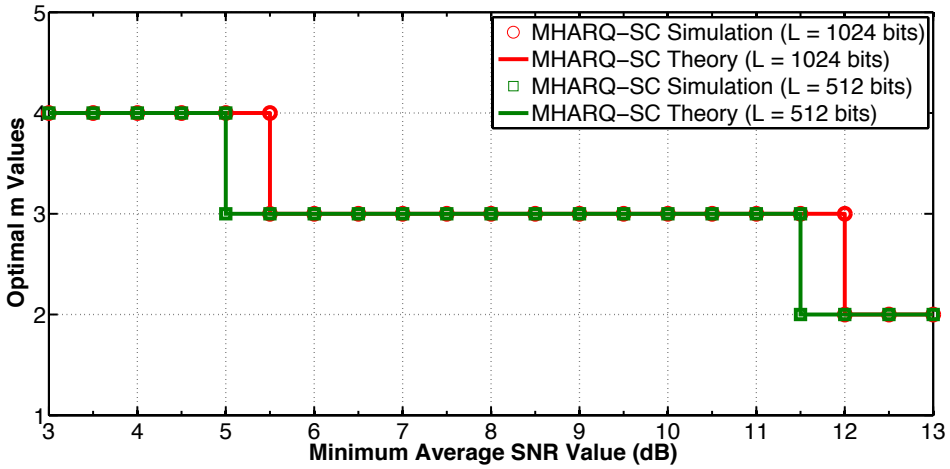


Figure 2.9: Optimal m values as a function of the minimum value of $\bar{\gamma}_i$ in the case of faded propagation conditions (scenario 2).

In addition, Figure 2.12 compares the performance of the optimized MHARQ-SC scheme with that achieved by the rate optimized HARQ-SC scheme proposed by Kim *et al.* [22] (indicated in the figure as Opt. HARQ-SC). The figure considers the scenario 1 with $\bar{\gamma}_i = 3$ dB (for $i = 1, \ldots, M$) and (frequency-non selective slow Rayleigh) faded propagation conditions. Also in this case the proposed MHARQ-SC clearly outperforms the other alternatives.

Briefly, in this section we investigated the performance of the proposed MHARQ-SC scheme in a multicast wireless communication system. The performance of the proposed solution has been validated under AWGN and frequency-non-selective slow Rayleigh faded propagation conditions. The MHARQ-SC approach has been compared to multiple alternatives, such as: the classical HARQ-SC and optimized HARQ-SC

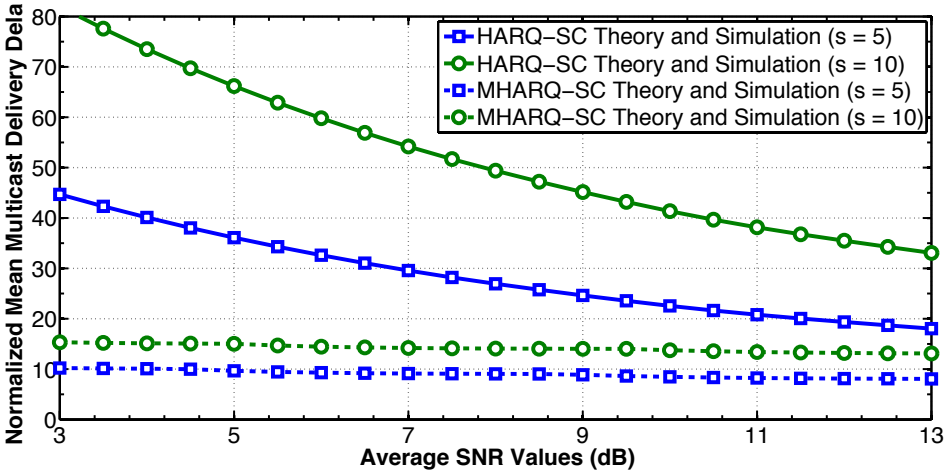


Figure 2.10: Normalized mean multicast delivery delay for information packet as a function of $\bar{\gamma}$ for faded propagation conditions (scenario 1).

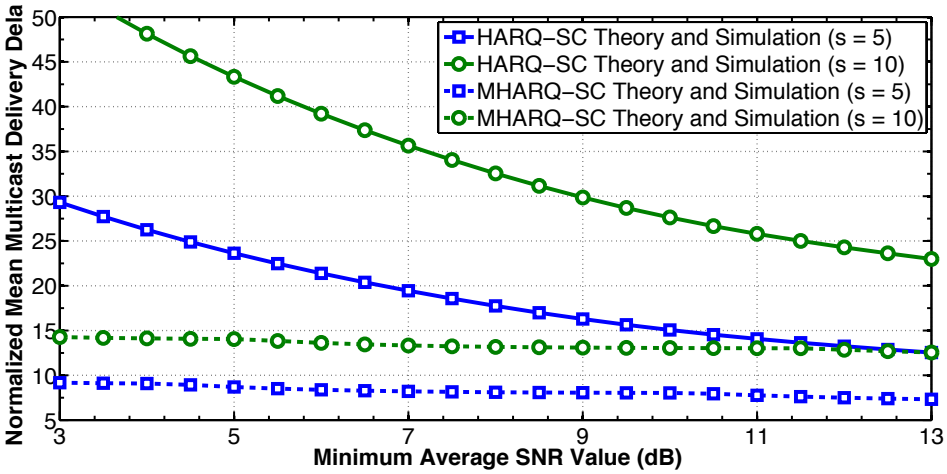


Figure 2.11: Normalized mean multicast delivery delay for information packet as a function of the minimum $\bar{\gamma}_i$ for faded propagation conditions (scenario 2).

schemes. The effectiveness of the proposed optimization and analytical model has been clearly shown.

2.2 NC Error Control Strategies with Symbol Combining

NC principle [28] is receiving a great attention as an effective way of improving the capacity of both wired and wireless networks, also including sensor [29] and vehicular [30] networks. In particular, Wang [31] proved that the NC achieves the min-cut flow in broadcast scenarios, and hence, makes possible improving the network capacity [32]. This section deals with the RLNC [33] approach which represents the simplest and the

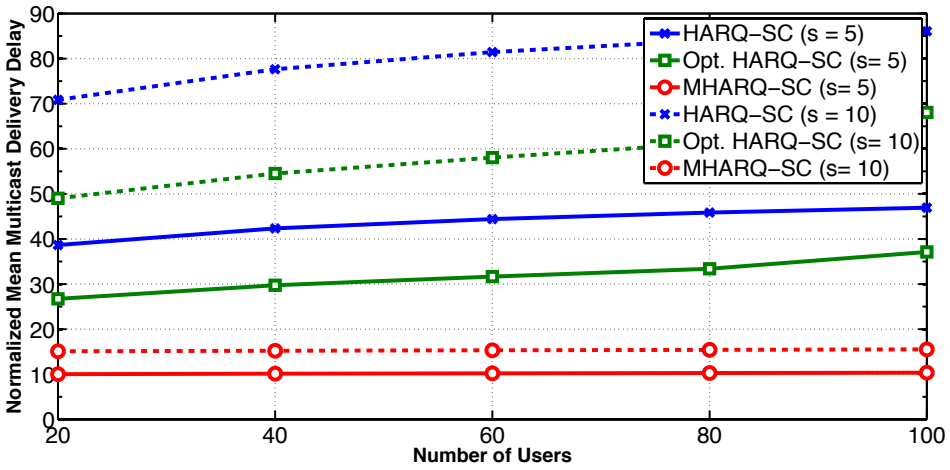


Figure 2.12: Normalized mean multicast delivery delay as a function of M for $\bar{\gamma}_i = 3$ dB.

most efficient way to implement the NC communication principle in a communication network.

As for a wireless communication network, one of the most important issue is how to counteract packet erasures. Several approaches have been proposed in the literature, including the integration of the NC with ARQ [33] or even HARQ schemes [34, 35, 36]. Among novel proposals, Chiti *et al.* [37] outline a power adaptation which aims at increasing the communication reliability. In particular, this approach can be considered as an ideal realisation of the well known Chase combining principle [38] widely adopted in the case of ARQ systems. Unlike the other proposals, this section deals with the performance evaluation and optimization of a novel RLNC scheme which is suitable for burst communications over lossy links. In the rest of the section, we will refer to the proposed approach as Symbol Duration Increased NC (SDI-NC). In brief, the SDI-NC scheme leads one transmitting end to deliver packets where the symbol duration is increased of a fixed (and integer) factor as stated by the SC principle [38]. In the rest of the section, we will show how to optimize and implement the proposed SDI-NC scheme in a broadcast and butterfly network model [33].

The remaining part of the section is organized as follows. Section 2.2.1 provides a quick overview about the RLNC and related works. Section 2.2.2 describes the SDI-NC principle on a link-to-link basis by considering an AWGN and a slow fading regime; two optimization methods are proposed in this Section. Section 2.2.2.1 generalises the proposed scheme to the case of a butterfly network model. At the end, Section 2.2.3 presents an extensive performance comparison among the classic RLNC and the proposed SDI-NC schemes.

2.2.1 Background and Related Works

2.2.1.1 RLNC Communication Strategy

Let $\mathbf{E} = \{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_K}\}$ be a message composed of K packets (in the rest of the section we will refer to it as "generation length"), each formed by J elements (belonging to a finite field and L/J bits long), where L is the packet length (in bits). Note that $\mathbf{E}$ can be alternatively modelled as a $J \times K$ matrix ($\mathbf{M_E}$) where the i-th column is defined by the i-th information packet, (for $i = 1, \ldots, K$). In this section we will refer to the RLNC, where coded packets are generated by a linear combination of the original ones. The j-th coded packet $\hat{\mathbf{e}}_\mathbf{j}$ can be computed as follows:

$$\hat{\mathbf{e}}_\mathbf{j} = \mathbf{M_E} \cdot \mathbf{c_j}, \tag{2.17}$$

where $\mathbf{c_j} = [c_{1,j}, c_{2,j} \ldots, c_{K,j}]^T$ is an N-dimensional column vector (called "coding vector") whose elements are randomly chosen. Finally, $\mathbf{c_j}$ and $\mathbf{e_j}$ (for $j = 1, 2, \ldots, K$) belongs to a large enough finite field $\mathcal{G}_q$ of size q [39]. According to this, a transmitting node can compute $N = K + G$ coded packets where G represents the introduced redundancy. The term G is directly related to three parameters [40, 41]: q, K and the packet delivery probability. From the main theorem of NC [39] and the basic properties of the RLNC [42], we have that a receiving node needs to collect at least K linearly independent coded packets to successfully recover the message $\mathbf{E}$.

Since coding vectors are randomly chosen, the probability that two coded packets are linearly dependent is nonzero. Hence, N coded packets are sent out in order to make possible recovering a generation [43]. In order to recover the transmitted message, each receiver has to collect the coding vector associated to each given coded packet (which are delivered together). Hence, each coded packet has to deliver an extra amount of bits which is equal to of smaller than $K \log_2(q)$ bits [44]. However, due to the fact that coding vectors are quite sparse, several approaches have been proposed [44, 45] to reduce such kind of overhead. However, this particular aspect is out of the scope of this section. Hence, we assumed that the coding vectors are known at the receiving ends. At the receiver side each linearly independent coded packet and its corresponding coding vector, defines a column of the matrices $\mathbf{M_{\hat{E}}}$ and $\mathbf{C}$ (of dimensions $J \times K$ and $K \times K$), respectively [39]. When the number of linearly independent coded packets is (at least) equal to K, the original message can be computed as follows:

$$\mathbf{M_E} = \mathbf{M_{\hat{E}}} \cdot \mathbf{C}^{-1}. \tag{2.18}$$

2.2.1.2 Related Works

The multimedia broadcasting (or multicasting) over 4G networks will be in the near future a key commercial service [46]. Usually, users belonging to the same broadcast group are spread over wide areas and suffer of different propagation conditions. Efficient strategies, which aim at preserving data integrity without loosing so much communication throughput, are more than welcome.

To this end, Kim *et al.* [22] propose an optimized version of a well known HARQ strategy which relies on the SC principle. However, in that case the strategy has to face with a severe limitation: each packet has to be acknowledged to the transmitting node. The aforementioned drawback can be overcome by resorting to the RLNC as proposed by Eryilmaz *et al.* [47]. In particular, authors of [47] describe a strategy to counteract channel erasures in PtM communication by resorting to an optimized scheduling scheme. In addition, authors derive theoretical bounds for communications that occur over an ON/OFF channel.

After that, the theoretical derivation of Ghaderi *et al.* [32] remarks that the RLNC can be efficiently used as an error control strategy in a broadcast and multicast context. Finally, [48] and [49] propose an optimized NC scheme which aims at minimizing the packet losses. Unfortunately, the aforementioned NC-based communication schemes takes place only in these cases where packets are received with errors and have to be re-transmitted. Hence, that reasoning cannot fully exploit the main features of the NC benefits highlighted in [32].

2.2.2 Symbol Combining NC Principle

This section investigates the performance gain of the proposed integration of the SC principle to RLNC. The original formulation of the SC principle foresees a bit-by-bit combination of all the received copies of the same data packet (also including those received with errors). In particular, it aims at implementing a soft-detection at each receiving ends [21]. It is straightforward to note that if RLNC scheme is used as an error control strategy, each coded packet is potentially different to the others. Hence, the basic SC approach [38] cannot be directly applied. That drawback can be overcome by resorting to the alternative formulation of the SC principle proposed in [21]. In this case we propose to transmit each coded packet such that the duration of each symbol is increased by a factor m (i.e., the SDI factor). It is worth noting that the proposed approach is characterized by a an implementation complexity which is not greater than that characterises the basic SC approach [38].

In the rest of this section we compare the proposed SDI-NC approach with the classic

RLNC. However, the performance of the SDI-NC scheme depends on the parameter m, hence in the rest of the section, we propose an efficient optimization procedure for it. In particular, the performance analysis focuses on an objective function which expresses both the mean delivery delay (i.e., the mean time needed to recover N coded packets by all the receiving ends), and mean energy consumption.

Let us start our analysis by focusing on the broadcast network model, shown in Figure 2.1, where a source (the AP) broadcasts packets to a set of $\{W_i\}$ nodes (where $i = 1, \ldots, M$) over M lossy independent channels. Whenever W_i successfully collects N coded packets (i.e., whenever it collects K linearly independent coded packets), it transmits to the AP an ACK[1]. The AP starts the transmission of a new generation whenever it collects an ACK message from all the network nodes. Moreover, we assume that the burst of N coded packets is sent through an AWGN channel by means of a Quadrature Phase-Shift Keying (QPSK) modulation[2] [50].

Let L be the size of a coded packet (expressed in bits), under the assumption of an ideal error detecting code. The Packet Error Probability (PEP) characterizing the reception of W_i can be defined as follows:

$$P_{B,i}(m) = 1 - \Big[1 - P_{e,i}(m)\Big]^{L}, \tag{2.19}$$

where $P_{e,i}(m)$ represents the bit error probability which affects the reception of W_i. In particular, $P_{e,i}(m)$ can be expressed as follows [50]:

$$P_{e,i}(m) = \mathbf{Q}\Big(\sqrt{m\,\gamma_i}\Big) \quad \text{i=1, } \ldots \text{,M,} \tag{2.20}$$

where

- γ_i is the ratio between the energy associated to each transmitted symbol and the one side AWGN spectral density (at the i-th receiver side); here after named as SNR for the i-th link;
- $\mathbf{Q}(x)$ is the well known Q function.

It is worth noting that, (2.20) has been derived by assuming that the effects of the channel interferences are negligible. Finally, if the overall contribution of the channel interferences can be modelled like an independent Gaussian noise, (2.20) holds[3].

[1] Also in this case, without loss of generality we assumed that the transmission of ACKs occurs instantaneously and through a fully reliable communication channel.

[2] Note that the derived results are quite general and they can be easily extended to different modulation schemes and communication channel models.

[3] For the sake of clarity, in that case γ_i represents the SINR of the i-th receiving end.

Let us consider the random variable V_i which represents the number of coded packets that the AP has to broadcast, to ensure the correct reception of N packets by W_i. Ghaderi *et al.* [32] show that V_i follows a negative binomial distribution. In particular, for $x \geq N$, the probability mass function $f_{V_i}(x; N)$ of V_i (for $i = 1, 2, \ldots, M$) can be defined as:

$$f_{V_i}(x; N) = \binom{x-1}{N-1} \Big[1 - P_{B,i}(m)\Big]^K P_{B,i}^{x-N}(m). \tag{2.21}$$

$f_{V_i}(x; N)$ is equal to zero if $x < N$. For this reason, the number of coded packets that the AP has to deliver (such that all the receiving ends collect N coded packets) can be defined as: $W = \max_{i=1,\ldots,M} \Big\{V_i\Big\}$. The probability mass function of W results to be [51]:

$$\begin{aligned} f_W(x; N) &= \text{Prob}\Big\{W \leq x\Big\} - \text{Prob}\Big\{W \leq x-1\Big\} \\ &= \prod_{r=1}^{M} \left[\sum_{i=N}^{x} f_{V_r}(i; N)\right] - \prod_{r=1}^{M} \left[\sum_{i=N}^{x-1} f_{V_r}(i; N)\right], \end{aligned} \tag{2.22}$$

where $\text{Prob}\Big\{W \leq x\Big\}$ is the probability that the value of W is equal to or smaller than x (for $x = 1, 2, \ldots, \infty$). According to (2.22), the average value of W can be defined as follows:

$$\Upsilon(m; N) = \sum_{i=N}^{\infty} i\, f_W(i; N). \tag{2.23}$$

The goal of the proposed optimization is to find the value of m (namely, m_o) which minimises the mean delivery delay and energy consumption (needed to successfully recover a generation). Hence, m_o can be derived by solving the optimization problem oSDI-NC[1]:

$$\begin{aligned} \text{(oSDI-NC)} \qquad &\text{minimize} \quad \Upsilon(m; N) && (2.24) \\ &\text{subject to} \quad m \in \mathbb{N}. && (2.25) \end{aligned}$$

Let us consider again (2.22) and (2.23), we note that the oSDI-NC problem is nonlinear. Moreover, because of the constraint (2.25), it also is an integer optimization problem. For these reasons, it is hard to solve the oSDI-NC problem in a closed form, therefore, we resorted to a suitable numerical approach[2].

In order to reduce the computational load of the optimization process, in the following paragraphs we will propose an efficient suboptimal optimization method whose

[1] In the rest of this section with the symbol $\mathbb{N}$ we will refer to the set of non-null integer numbers.
[2] In particular, also in this case we resorted to the NOMAD solver [27].

accuracy will be validated in Section 2.2.3. Let us define the mean Link-to-Link (L2L) delivery delay for W_i (i.e., the mean time needed by the i-th node to collect K coded packets) as

$$\hat{\delta}_i(m; N) = \frac{m\, N\, L\, T_b}{1 - P_{B,i}(m)}, \tag{2.26}$$

where T_b is the transmission time duration of a bit for $m = 1$. From (2.26), we have that the mean time needed to correctly receive a coded packet (normalized to the nominal coded packet duration $L\, T_b$) is

$$\delta_i(m) = \frac{m}{1 - P_{B,i}(m)}. \tag{2.27}$$

From (2.27), it is straightforward to note that the mean communication throughput (normalized to $L\, T_b$) can be expressed as $1/\delta_i(m)$. In addition, let us define the L2L mean energy $\epsilon_i(m)$ needed to successfully deliver a coded packet, normalized with respect to $L\, E_b$ (i.e., the energy associated to the transmission of a coded packet composed by L bits). In particular, we have

$$\epsilon_i(m) = \frac{m}{1 - P_{B,i}(m)}. \tag{2.28}$$

It is worth noting that, the normalized L2L mean delay and energy consumption, associated to the reception of a coded packet, in the case of the classical RLNC scheme can be found by setting $m = 1$ in (2.27) and (2.28), respectively. Moreover, from (2.27) and (2.28), we note that $\delta_i(m) = \epsilon_i(m)$ (for $i = 1, \ldots, M$). Hence, we simplify our analysis by defining as objective function, the function $\Gamma_i(m) : \mathbb{N} \longrightarrow \mathbb{R}^+$ given by[1]:

$$\Gamma_i(m) \doteq \delta_i(m) \;=\; \epsilon_i(m) = \frac{m}{1 - P_{B,i}(m)} \;= \frac{m}{[1 - P_{e,i}(m)]^L}. \tag{2.29}$$

In particular, form (2.29), we note that minimizing $\Gamma_i(m)$ is equivalent to minimize, at the same time, the mean L2L delay and energy consumption. It is worth noting that, from (2.20) and (2.29), $\Gamma_i(m)$ is monotonically decreasing (for $i, j = 1, \ldots, M$ and $i \neq j$):

$$\Gamma_i(m) \geq \Gamma_j(m) \quad \text{iff} \quad \gamma_i \leq \gamma_j. \tag{2.30}$$

Therefore, for the sake of simplicity, in the rest of this section we consider the minimization of the function $\Gamma_i(\cdot)$.

We propose a min-max-based optimization strategy as heuristic optimization crite-

[1]In the rest of the section with $\mathbb{R}^+$ symbol we will refer to the set of not negative real numbers.

rion. Let us consider the problem below

$$\text{(sSDI-NC)} \qquad \min \max_{i=1,2,\dots,M} \quad \Gamma_i(m) \tag{2.31}$$

$$\text{subject to} \quad m \in \mathbb{N}. \tag{2.32}$$

The accuracy of the sSDI-NC model is validated in Section 2.2.3 by comparing it to the oSDI-NC approach. In addition, we prove that the sSDI-NC is convex.

As for (2.30), the sSDI-NC problem is equivalent to the following one:

$$\text{(esSDI-NC)} \qquad \text{minimize} \quad \Gamma_h(m), \tag{2.33}$$

$$\text{where } h := \arg\min\{\gamma_i | i = 1, \dots, M\}$$

$$\text{subject to} \quad m \in \mathbb{N}, \tag{2.34}$$

where W_h is the node experiencing the worst propagation conditions. Let $\hat{\Gamma}_i(\hat{m}) : \mathbb{R}^+/\{0\} \longrightarrow \mathbb{R}^+$ be the continuous extension of $\Gamma_i(m)$ (for $i = 1, \dots, M$). The following proposition holds:

Proposition 2.1. The function $\hat{\Gamma}_i(\hat{m})$, for $i = 1, \dots, M$ is convex, continuously differentiable and admits an unique minimum in $\mathbb{R}^+/\{0\}$.

Proof. See Applendix C.1.1. □

Let us solve the esSDI-NC problem by relaxing the constraint (2.34). With other words, we can restate it as

$$\text{(reSDI-NC)} \qquad \text{minimize} \quad \hat{\Gamma}_h(m), \tag{2.35}$$

$$\text{where } h := \arg\min\{\gamma_i | i = 1, \dots, M\}$$

$$\text{subject to} \quad m \in \mathbb{R}^+/\{0\}. \tag{2.36}$$

From Proposition 2.1, the solution of the (reSDI-NC) problem is the root ($\hat{r}$) of the equation:

$$\frac{d}{d\hat{m}}\Big(\hat{\Gamma}_h(\hat{m})\Big) = 0 \Leftrightarrow 1 - \mathbf{Q}\Big(\sqrt{\hat{m}\gamma_h}\Big) - L\sqrt{\frac{\gamma_h \hat{m}}{2\pi}} e^{-\frac{\hat{m}\gamma_h}{2}} = 0. \tag{2.37}$$

Finally from Proposition 2.1, if $\hat{r}$ exists, $\tilde{m}$ is equal to $\min\big\{\Gamma_h\big(\lfloor \hat{r} \rfloor\big), \Gamma_h\big(\lceil \hat{r} \rceil\big)\big\}$.

In order to validate the proposed optimization strategies, in the rest of this section we address the case of slow faded propagation conditions. For the sake of the analysis, we assume that the propagation conditions are constant for all nodes during the trans-

mission of a coded packet (regardless of m)[1] and statistically independent duting each coded packet transmission. Hence, $P_{B,i}(m)$ can be defined as [50]:

$$P_{B,i}(m) = 1 - \frac{1}{\overline{\gamma}_i} \int_0^{\infty} \Big[1 - P_{e,i}(m, \gamma_i)\Big]^L e^{-\frac{1}{\overline{\gamma}_i}\gamma_i} d\gamma_i, \tag{2.38}$$

where $\gamma_i = |\alpha_i|^2 \frac{2\,E_b}{N_{0,i}}$. The parameter $|\alpha_i|$ is the channel attenuation (Rayleigh distributed), $N_{0,i}$ is the one side AWGN spectral density (at the i-th receing side), and $\overline{\gamma}_i$ is the mean SNR per symbol experienced by W_i.

Also in this case the oSDI-NC optimization problem cannot be efficiently solved in real time. However, the value of m can be optimized by resorting to the sSDI-NC model. In addition, Proposition 2.1 still holds (see Appendix C.1.2). Hence, we have that:

- the (sSDI-NC) optimization problem can be equivalently rewritten as reported in (esSDI-NC);
- the solution ($\hat{r}$) of the (rsSDI-NC) problem can be derived by solving the following equation

$$\frac{d}{d\,\hat{m}}\Big(\hat{\Lambda}(\hat{m})\Big) = 0 \Leftrightarrow \int_0^{\infty} \Big[1 - Q(\sqrt{\hat{m}\,\gamma_i})\Big]^L e^{-\frac{1}{\overline{\gamma}_i}\gamma_i} d\gamma_i \\ -\frac{L\sqrt{\hat{m}}}{2\sqrt{2\pi}} \int_0^{\infty} \sqrt{\gamma_i} \Big[1 - Q(\sqrt{\hat{m}\,\gamma_i})\Big]^{L-1} e^{-\frac{2+\hat{m}\overline{\gamma}_i}{2\overline{\gamma}_i}\gamma_i} d\gamma_i = 0; \tag{2.39}$$

- from Proposition 2.1, the solution ($\tilde{m}$) of sSDI-NC can be selected between $\lfloor \hat{r} \rfloor$ and $\lceil \hat{r} \rceil$, by choosing the value that minimizes the objective function (2.31) .

2.2.2.1 Application to the Butterfly Network Topology

This section proposes a generalisation to the results provided in Section 2.2.2 to the case of butterfly network models (shown in Figure 2.13). Even though the considered network model is purely theoretical, it is useful to inspect the performance of the SDI-NC approach in a multi-hop network.

We considered two sources, A and B, which transmit independent information messages. Each message is N packets long (message **a** for the node A and message **b** for the node B). They are directed to three different destinations: C, D and R. Moreover, the node R acts as a relay for nodes C and D. The node A transmits coded packets

[1] This occurs, for example, whenever data transmissions are organized on a frame-basis (as in LTE [4] or WiMAX [52] systems).

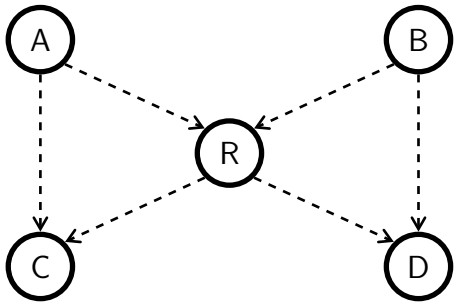


Figure 2.13: Butterfly network model.

obtained from the message $\mathbf{a}$ to nodes R and C. The node B does the same for nodes R and D. For the sake of simplicity, we assume that:

- all the transmitting nodes access the medium in a contention free fashion;
- coding/decoding operations are performed by consenting the same finite field $\mathcal{G}_q$. Hence, the same number (N) of coded packets (L bits long) are needed (for each information flow) by all the receiving ends to recovered transmitted generations.

Whenever the relay node R has recovered $\mathbf{a}$ and $\mathbf{b}$, it can start the transmission of coded packets $\hat{\mathbf{r}}_\mathbf{i}$ (L bits long) obtained as follows:

$$\hat{\mathbf{r}}_i = \mathbf{M}_\mathbf{r} \cdot \mathbf{c}_\mathbf{i}, \tag{2.40}$$

where $\mathbf{c}_\mathbf{i}$ is the i-th K-ary coding vector. $\mathbf{M}_\mathbf{r}$ is a $P \times K$ matrix where the q-th column is defined as $\mathbf{a}_q \otimes \mathbf{b}_q$ (i.e., the q-th coded packets of $\mathbf{a}$ and $\mathbf{b}$ are XORed bit-by-bit). Moreover, $\mathbf{M}_\mathbf{r}$ defines the original message $\mathbf{r} = \{\mathbf{r_1}, \ldots, \mathbf{r_K}\}$ transmitted by the node R. It is straightforward to note that in a butterfly network, each destination node (C or D) receives two coded packets, one from a source node (A or B) and the other one from the relay node R. The decoding process operated by node C (or D) can be summarized as follows:

1. to recover message $\mathbf{a}$ ($\mathbf{b}$) by decoding the packets received from node A (B), see (2.18);
2. to recover message $\mathbf{b}$ ($\mathbf{a}$) by decoding the packet received from node R. In particular the q-th plain packet of the message $\mathbf{b}$ ($\mathbf{a}$) is given by $\mathbf{r}_q \otimes \mathbf{a}_q$ ($\mathbf{r}_q \otimes \mathbf{b}_q$).

In order to implement the optimization criterion presented in Section 2.2.2 the butterfly network has to be split into three broadcast subnetworks:

- $\overbrace{ACR}$, subnetwork A, where A is the AP for C and R;

- $\widetilde{BRD}$, subnetwork B, in this case R is AP for R and D;
- $\widetilde{RCD}$, subnetwork R, in this case R is the AP for C and D.

Finally, we can note that R is both a receiver (for A and B) and an AP (for the $\widetilde{RCD}$ broadcast network). We assume that each node can receive coded packets coming from just a transmitting node at a time. In particular, a receiving end starts the reception of a new message only if it has successfully decoded (and acknowledged) the previous one.

It is worth noting that, also in this network context, both the proposed SDI-NC scheme and the optimization strategies are still valid. In particular, the optimal value of m, namely, m_X (where $X \in \{A, B, R\}$), can be derived for each broadcast subnetwork.

In order to compare the performance of the SDI-NC scheme with that of the classical RLNC, let us consider the mean End-to-End (E2E) delivery delay (normalized to KLT_b). Let $\widetilde{\delta}$ be the mean time required by C and D to successfully recover an information packet belonging to **a** and **b**. The mean normalized E2E delivery delay results to be

$$\widetilde{\delta} = \overline{\delta}(m_A) + \overline{\delta}(m_B) + \overline{\delta}(m_R), \tag{2.41}$$

where m_A, m_B and m_R are the optimized SDI values for subnetworks A, B and R, respectively, derived according to the oSDI-NC or sSDI-NC optimization criteria. The parameter $\overline{\delta}(m_X)$ is the overall mean delay (normalized to $L\,T_b$) needed by all the receiving nodes in the subnetwork X, (with $X \in \{A, B, R\}$), to successfully recover a coded packet. Finally, (2.41), with $m_X = 1$ (for $X \in \{A, B, R\}$), defines the mean normalized E2E delivery delay (for a coded packet) in the case of the classical RLNC scheme.

2.2.3 Preliminary Results

This section deals with the performance evaluation of the optimized SDI-NC scheme in the case of broadcast and butterfly network models. In particular, we focus on AWGN and slow faded propagation conditions. In addition, the section compares the oSDI-NC model to the sSDI-NC one. Results are provided in terms of normalized mean delivery delay.

2.2.3.1 Broadcast Network Scenario

In order to compare the oSDI-NC and sSDI-NC optimization models, the performance of the optimized SDI-NC scheme has been inspected by resorting to computer simulations. In particular we considered the following network scenarios:

Table 2.1: Probability of nonsingularity for $N = K$ in the case of $N = 10$.

q	2^4	2^5	2^6	2^7	2^8
p_{ns}	0.934	0.968	0.984	0.992	0.996

I The number of network nodes M is equal to 30. The maximum SNR (AWGN case) and mean SNR (slow fading case) offsets among all the nodes is 10 dB. Without loss of generality, we assume that W_M experiences the best propagation conditions while node W_1 the worst ones[1]. The SNR (AWGN case) and mean SNR values (slow fading case) associated to the remaining nodes span the interval [5, 15] dB or [0, 10] dB, respectively. The SNR offset between any couple of nodes is constant.

II The number of receiving nodes is $M \in [2, 32]$. All the receiving ends experience the same propagation conditions (namely, $\gamma_i = 9$ dB and $\overline{\gamma}_i = 5$ dB in an AWGN and slow fading regime, respectively).

Moreover, regardless of the scenario in use, we assume that all the coding operations are performed within a (large enough) finite field so that two coded packets can be considered linearly independent with high probability [53]. In particular, Table 2.1 shows the probability that a node has received $K = 10$ over $N = K$ linearly independent coded packets (without considering the channel effects) [40]. In particular, if coding vectors are uniformly selected over a finite field with a dimension greater than or equal to 2^8, the probability that two coding vectors are linearly dependent is less than $4 \cdot 10^{-3}$. Hence, we considered that: i) $K \cong N$ and $K = 10$.

Considering the scenario I and AWGN propagation conditions. Figure 2.14 shows the normalized mean delivery delay (for two different packet lengths) as a function of the mean SNR value among users ($\overline{\gamma}$) defined as:

$$\overline{\gamma} = \frac{1}{M} \sum_{i=1}^{M} \gamma_i. \tag{2.42}$$

In particular, the SDI values have been derived by the oSDI-NC and sSDI-NC criteria. The figure clearly shows that, regardless of the optimization strategy in use, the optimum and heuristic SDI values are almost the same (there is a slightly difference for $\overline{\gamma} \in [12.2, 12.4]$ dB).

Figure 2.15 compares the performance of the SDI-NC scheme (optimized according to the oSDI-NC and sSDI-NC criteria) to that of the classical RLNC scheme. This figure highlights the same behaviour for both the optimization criteria. Moreover, it

[1]It corresponds to W_h in esSDI-NC.

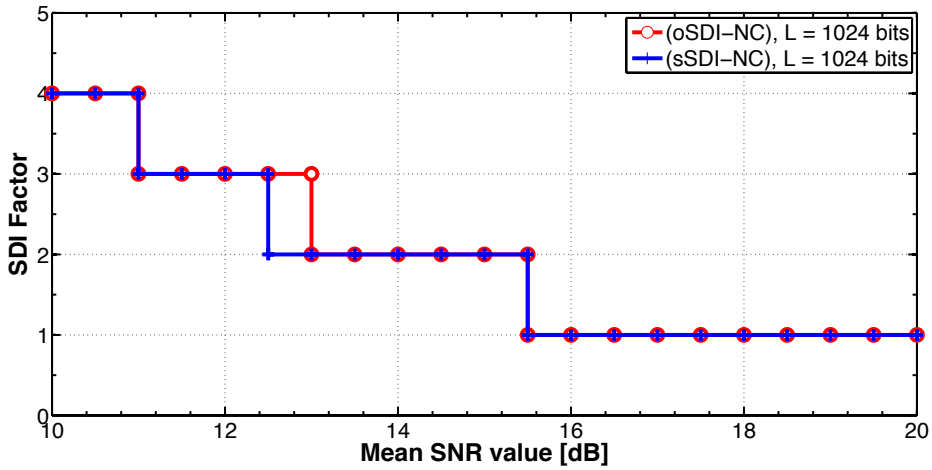


Figure 2.14: Optimal SDI factors vs. the mean SNR among users (AWGN propagation conditions and scenario I).

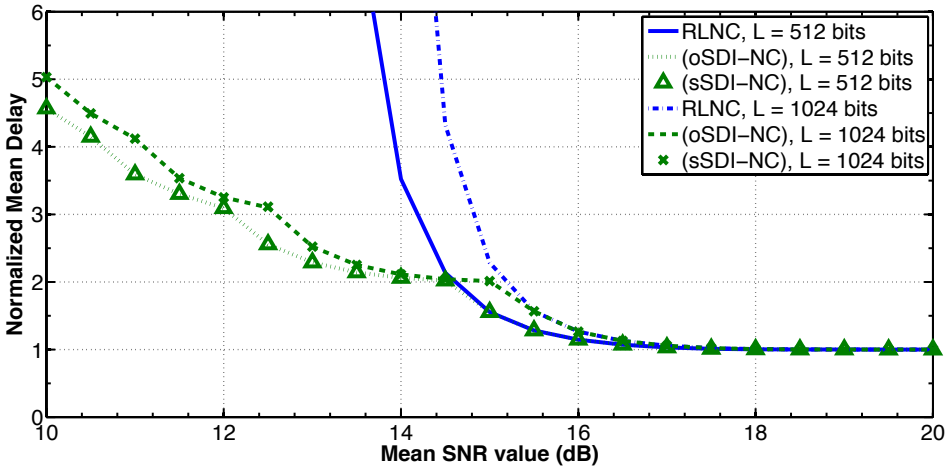


Figure 2.15: Normalized mean delivery delay vs. the mean SNR among users (AWGN propagation conditions and scenario I).

is worth noting that the optimized SDI-NC clearly outperforms the classical RLNC scheme for mean SNR values which are less than 15 dB. In addition, Figure 2.15 shows that the performance of the oSDI-NC and sSDI-NC models are exactly the same for $L = 512$ bits. On the other hand, for $L = 1024$ bits, the sSDI-NC strategy is characterized by a normalized mean delivery delay that is (at most) 0.4 % greater than that characterizing the oSDI-NC approach.

Likewise, Figure 2.16 compares the normalized mean delivery delay as a function of the number of receiving nodes in an AWGN regime, in the case of scenario II. Also in this case, we can note that the optimization model oSDI-NC performs as well as the sSDI-NC one. Finally, Figure 2.17-2.19 reports the performance of the proposed scheme

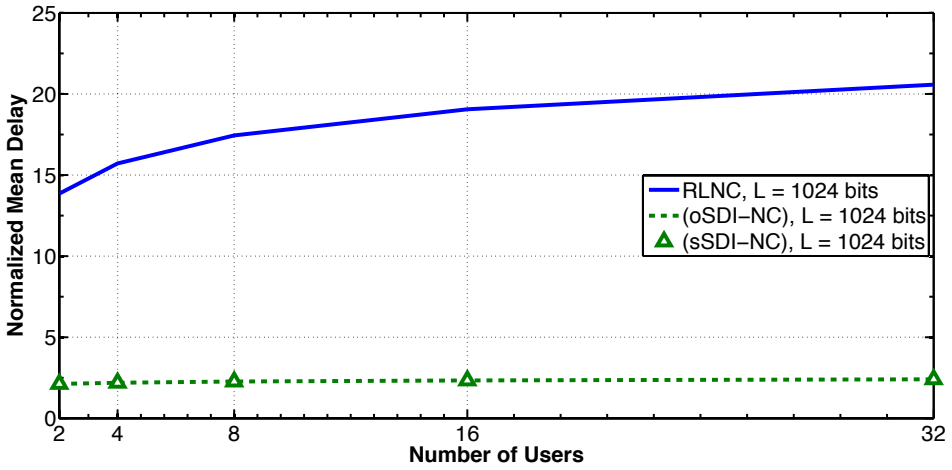


Figure 2.16: Normalized mean delivery delay vs. the number of receiving nodes for AWGN propagation conditions (scenario II).

in a slow fading regime. Numerical results are reported as a function of the parameter

$$\tilde{\gamma} = \frac{1}{M} \sum_{i=1}^{M} \overline{\gamma}_i, \tag{2.43}$$

and as a function of the number of receiving nodes. It is worth noting that: (i) the optimized SDI-NC scheme outperforms the classical RLNC, and (ii) the oSDI-NC performance is very close to the sSDI-NC one. In particular, Figure 2.18 compares the performance of sSDI-NC and oSDI-NC. Both the proposed optimization strategies are almost characterised by the same normalized mean delivery delay. The only exception is represented by the case of $L = 1024$ bits where the sSDI-NC approach is characterized by a normalized mean delivery delay that is (almost) 0.34 % greater than that of the oSDI-NC model.

2.2.3.2 Butterfly Network Scenario

In this case we consider a butterfly network scenario (shown in Figure 2.13) where:

- γ_i values of the end points of A-C and A-R links are equal to 9 dB in AWGN regime, and $\overline{\gamma}_i$ values are equal to 5 dB in the case of a slow fading regime;
- γ_i values of the end points of B-D and B-R links are equal to 10 dB, and $\overline{\gamma}_i$ values are equal to 6 dB in the case of a slow fading regime;
- SNR values at the end points of R-C and R-D links are equal and take values within $[5, 15]$ dB (likewise, for the case of a slow fading regime, the mean SNR

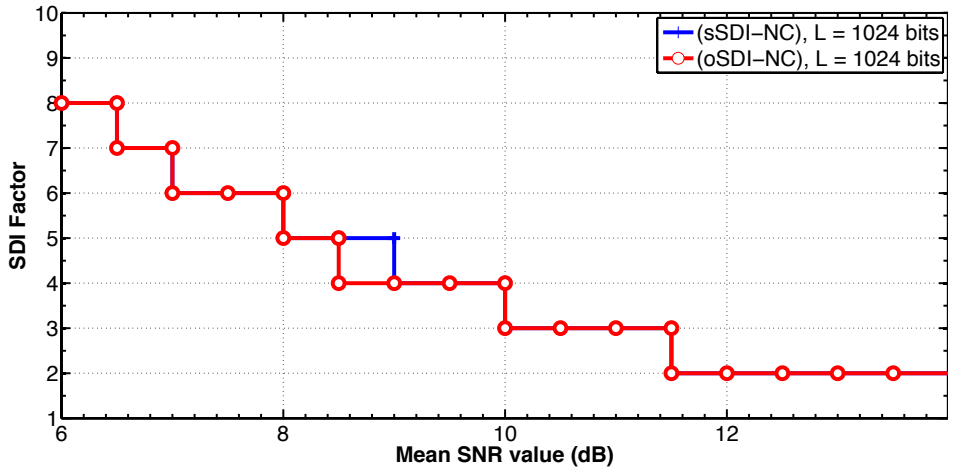


Figure 2.17: Optimal SDI factors vs. the mean SNR among users in faded propagation conditions and scenario I .

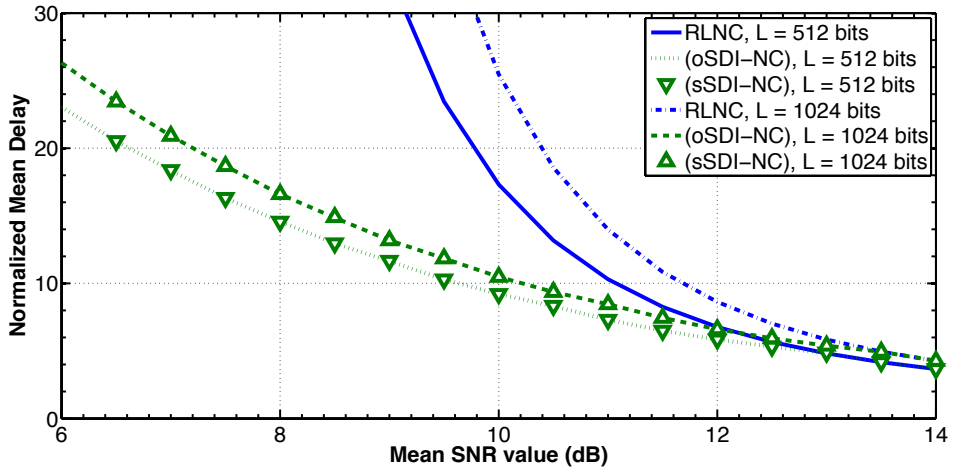


Figure 2.18: Normalized mean delivery delay vs. the mean SNR among users (faded propagation conditions and scenario I).

values are equal and belong to the $[0, 10]$ dB set);

- communications rely on the SDI-NC approach are characterized by a generation length of 10 packets with $K \cong N$;
- the packet length has been set equal to 512 and 1024 bits.

Numerical results (under AWGN and a slow fading regime), obtained by resorting to computer simulations, are given in Figure 2.20. This figure shows the normalized mean E2E delivery delay (per-information packet) as a function of the (mean) SNR values associated to the receiving nodes of the $\overbrace{RCD}$ network. Figure 2.20 clearly points out

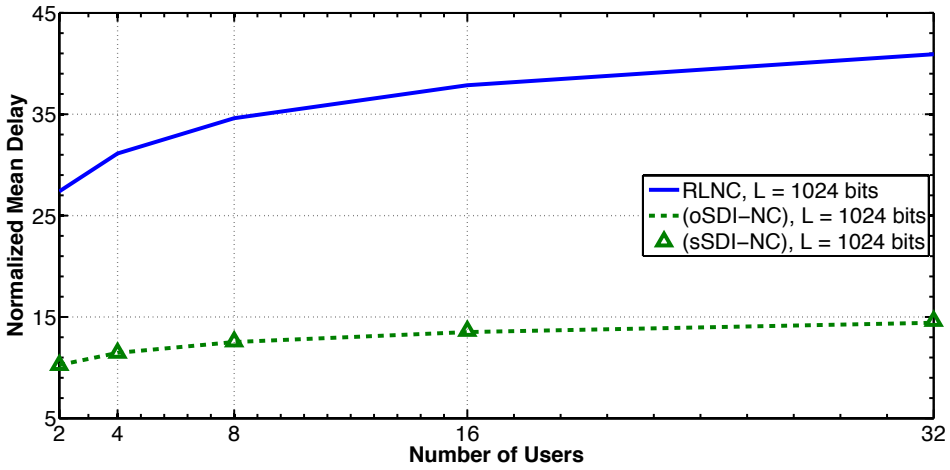


Figure 2.19: Normalized mean delivery delay vs. the number of receiving nodes for faded propagation conditions (scenario II).

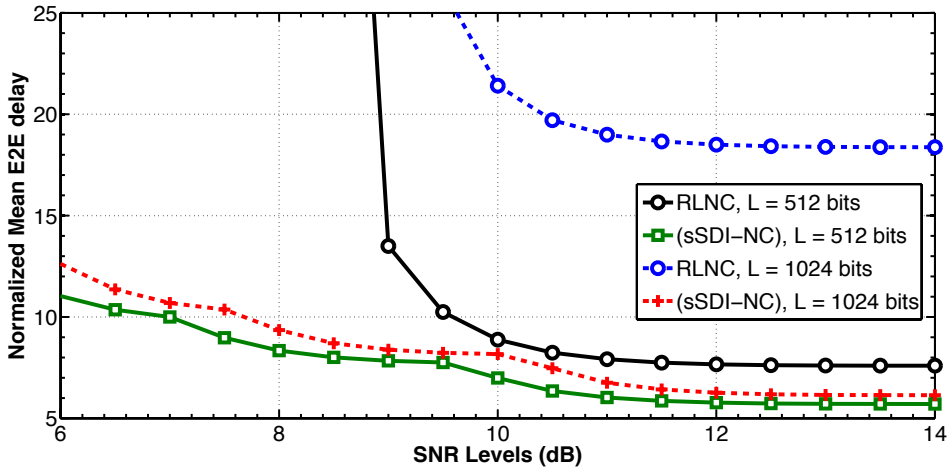
that, regardless of the chosen optimization and channel model, the SDI-NC scheme outperforms the classical RLNC scheme.

In this section we proposed a novel RLNC scheme based on the SC approach. Results presented in the section clearly show that the proposed SDI-NC scheme: (i) can be easily integrated within an existing RLNC implementation, (ii) is characterized by the same implementation complexity of a system adopting the SC principle, and (iii) can be successfully optimized by resorting to a convex heuristic approach. Finally, we demonstrated that the optimized SC-NC schemes significantly outperforms the classical RLNC scheme.

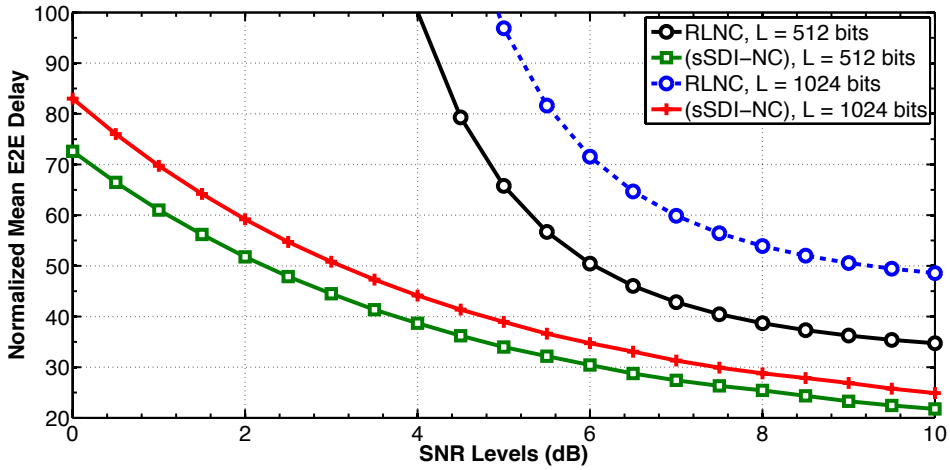
2.2.4 Short Message Transmission over Satellite Networks

This section draws inspiration from the proposal presented in Section 2.2.2. In particular, we extend the aforementioned proposal to a generic satellite network. More precisely: (i) we focus on the issue of reliably deliver of short messages through satellite links, and (ii) we aim at minimizing the overall communication delay on a MG basis.

In the field of satellite networks, reliability of broadcast communication has been addressed by means of Forward Error Correction (FEC) codes. Modern systems can have a feedback channel, e.g., DVB-RCS or DVB-RCS2, allowing the adoption of an ARQ-based error control protocol. However, the existing approaches are not suitable for short-duration communications, for e.g., traffic information, meteorological conditions and alerting systems, as in this case the message-based services are usually characterized by short, sporadic messages that have to be broadcasted reliably to a large number of clients [54] and the ARQ-based systems suffer from scalability issues. This can be



(a) AWGN regime



(b) Slow fading regime

Figure 2.20: Mean normalized end-to-end delivery delay as a function of the SNR (or mean SNR) values at the C and D sides.

overcome by using the NC principle as error control strategy [55].

Differently from the approaches previously proposed in the literature, this section proposes a Modified NC (MNC) scheme where the transmission of each packet is iteratively repeated m times, with m assumed as a system parameter. It is straightforward to note that this is equivalent to increase the duration of each symbol of a factor m with respect to the solution proposed in Section 2.2.2. However and unlike Section 2.2.2, we assume that m copies of the same packet are soft-combined symbol-by-symbol by each user during the receiving phase, then the network client decides if the packet has been correctly received (or not).

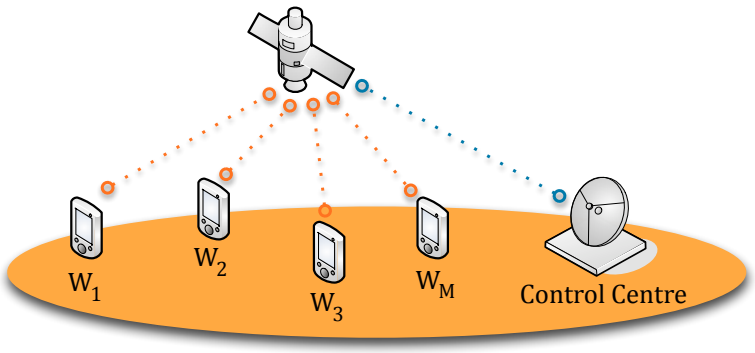


Figure 2.21: Considered GEO satellite system model.

2.2.4.1 System Model

In this section we will consider a Geostationary satellite (GEO) system (sketched by Figure 2.21) where a Control Centre (CC) can send short data messages to M users (namely, $W_1, W_2, \ldots, W_M$) spread across a region by mean of the satellite system. It is outside of the scope of the present work to identify the particular satellite standard, as the proposed approach can work with any satellite system able to send short messages.

It is worth noting that the satellite transmits each information message (i.e., the short messages) by using the NC principle (see Section 2.2.1.1). Whenever a receiving node has successfully recovered an information message, it sends back an ACK[1] to the satellite. As soon as all the nodes have successfully recovered the same message, the satellite can start the transmission of a new one. For the sake of the analysis, we assumed that NC operations are implemented on-board the satellite. However, the provided theoretical framework remains valid even when the NC process is implemented in the CC. In carrying out our analysis we have referred to the QPSK modulation scheme[2] for the data transmission in the considered scenarios.

2.2.4.2 MNC Principle

This section deals with the performance evaluation of the proposed MNC scheme by considering the broadcast network topology (sketched in Figure 2.21), under Rician propagation conditions. The optimization strategy proposed in this section comes from the model proposed in Section 2.2.2. However, in this case we will show how to optimise the value of m in order to minimize the mean overall communication delay. In the rest

[1]We have assumed here, without any loss of generality, that ACK messages are sent across a fully reliable channel.

[2]Results provided in this section are quite general and can be easily extended to different modulation schemes, like 8PSK, 16APSK, or 32APSK, etc.

of this section, the parameter m will be called Repetition Index (RI).

Let us consider the case of a slow Rician fading regime. In particular let us assume that: (i) propagation conditions are kept constant during the transmission of m copies of the same coded packet (regardless of the value of m), (ii) the M communication channels are independent, and (iii) by using suitable radio resource allocation scheme, they can be considered statistically independent. For these reasons we can assume that losses of coded packets at each receiving end occur as statistically independent events. Due to the fact that the channel fading is slow, we can assume an ideal coherent detection at each receiving node side.

Let $\gamma_i = \alpha_i^2 \frac{E_{b,i}}{N_0}$ be the SNR at the W_i side where: α_i^2 is a noncentral χ^2-distributed random variable (with two degrees of freedom) representing the squared magnitude of a Rice channel coefficient (α_i). The parameter $E_{b,i}$ is the energy associated to each symbol received by W_i. The probability density function of γ_i [56] is given by

$$r(\gamma_i) = \left(\frac{1+V}{\overline{\gamma}} \right) e^{-\frac{(1+V)\gamma_i + V\overline{\gamma}}{\overline{\gamma}}} I_0 \left(2\sqrt{\frac{V(1+V)\gamma_i}{\overline{\gamma}}} \right) \tag{2.44}$$

where $\overline{\gamma}$ is the average SNR characterizing each receiving node, V is the well known Rician parameter [56], and $I_0(\cdot)$ is the 0-th order modified Bessel function of the first kind. For these reasons the $P_C(m)$ in this case can be expressed as:

$$P_C(m) = \int_0^\infty \Big[1 - P_e(m,\gamma) \Big]^L r(\gamma) \;\; d\gamma. \tag{2.45}$$

where $P_e(m,\gamma) = Q\Big(\sqrt{m\,\gamma}\Big)$ is the expression of the bit error probability in the case of the QPSK modulation.

Let m and K be the chosen RI factor and the generation length, respectively. Let us extend the combinatorial analysis presented in Section 2.2.2 by considering the random variable S_i (for $i = 1, \ldots, M$) which represent the number of transmission attempts performed by the satellite to ensure the correct reception of an information message by W_i (i.e., to ensure the correct reception of K linearly independent coded packets). Moreover, let

$$H = \max_{i=1,\ldots,M} \Big\{ S_i \Big\} \tag{2.46}$$

be the random variable representing the number of transmission attempts performed by the satellite to ensure the correct reception of a generation by all the receiving nodes. The function expressing the probability that $S_i \leq j$ (for $j \geq K$) can be expressed

as [40, 57]:

$$f_i(j) = \sum_{a=K}^{j} \binom{j}{a} P_C^a(m) \Big[1 - P_C(m)\Big]^{j-a} p_{NC}(a, K) \tag{2.47}$$

where

$$p_{NC}(a, K) = \prod_{b=0}^{K-1} \Big[1 - \frac{1}{q^{a-b}}\Big]. \tag{2.48}$$

The term $p_{NC}(a, K)$ expresses the probability that at least K over $a \geq K$ coded packets (belonging to the same generation) are linearly independent [40]. For $j < K$, $f_i(j)$ is null.

From (2.46) and (2.47), the average value of the random variable H can be expressed as [51]:

$$\begin{aligned} \zeta_{NC}(m) &= \sum_{n=0}^{\infty} n \Bigg\{ \text{Prob}\Big\{H \leq n\Big\} - \text{Prob}\Big\{H \leq n-1\Big\} \Bigg\} \\ &= \sum_{n=N}^{\infty} n \Bigg\{ \prod_{r=1}^{M} f_r(n) - \prod_{r=1}^{M} f_r(n-1) \Bigg\}. \end{aligned} \tag{2.49}$$

From (2.49) we define the mean broadcast delay as the mean time needed by all the network nodes to successfully receive an information message (K packets long). It can be expressed as follows:

$$\tilde{\Lambda}_{NC}(m) = m \; \zeta_{NC}(m) \; T_s \tag{2.50}$$

where T_s is the mean propagation delay of a coded packet when the RI factor is equal to one (i.e., $m = 1$). In order to make our analysis general, we refer in what follows to the mean broadcast delay ($\Lambda_{NC}(m)$), normalized with respect to the parameter K and T_s:

$$\Lambda_{NC}(m) = \frac{\tilde{\Lambda}_{NC}(m)}{K \; T_s} = \frac{m \; \zeta_{NC}(m)}{K}. \tag{2.51}$$

Further, it is important to note that the normalized mean broadcast delay of the classical RLNC scheme is equal to $\Lambda_{NC}(1)$, i.e., it is equal to $\Lambda_{NC}(m)$ for $m = 1$.

As a consequence, we can derive the optimal RI factor by solving the following optimization problem:

$$\text{(oMNC)} \qquad \text{minimize} \quad \Lambda_{NC}(m) \tag{2.52}$$

$$\text{subject to} \quad m \in \mathbb{N} \tag{2.53}$$

The oMNC is an integer nonlinear optimization problem whose solution is hard to derive

in a closed form, hence we need to resort to derivative-free methods[1].

As an alternative to the previous approach, we propose to consider a novel heuristic model. It is characterized by a reduced computational complexity, and at the same time, the obtained solution is close enough to the optimal one. It is useful to define the mean link delay ($\hat{\lambda}(m)$) as: the mean time required by a given receiving node to successfully collect an arbitrary number $N \geq K$ of coded packets. It can be expressed by:

$$\hat{\lambda}(m) = \frac{m\, N\, T_s}{P_C(m)}. \tag{2.54}$$

Let $\mathcal{P}$ be the set of real numbers equal to or greater than one; the mean link delay function, normalized to $G\, T_s$, $\Lambda_{L2L}(m) : \mathbb{N} \to \mathcal{P}$ results to be:

$$\Lambda_{L2L}(m) = \frac{\hat{\lambda}(m)}{G\, T_s} = \frac{m}{P_C(m)}. \tag{2.55}$$

It is worth noting that the (2.55) is equivalent to (2.27). However, in this case, m represents the RI. In addition, let us consider the idea underlying the esSDI-NC problems: it aims at optimizing the SDI factor only considering the network user affected by the worst propagation conditions. In this case we assume that all the users are characterized by the same average SNR. Hence, we proposed to heuristically optimize the RI factor by the following model:

$$\text{(hMNC)} \qquad \text{minimize} \quad \Lambda_{L2L}(m), \tag{2.56}$$

$$\text{subject to} \quad m \in \mathbb{N}. \tag{2.57}$$

In order to solve the hMNC optimization problem, let us prove the proposition below:

Proposition 2.2. Let $\hat{\Lambda}_{L2L}(\hat{m}) : \mathcal{P} \to \mathcal{P}$ be the continuous expansion of the function $\Lambda_{L2L}(m) : \mathbb{N} \to \mathcal{P}$. The function $\hat{\Lambda}_{L2L}(\hat{m})$ is continuously differentiable and convex on its domain.

Proof. See the Appendix C.2.1. □

For the sake of clarity, proving Proposition 2.2 is equivalent to prove Proposition 2.1 under Rician propagation conditions.

In order to solve the hMNC problem, let us consider the following one:

$$\text{(rMNC)} \qquad \text{minimize} \quad \hat{\Lambda}_{L2L}(\hat{m}), \tag{2.58}$$

$$\text{subject to} \quad \hat{m} \in \mathcal{P}. \tag{2.59}$$

[1]In particular, in this section we have resorted to the NOMAD solver [27].

Since $\hat{\Lambda}_{L2L}(\hat{m})$ is convex in its domain see (Proposition 2.2), the solution $(\hat{m}_o)$ of the rMNC problem is the real root (if it exists) of the following equation:

$$\frac{d}{d\,\hat{m}}\Big(\hat{\Lambda}_{L2L}(\hat{m})\Big) = 0 \Leftrightarrow w(\hat{m}) - t(\hat{m}) = 0 \tag{2.60}$$

where

$$w(\hat{m}) \doteq \int_0^{\infty} \Big[1 - Q(\sqrt{\hat{m}\,\gamma})\Big]^L r(\gamma)\ \ d\gamma \tag{2.61}$$

and

$$t(\hat{m}) \doteq \frac{L\sqrt{\hat{m}}}{2\sqrt{2\pi}} \int_0^{\infty} \sqrt{\gamma}\,\Big[1 - Q(\sqrt{\hat{m}\,\gamma})\Big]^{L-1} e^{-\frac{\hat{m}\gamma}{2}} r(\gamma)\ d\gamma. \tag{2.62}$$

In particular, the solution (m_o) of the hMNC problem is represented by $\lceil \hat{m}_o \rceil$ or $\lfloor \hat{m}_o \rfloor$. Hence, the hMNC problem can be easily solved by choosing the value minimizing the (2.56). If (2.60) has no real root, the solution of the hMNC problem is $m_o = 1$. Even though a closed-form solution to the (2.60) is not achievable, its convexity ensures that the rMNC problem can be efficiently solved by resorting to suitable numerical approaches [58]. In particular, we have resorted here to the CVX solver [59].

2.2.4.3 Numerical Results

This section, compares the performance of the classical RLNC with the optimized MNC schemes. We have considered two different scenarios:

I The satellite transmits packets 42 or 21 bytes long to 20 receivers (i.e., $M = 20$). Each receiving node is characterized by the same average SNR value $\overline{\gamma} \in [0, 10]$ dB;

II The satellite transmits to a variable number of nodes packets 42 bytes long. Also in this case, we have that each node is characterized by the same mean SNR value: $\overline{\gamma} = 5$ dB.

Moreover, we assumed for all the RLNC-based communications the finite field size $q = 2^2$ (i.e., all the items of coding vectors are 2 bits long) and a generation length of 20 information packets. Finally, we set the parameter V to 5 dB.

Figure 2.22 shows the normalized mean broadcast delay as a function of the SNR value. On the other hand, Figure 2.23 shows the same performance metric as a function of the number of receiving nodes in the network scenario II. We can note that both the oMNC and hMNC models minimize the normalized mean broadcast delay of the MNC scheme if compared to the RLNC. For instance, Figure 2.23 shows that in a

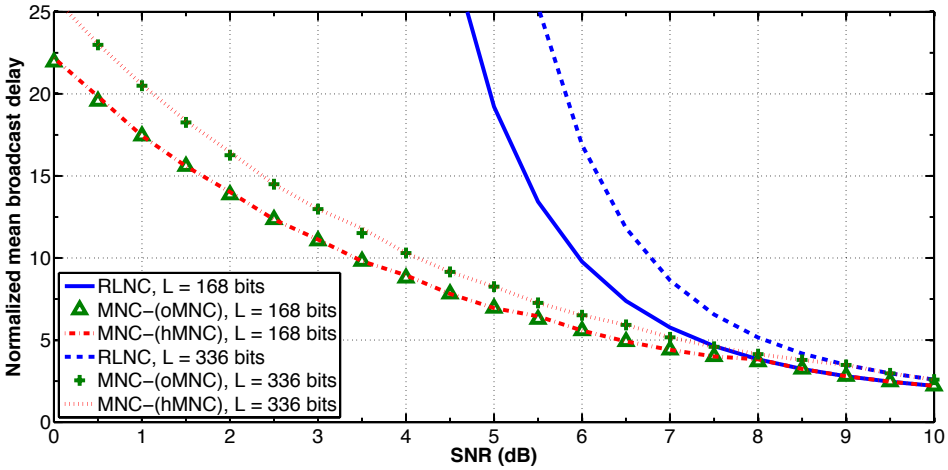


Figure 2.22: Normalized mean broadcast delay of RLNC and MNC vs. the average SNR value (Rician propagation conditions).

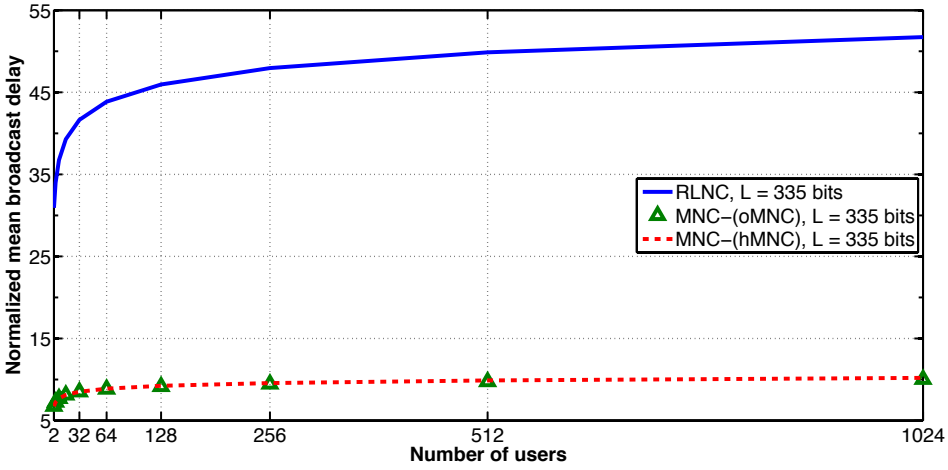


Figure 2.23: Normalized mean broadcast delay of RLNC and MNC vs. the number of receiving nodes (Rician propagation conditions).

network composed by $M = 1024$ users, the MNC (optimized by the hMNC model) is characterized by a (normalized) mean broadcast delay that is almost 5-fold smaller than that of the RLNC. In addition to that, the hMNC model is characterized by almost the same performance that we would have by using the oMNC model. Finally, numerical results show that the optimized MNC scheme is characterized by a performance gain of almost 5-12 fold if compared to the classical RLNC scheme.

2.2.5 Cross-Layer Design of Reliable Multicasting in LTE

The theoretical model presented by Section 2.2.2 can also be used to implement an energy-aware communication scheme that is suitable for multicast and broadcast data delivery over LTE and LTE-Advanced networks. In this case, the proposed communications scheme minimises the average energy consumption of the macro base station that is required to deliver a message to all users in a multicast group.

In Section 1.3.2, we outlined the basic principles of the eMBMS communication framework. In brief, 3GPP standards have proposed to eMBMS transmission schemes: the SC-eMBMS and the MBSFN-eMBMS one. The latter one is very beneficial for users positioned close to the cell-edge but MCSs of delivered services are chosen by the core network and are same for all eNBs [17, 4]. It is straightforward to note that the single-cell approach is more convenient than the other if: (i) users targeted by the multicast/broadcast services belong to the same cell, or (ii) carrier needs to control the service level of communications on a per-cell basis. In this section, we consider a SC-eMBMS deployment where the eNB delivers broadcast services to all UEs that belong to one cell.

The energy consumption of 3G and 4G cellular networks is mainly attributed to macro base stations (eNBs for short) and represents a major challenge to mobile network operators [60]. Nowadays, political and national initiatives currently support trends towards energy-saving in information technology and telecommunications, with the specific aim to lower the CO_2 emissions. Traditional solutions to reduce the energy demand are focused on reducing the eNB transmission power or even switching-off macro eNBs with low or no traffic load services [61]. These solutions should take into consideration the user QoS and the channel propagation conditions in order to avoid possible degradation in the user QoS with reduced energy usage. For example, in [62] the authors propose an energy-efficient optimization strategy that jointly considers the radio resource allocation process and the user admission control.

Reduced complexity Application Layer Random Linear Network Coding (AL-RLNC) solutions are currently proposed as alternative to the classical Application Layer-FEC (AL-FEC) schemes [32] that are used for multimedia delivery in MBMS [63]. AL-RLNC solutions already provide reliable multimedia delivery over wireless networks [64]. However, given the large end-to-end delays between the application-layer entities compared to the short message transmission time, authors in [65] proposed the integration of the RLNC solution into the MAC layer of the LTE/LTE-A RAN protocol stack. MAC-RNC exploits the very small round-trip delay at the MAC layer to reduce the amount of redundancy produced by AL-RLNC. This makes MAC-RNC a suitable solution for multimedia delivery over LTE-A [66].

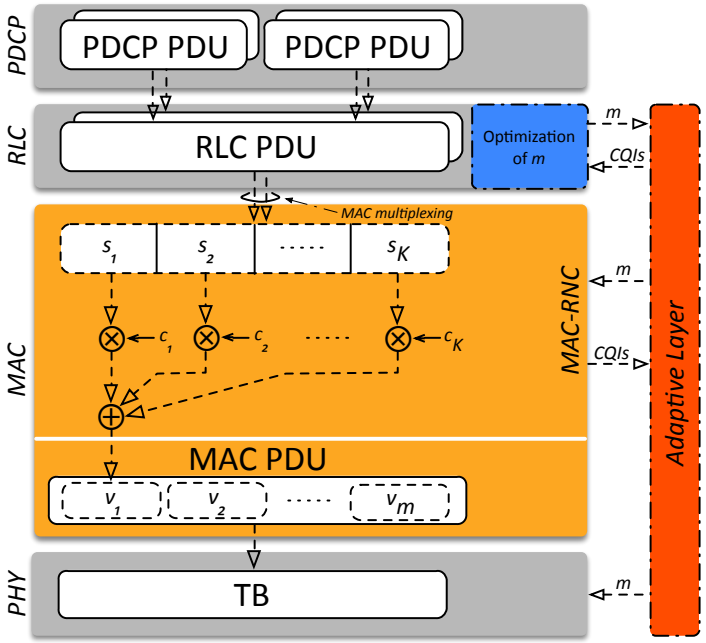


Figure 2.24: E-RLNC optimized architecture for eNB and UE.

In this section, we extend the work presented in [65] to realise a fully reliable and energy-aware communication protocol that is suitable for eMBMS delivery over LTE-A. The proposed Extended-RLNC (E-RLNC) represents a possible implementation of the SDI-NC principle presented in Section 2.2.2. Briefly, the proposed E-RLNC aims at minimising the average energy consumption required to deliver a message to all users in a multicast group, by leveraging all available information on users' channel conditions and requested QoS to optimise the overall number of transmitted packets that ensure correct delivery of the transmitted message.

2.2.5.1 System Model

The integration of the RLNC communication scheme within the MAC layer requires some changes to the LTE-A protocol stack. Figure 2.24 shows the integration of the E-RLNC scheme in the LTE-A protocol structure and the cross-layer interactions to process a downlink data flow at the macro eNB side.

Let us consider a similar information flow to that in Figure 2.24 that is directed to a MG composed of several mobile UEs. Starting at the PDCP layer the PDUs are segmented/concatenated at the RLC to produce RLC PDUs of suitable length. A given RLC PDU directed to a MG will be kept in an appropriate buffer until all the UEs belonging to that MG will have successfully acknowledged (by an ACK message) the reception of the PDU itself. The RLC PDUs are then forwarded to the MAC layer. Each

MAC SDU represents the source message of the MAC-RNC sublayer. This sublayer, proposed in [65], splits the MAC SDU into K equal-length source symbols (namely, $s_1, s_2, \ldots, s_K$) and then encodes them in a RLNC fashion [67] (see Section 2.2.1.1). The output of the coding process (i.e., the output of the MAC-RNC sublayer) is a stream of coded symbols $v_i = \sum_{j=1}^{K} c_j \cdot s_j$ (where c_j, for $j = 1, 2, \ldots, K$, is the j-th coding vector).

Unlike [65], the protocol stack we are proposing in this section is characterized by a MAC layer that generates a stream of MAC PDUs, where each PDU comprises m copies of the same coded symbol. All the MAC PDUs are forwarded to the PHY layer and mapped on different TBs[1]. Each UE can recover the original coded symbol (v_i) by soft-combining (according to the maximum likelihood principle) the m copies of v_i forming the MAC PDU. Whenever all the UEs have recovered the data message (i.e., when all the UEs has collected at least K linearly independent coded symbols), the eNB starts the transmission of the next one (it starts the transmission of the next MAC SDU). Finally, it is straightforward to note that the proposed approach draws inspiration from the proposal described in Section 2.2.4.

In Section 2.2.5.2 will be discussed how to efficiently optimise the number of copies of the same coded symbols held by a TB. That is a key aspect in the energy efficient error control protocol we are proposing. If the m value is too small, the error probability of each coded symbol cannot be effectively reduced. However, as the value of m increases, the energy need to transmit a TB becomes bigger. The optimum value of m is a trade-off between the minimisation of the TB error probability and energy needed to broadcast the TB itself.

It is important to note that due to the fact that a multicast/broadcast communication flow adopts the Unacknowledged Mode at the RLC layer, it cannot adopt any classical error control strategy such as ARQ or HARQ strategies. However, [68] shows that an eNB can collect ACK messages from UEs targeted by eMBMS services. This can be achieved for the UEs transmitting through the PUCCH channel that is commonly used in the case of the HARQ [69].

A key aspect of the error control strategy that we are proposing is represented by the optimization of the index m as a system parameter (i.e., a parameter shared among the eNB and the UEs belonging to the same MG). As will be described in the next section, the optimization algorithm should determine the best choice of m by considering the following inputs:

- the channel conditions reported by each UE, using the CQI, in the MG

[1]We assume here that a TB can hold just one MAC PDU.

- the length L (expressed in bits) of each information symbol
- the Modulation and Coding (MC) scheme used to transmit a given eMBMS flow.

In our model the optimization of the index m is in charge of the RLC layer as this layer is able to control and thus optimize different multicast communication flows independently (i.e., on a MG basis). Moreover, during the optimisation process the following cross-layer interactions are performed between RLC, MAC and PHY layers:

- the MAC and PHY layers have to know the index m in oder to provide a feasible allocation of TBs within subframes, and to correctly perform soft-combining of different copies of the same coded symbol, respectively;
- the MAC and RLC layers share all information related to the propagation conditions of the UEs.

The optimized m value and the reported CQI values of all UEs in the same MG are shared among layers of the communication stack by the adaptive layer [70].

2.2.5.2 Energy Efficient E-RLNC Scheme

A multicast communication in a eMBMS network can be modelled efficiently as a classical AP where a node (namely, the eNB) transmits the same data to a set of devices (the UEs). Figure 2.1 shows the network topology of a MG composed by M UEs. Let W_i (for $i = 1, \ldots, M$) be the i-th UE of a MG. Whenever W_i is able to recover the overall information message (transmitted according to the RLNC principle), it sends an ACK to the eNB. The eNB continues to transmit coded symbols belonging to the same message until each UE of the MG has successfully recovered the information message (i.e., until the eNB has collected a number of ACKs equal to M).

The theoretical derivation proposed in the rest of the section assumes that all the downlink transmissions directed to the members of a given MG adopt QPSK modulation scheme[1].

The E-RLNC scheme described in Section 2.2.5.1 is characterized by only one optimization variable: the index m. In this section we define an energy efficient model for the optimization of m that is able to minimize the energy related to the transmission of a message.

Let us consider again a given MG whose network topology is reported in Figure 2.1. In addition to this, we have assumed that: (i) the losses of different TBs are statistically

[1]It is important to note that the proposed analysis is general and can be easily extended to different modulation schemes, like 16-QAM or 64-QAM.

independent events, (ii) the propagation conditions are constant within a TB but are statistically independent between different downlink communication links. The i-th UE belonging to the considered MG receives TBs with the Signal-to-Noise Ratio (SNR) $\gamma_i = \alpha_i^2 \frac{E_b}{N_{0,i}}$ where: (i) α_i^2 is a random variable equal to the square of the Rayleigh channel coefficient, (ii) E_b is the energy associated to each symbol, and (iii) $N_{0,i}$ is the total noise power associated to each symbol received by the i-th UE.

Hence, as expressed in (2.38), the TB error probability characterizing the i-th UE can be given by:

$$P_i(m) = 1 - \int_0^{\infty} \frac{1}{\overline{\gamma}_i} \Big[1 - p_i(m)\Big]^L e^{-\frac{1}{\overline{\gamma}_i}\gamma_i} \; d\gamma_i \tag{2.63}$$

where L is the coded symbol length (expressed in bits), $\overline{\gamma}_i$ is the average SNR (at the u_i side), and the $f(\cdot)$ is the probability density function of γ_i. Finally, the term $p_i(m)$ defines the bit error probability (as a function of the index m) affecting the reception at the u_i side. In the rest of this section, we refer to the bit error probability of the QPSK modulation (see (2.20)): $p_i(m) = \frac{1}{2}\,\mathrm{erfc}\Big(\sqrt{\frac{r\gamma_i}{2}}\Big)$.

Let us consider the combinatorial analysis reported in Section 2.2.4.2. In addition, considering that U_i (for $i = 1, \ldots, M$) is the number of TBs transmitted by the eNB until W_i has successfully recovered the information message, and that the random variable T gives the total number of TBs required to deliver the same message to all UEs in the MG. Hence, $T = \max_{i=1,\ldots,M} \Big\{U_i\Big\}$.

Unlike 2.47, in this case we have that the i-th UE recovers the information message within N transmissions if: (i) W_i will be able to successfully receive a given amount of error-free TBs (i.e., error-free coded symbols), (ii) these ones are linearly independent, and (iii) the ACK will be successfully received by the eNB. Hence, the probability that U_i is equal to or less than N (for $N \geq K$) is [40]:

$$F_i(N) = h_i(N) \sum_{a=K}^{N} \binom{N}{a} P_i^{N-a}(m) \Big[1 - P_i(m)\Big]^a g(a). \tag{2.64}$$

The probability that an ACK message (transmitted by u_i) is successfully received by the eNB within N trials is

$$h_i(N) = 1 - P_{ack,i}^{N-K+1} \tag{2.65}$$

where $P_{ack,i}$ is the error probability of an ACK message transmitted by the i-th UE. Likewise (2.48), the term $g(a)$ expresses the probability that at least K symbols over

a, for $a \geq K$, are linearly independent, and in this case, can be restated as:

$$g(a) = \prod_{t=0}^{K-1} \Big[1 - \frac{1}{q^{a-t}}\Big].$$

We remark that q is the finite field size.

From the (2.64), and likewise to the result reported in (2.49), the mean value of T is:

$$\overline{t}_{nc}(m) = \sum_{n=K}^{\infty} n \Bigg\{ \prod_{v=1}^{M} F_v(n) - \prod_{v=1}^{M} F_v(n-1) \Bigg\}. \tag{2.66}$$

Let $\hat{E}_b$ be the energy associated to the transmission of a TB using $m = 1$. The mean energy consumption $e_{nc}(m)$, normalized by $\hat{E}_b$ is

$$\overline{e}_{nc}(m) = r \; t_{nc}(m). \tag{2.67}$$

For $m = 1$, (2.67) expresses the (normalized) mean energy consumption of an information message by using RLNC in the place of the E-RLNC.

The E-RLNC scheme improves energy efficiency by choosing the index m minimizing the mean energy consumption of a MG. This is the aim of the following optimization problem:

$$\begin{aligned} \text{(P1)} \qquad & \text{minimize} \quad \overline{e}_{nc}(m) && (2.68) \\ & \text{subject to} \quad r \in \mathbb{N} \setminus \{0\}. && (2.69) \end{aligned}$$

Due to the fact that P1 is a nonlinear problem, it is not feasible, from a computational point of view, to solve it in real time. In the rest of this section we will propose an effective heuristic approach to overcome this problem.

A Convex Heuristic Model Let $\overline{Z}$ be the mean number of coded symbols (i.e., $\{v_1, v_2, \ldots, v_{\overline{Z}}\}$) that the eNB has to generate to successfully deliver a linearly independent set of MAC PDUs.

Let u_h be the UE of the MG experiencing the worst propagation conditions, we can approximate the mean energy consumption (normalized by $\hat{E}_b \, \overline{Z}$) as follows:

$$\overline{l}(m) = \frac{r}{1 - P_h(m)}. \tag{2.70}$$

As we can note, (2.70) approximates the (2.67).

Let $\tilde{l}(\tilde{m})$ be a function defined similar to (2.70) over a set of real numbers that are equal to or greater than one. Regardless of the value of $\overline{Z}$, from (2.70), we can define the following heuristic strategy for optimizing the index m:

$$\begin{aligned} \text{(P2)} \qquad & \text{minimize} \quad \bar{l}(m), && (2.71)\\ & \text{subject to} \quad r \in \mathbb{N} \setminus \{0\}. && (2.72) \end{aligned}$$

In order to solve P2, we can relax the constraint (2.72) and restate P2 in terms of a simple minimization of $\tilde{l}(\tilde{r})$. It is proved in the Appendix C.1.2 that $\tilde{l}(\tilde{r})$ is convex, hence, it can be minimized for that value of $\tilde{r}$ (namely, $\tilde{r}_o$) that is a solution of the following equation:

$$\begin{aligned} \frac{d}{d\,\tilde{r}}\Big(\tilde{l}(\tilde{r})\Big) = 0 \Leftrightarrow & \sum_{n=0}^{\infty} n\Bigg\{\text{Prob}\Big\{H \leq n\Big\} - \text{Prob}\Big\{H \leq n-1\Big\}\Bigg\} \\ \Leftrightarrow & \int_0^{\infty} \Big[1 - \frac{1}{2}\,\text{erfc}\Big(\sqrt{\frac{r\,\gamma_h}{2}}\Big)\Big]^{L} e^{-\frac{1}{\overline{\gamma}_h}\gamma_h} d\gamma_h \\ & - \frac{L\sqrt{\tilde{r}}}{2\sqrt{2\pi}} \int_0^{\infty} \sqrt{\gamma_h}\,\Big[1 - \frac{1}{2}\,\text{erfc}\Big(\sqrt{\frac{r\,\gamma_h}{2}}\Big)\Big]^{L-1} e^{-\frac{2+\tilde{r}\overline{\gamma}_h}{2\overline{\gamma}_h}\gamma_h} d\gamma_h = 0. \quad (2.73) \end{aligned}$$

Finally, the objective function (2.71) of P2 problem is minimized by that value of $r = r_o$ such that:

- $r_o = 1$, if (2.73) has no real root;
- $r_o = \text{argmin}\{\bar{l}(m)\,|\,r = \lceil \tilde{r}_o \rceil, \lfloor \tilde{r}_o \rfloor\}$, otherwise.

2.2.5.3 Numerical Results

This section shows the numerical results of the proposed E-RLNC scheme by comparing the performance of the classical RLNC (i.e., the E-RLNC scheme with the index m equal to one) to the one of the E-RLNC optimized by resorting to the P1 or P2 optimization model. We remark that the P2 model represents a heuristic strategy able to provide a feasible but suboptimal solution to the P1 problem.

The performance of the proposed E-RLNC scheme is investigated for a MG composed of a variable number of UEs. All the eMBMS communication flows that adopt the RLNC or E-RLNC are characterised by a finite field of size $q = 2^8$, and a generation length $K = 20$. The probability of correct delivery of the ACK message to 99%. Finally, two information symbols lengths L equal to 20 and 40 bytes are used.

In order to inspect the performance of the E-RLNC over the classic RLNC and to show effectiveness of the proposed heuristic approach (i.e., the optimization model P2)

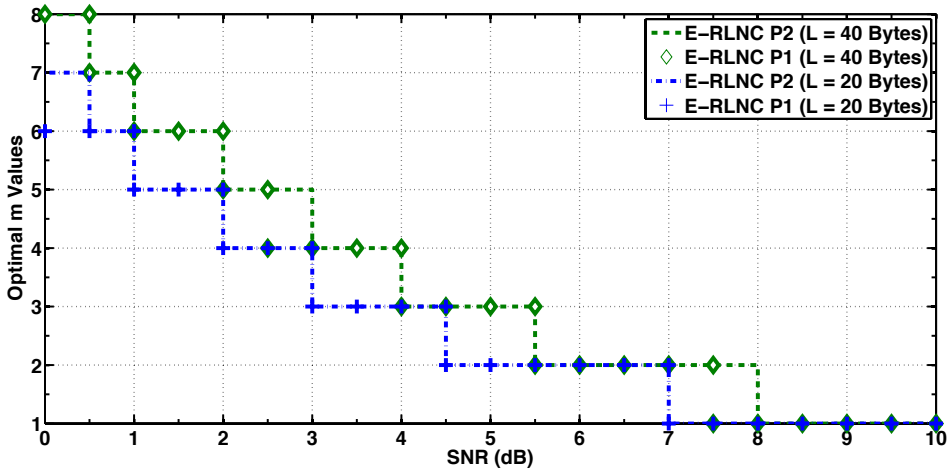


Figure 2.25: Value of m as a function of $\overline{\gamma}_h$ value (scenario A).

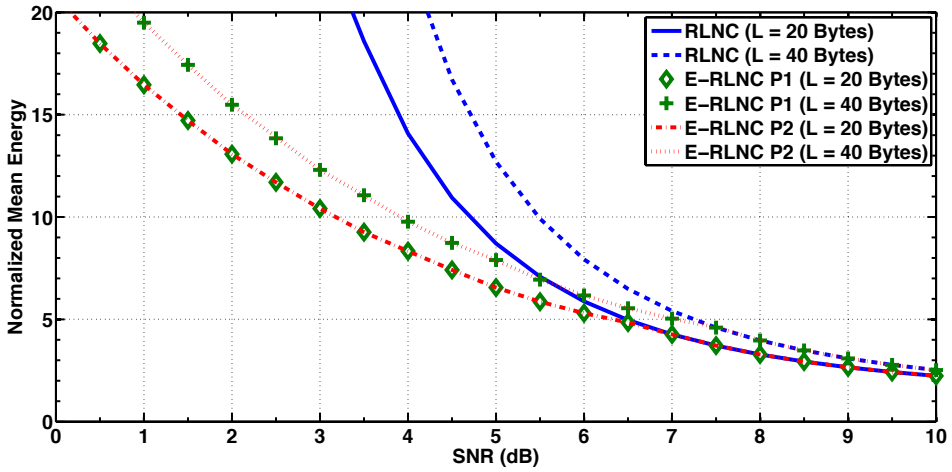


Figure 2.26: Mean energy consumption as a function of $\overline{\gamma}_h$ value (scenario A).

over the optimization model P1, we considered the following scenarios:

A $M = 30$ UEs experience different channel conditions. The 1-st UE and the 30-th are characterized by $\overline{\gamma}_1 = s_o + \Delta$ dB and $\overline{\gamma}_{30} = 10 + \Delta$ dB, respectively. Moreover, $\overline{\gamma}_i = \overline{\gamma}_{i-1} + \Delta$ dB (for $i = 2, \ldots, M-1$). Finally, s_o takes values in the interval $[0, 10]$ dB, and $\Delta = \frac{10}{M-1}$ dB.

B M takes values in the $[2, 128]$ interval, $\overline{\gamma}_i = \tilde{\gamma}$ (for $i = 1, \ldots, M$). The parameter $\tilde{\gamma} = 3.5$ dB.

Let us focus on the network scenario A. Figure 2.25 shows the optimum value of m as a function of the value of $\overline{\gamma}_1$. The figure clearly shows that the indices of m selected by the proposed convex heuristic (i.e., P2) strategy typically overlaps those derived by

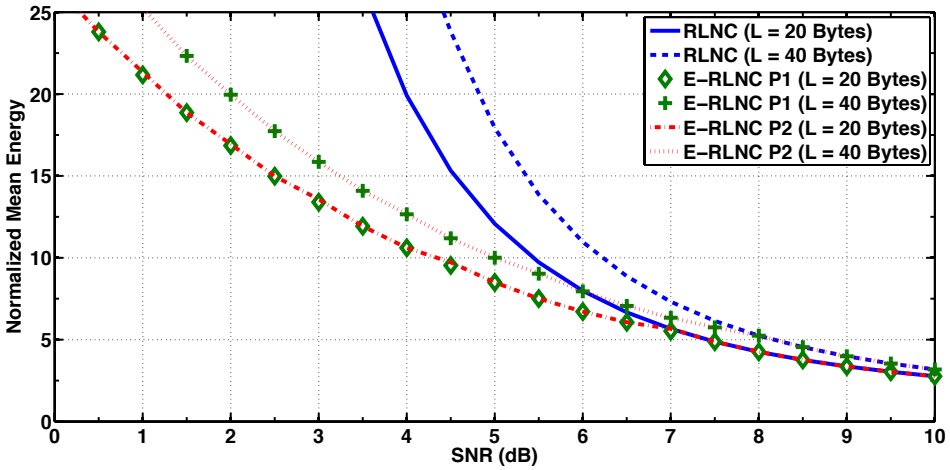


Figure 2.27: Mean energy consumption as a function of $\tilde{\gamma}$ value (scenario B, $M = 30$).

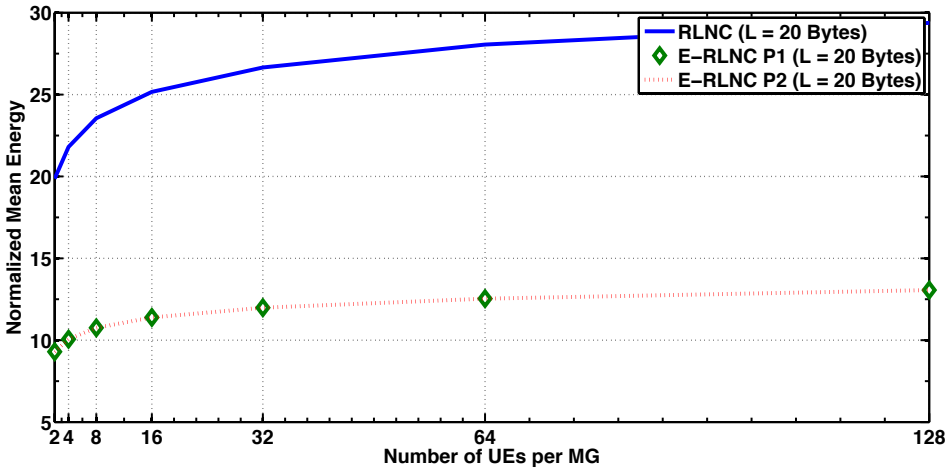


Figure 2.28: Mean energy consumption as a function of number of UEs belonging to the MG (scenario B).

the optimum model (i.e., P1), regardless of the chosen information symbol length. In particular, even if there are some slight differences, it does not have any impact on the overall system performance. This is clearly highlighted by Figure 2.26 showing the normalized mean energy (of a single information symbol) as a function of the $\overline{\gamma}_1$ value. In addition to that, we can note that the E-RLNC clearly outperforms the classical RLNC strategy.

Let us consider network scenario B, where all of the UEs experience virtually the same propagation conditions. Figure 2.27 shows the normalized mean energy consumption (on an information symbol basis) as a function of the $\tilde{\gamma}$. Here again, the performance of the proposed convex heuristic strategy is close to that of the optimum one. Moreover, also in this case we can note that the E-RLNC strategy outperforms

the classical RLNC one.

Finally, let us consider Figure 2.28 it shows the same performance metric as a function of the M value. We can note that: (i) the performance gain of the heuristic strategy overlaps that of the optimum one, regardless of the M value, and (ii) the E-RLNC scheme is characterized by a performance gain of almost two-fold compared to the classical RLNC.

Finally, the numerical results show that the performance of the heuristic strategy is close to that of the optimal model, and that the proposed E-RLNC scheme significantly outperforms the classical RLNC scheme regardless of the optimisation models in use. E-RLNC performance gain of almost two-fold is achieved compared to the classical RLNC scheme.

Chapter 3

Unacknowledged Rate Allocation Strategies

In the previous chapter we inspected a rate adaptation strategy integrated to an error control strategy. In the following, we propose an completely unacknowledged version of the Modified RLNC (MNC) suitable for delay sensitive service delivery. In particular, Section 3.1 provides a couple of multicast communication strategies which aim at significantly improving system performance both in terms of transmission energy and delivery delay associated to the transmission of the whole data flow. Reported analytical results clearly show that both the strategies achieve the aforementioned goals in comparison with the RLNC alternative.

Section 3.2 proposes a rate allocation strategy which draws inspiration form that proposed at the beginning of Section 3.1. However, in this case we aim at optimizing the radio resource allocation process so that users, according to their propagation conditions, can receive video streams at the maximum achievable service level in a LTE-A cell. A key aspect of the proposed system model is that video streams are delivered as eMBMS flows by using the random linear network coding principle. We propose two resource allocation strategies based an optimal and heuristic resource allocation solutions. In particular, we prove that the heuristic strategy is characterized by a reduced computational load. Simulation results show that both strategies significantly outperform alternatives at the state of the art in terms of the service level perceived by users.

3.1 Standard Agnostic Rate and Energy Optimization

In 4G cellular systems, several error control strategies [22] has been proposed, with the aim of supporting downlink PtM flows according to the expected QoS levels. However, these solutions do not consider the following aspects: (i) the number of UEs associated to the same communication flow (i.e., the number of UEs belonging to the same MG), and (ii) the QoS level actually experienced by the MG as a whole. Due to the complexity of managing multicast/broadcast services through retransmission processes, usually 4G communication systems (such as the 3GPP's LTE-Advanced) avoid the use of ARQ or HARQ schemes, simply adopting AL-FEC techniques. Nevertheless, several seminal papers [71, 72] have shown that the overall system performance can be improved by using NC solutions as an alternative to the AL-FEC.

The NC principle, introduced by Ahlswede *et al.* [28], has been shown by Ghaderi *et al.* [32] to be advantageously adopted in PtM communications as error control protocol. Despite to the solid theoretical analysis, that proposal assumes that all the terminals acknowledge the reception of an information message. As a consequence, this approach can not directly applicable to 4G networks, due to the lack of a scalability.

This section aims at proposing a couple of MNC schemes which significantly outperform the RLNC alternative considered in [32] both in terms of the transmission energy cost and delivery delay of the whole data flow. It is worth noting that the proposed communication methods require that all the UEs of the MG have to recover the transmitted flow with a decoding probability which is equal to or grater than a given threshold value.

The proposed MNC schemes achieve the aforementioned goals in two ways: (i) by optimally selecting the transmission data rate, while the power associated to transmission of each packet is kept constant, or (ii) by optimizing the transmission power cost and keeping constant the transmission rate. In the rest of the section we refer to those strategies as Constant Power MNC (CP-MNC) and Constant Rate MNC (CR-MNC), respectively. In addition, a theoretical framework for characterizing and efficiently optimizing both the proposed approaches are presented.

A PtM communication pattern typical of a 3G/4G network can be modelled according to the multicast network model where a BS delivers a multicast service to a set of M UEs forming the MG. The NC communication strategy considered in this section can be summarised as follows.

Let $\mathbf{E} = \{e_1, e_2, \ldots, e_K\}$ be an information message composed by K information packets (see Section 2.2.1.1). The BS linearly combines (in a rateless mode) all the

information packets and transmits a stream of $N \geq K$ coded packets to the UEs of the MG. In the following, the widely known RLNC strategy [67] is considered. Whenever a UE successfully receives (at least) K linearly independent coded packets over N, it can retrieve the original information message $\mathbf{E}$. Conversely, if the number of the received and linearly independent coded packets is less than K the decoding operation fails, the UE can not retrieve the original information message, and it is definitively lost because retransmissions of the same information message are not allowed[1].

It is worth noting that in the case of the CP-MNC scheme each coded packet is transmitted at a rate that is m times smaller than the maximum allowed one (with $m \geq 1$) such that the symbol time duration is m times greater than the nominal value T. On the other hand, the CR-MNC scheme requires that transmission energy of each symbol is m times greater than a target value E (i.e., the power associated to the transmission of each coded packet is m times greater than a nominal value). Differently to CP-MNC, in this case the transmission rate does not change.

In spite of the increasing of the transmission time or transmission energy associated to each coded packet, this section shows that is possible to outperform the RLNC scheme, by means of a suitable optimization of the pair (m, N), both in terms of the overall delivery delay and energy consumption needed to complete the transmission of the set of N coded packets, for a given decoding probability of the information message (on a MG basis).

3.1.1 Optimization Frameworks

Let us consider again the multicast network model presented in Section 3.1, we have assumed slow-faded Rayleigh propagation conditions for all the communication channels linking each UE (of the MG) to the BS. In addition to this, we have also assumed that losses occur independently among UEs of the same MG, and the use of a BPSK modulation scheme[2] in transmission. Hence, in the case of the CP-MNC and CR-MNC approaches, the corresponding signals associated to the transmission of a generic coded packet (for $1 \leq i \leq L$) can be expressed as follows

$$s_i(t) = \begin{cases} \sqrt{\frac{2E}{T}} d_i \cos\left(2\pi f_o t\right) & \text{for CP-MNC with } 0 \leq t \leq mT \\ \sqrt{\frac{2mE}{T}} d_i \cos\left(2\pi f_o t\right) & \text{for CR-MNC with } 0 \leq t \leq T \end{cases} \tag{3.1}$$

[1] We assumed that the BS can not know if a specific UE has actually received an information message because the reception of information messages can not be acknowledged.

[2] However, the theoretical results presented in this section are quite general, as they can be easily extended to different modulation schemes.

where E is the energy of a transmitted signal with a time duration equal to T, f_o is the carrier frequency, L is the number of bits forming each coded packet, and d_i is equal to $+1$ or -1 if the i-th bit of the coded packet is 1 or 0, respectively.

From (3.1), in the case of the CP-MNC method it is important to stress that the amplitude of the transmitted signal is kept constant independently on the value of m. Likewise, for what concerns the CR-MNC, the transmission rate is kept constant regardless of the value of m. In addition, let us assume that the channel fading is sufficiently slow to make possible an exact estimation of the phase shift of the received signal at each receiving end, and hence, an ideal coherent detection can be performed. According to this, and whether CP-MNC or CR-MNC schemes are adopted, it is straightforward to prove that the bit error probability, as a function of m, can be expressed as [50, 73]:

$$p_u(m) = \frac{1}{2}\text{erfc}\Big(\sqrt{m\gamma_u}\Big) \quad \text{for } u = 1, \ldots, M \tag{3.2}$$

where γ_u is the instantaneous SNR related to the signal received by the u-th UE. Hence, according to the assumption of a Rayleigh fading, it follows that γ_u has a chi-square probability distribution with two degrees of freedom and a mean value equal to $\overline{\gamma}_u$. In addition, we have that the packet error probability (for $u = 1, \ldots, M$) for a L bits long coded packet can be expressed as[1]

$$P_u(m) = 1 - \int_0^{\infty} \frac{1}{\overline{\gamma}_u}\Big[1 - p_u(m)\Big]^L e^{-\frac{1}{\overline{\gamma}_u}\gamma_u} \; d\gamma_u. \tag{3.3}$$

As for the combinatorial model used to derive (2.47) and reported in Section 2.2.4.2, let us define from [40] the probability that the u-th UE recovers the information message after that the BS has transmitted N coded packets (where $N \geq K$):

$$F_u(m, N) = \sum_{j=K}^{N} \binom{N}{j} P_u^{N-j}(m)\Big[1 - P_u(m)\Big]^j g(j) \tag{3.4}$$

where

$$g(j) = \prod_{h=0}^{K-1}\Big[1 - \frac{1}{q^{j-h}}\Big] \tag{3.5}$$

which is the probability that at least K coded packets are linearly independent over j (for $j \geq K$) [40].

Hence, for a MG formed by M UEs, the probability that all the UEs of the MG are able to recover the overall information message upon the transmission of N coded

[1]Note that the analysis outlined here is quite general and it includes also the case of different $\overline{\gamma}_u$ values among the UEs of the MG.

packets results to be

$$\Phi(m, N) = \prod_{u=1}^{M} F_u(m, N). \tag{3.6}$$

Moreover, as for the parametric function $\hat{\Phi}_{\hat{N}}(m) = \Phi(m, N)\Big|_{N=\hat{N}}$ (for $\hat{N} \geq K$), it is a monotonically non-decreasing function. It is straightforward to note that the aforementioned statement holds because of the reasons below: (i) from (3.6) we have that $\hat{\Phi}_{\hat{m}}(N)$ is the product of the parametric functions $F_u(m, N)\Big|_{N=\hat{N}}$ (for $u = 1, \ldots, M$) which clearly are nonnegative nondecreasing functions[1], (ii) since the product of nondecreasing functions is nondecreasing too, the parametric function $\hat{\Phi}_{\hat{N}}(m)$ is nonnegative nondecreasing as well.

Let us consider the CP-MNC strategy, we have that $m\,L\,E$ and $m\,L\,T$ is the energy associated to the transmission of a single coded packet and its transmission time, respectively. On the other hand, in the case of CR-MNC the transmission time duration of a coded packet does not change, hence, it is equal to $L\,T$. In addition, in this case the energy cost associated to the transmission of a coded packet results to be equal to $m\,L\,E$. Hence, for both the considered approaches, we have that the overall transmission energy cost, i.e., the overall energy associated to the transmission of N coded packets, is

$$\epsilon(m, N) = m\,L\,E\,N \tag{3.7}$$

while the delivery delay, i.e., the overall time needed to transmit N coded packets, is

$$\delta(m, N) = \begin{cases} m\,L\,T\,N \text{for CP-MNC} \\ L\,T\,N \quad \text{for CR-MNC.} \end{cases} \tag{3.8}$$

From above, it is straightforward to note that the case of $m = 1$ is related to the RLNC scheme [67] (where both the transmission rate and power are kept constant).

3.1.1.1 CP-MNC Model

Let us focus on the CP-MNC strategy. In this Section, we propose an analytical procedure which aims at optimizing the CP-MNC approach. The optimization procedure aims at deriving the optimum values of the transmission rate (i.e., the optimum value of m) and N in order to minimize the overall transmission energy cost and delivery delay, if compared with the RLNC scheme (where m is always equal to 1). In particular, from (3.7) and (3.8) it is worth noting that the goal is achieved by minimizing the

[1]Due to the fact that the probability that the u-th UE recovers the overall information message can not decrease as the packet error probability decreases. In particular, it is straightforward to note that $P_u(m)$ is a non-decreasing function.

object function given by $m\,N$. Hence, our optimization problem can be formulated as follows[1]:

$$\text{(O1)} \qquad \min_{m,N} \; m\,N \tag{3.9}$$

$$\text{subject to} \quad \Phi(m,N) \geq \hat{\Phi} \tag{3.10}$$

$$1 \leq m \leq \hat{m}_{max}, \quad m \in \mathbb{R}^{+} \tag{3.11}$$

$$K \leq N \leq \hat{N}_{max}, \quad N \in \mathbb{N} \tag{3.12}$$

where $\hat{\Phi}$ is the target delivery probability of an information message, i.e., the probability that all the UEs of the MG recover the whole information message. In addition, due to the fact that the transmission time duration and energy associated to an information message are function of m and N, it follows that both m and N have practical upper bounds which dependent on the specific QoS constraints. These practical limits are modelled by constraints (3.11) and (3.12) where we assumed that the parameter $\hat{m}_{max}$ (for $\hat{m}_{max} \geq 1$) and $\hat{N}_{max}$ (for $\hat{N}_{max} > K$) is the upper bound for m and N, respectively.

Due to the fact that O1 is a mixed integer non-linear optimization problem, it can not be solved with reasonable computing efforts. For this reason, the rest of this section shows how to transform O1 into an equivalent problem that can be efficiently solved. In particular, in the rest of the section we show that O1 can be solved by a two-steps procedure which can be summarised as follows: (i) for each $N \in [K, \hat{N}_{max}]$ the optimum value of m is found, then (ii) the pair (m, N) which minimises the objective (3.9) is selected.

Let us consider the problem O1, it can be rewritten as

$$\text{(O2)} \qquad \min_{N} \; N \; \min_{m} \; m \tag{3.13}$$

$$\text{subject to} \quad m \in \mathcal{S}_N \tag{3.14}$$

where the set $\mathcal{S}_N$ is defined as $\mathcal{S}_N \doteq \left\{ m \in \mathbb{R}^{+} \;\middle|\; 1 \leq m \leq \hat{m}_{max} \;\wedge\; \hat{\Phi}_N(m) \geq \hat{\Phi} \right\}$. The problem O2 is a nested optimization model because it consists of two subproblems. In particular, for $N = N^*$ the innermost problem is

$$\text{(I1)} \qquad \min_{m} \; m \tag{3.15}$$

$$\text{subject to} \quad m \in \mathcal{S}_{N^*}. \tag{3.16}$$

In addition, let $\mathcal{S}$ be the feasible set of O1, of course, the relation $\mathcal{S}_{N^*} \subseteq \mathcal{S}$ holds. In

[1] In this section we refer with $\mathbb{R}^{+}$ and $\mathbb{N}$ to the set of real positive and natural numbers, respectively.

particular, due to the fact that $\hat{\Phi}_{N^*}(m)$ is a monotonically non-decreasing function, the solution of I1 is the smallest value of m (for $m \geq 1$) such that the relation $\Phi(m, N^*) \geq \hat{\Phi}$ holds. As a consequence, the problem O2 can be solved by the two-steps procedure below:

i For each $N = K, \ldots, \hat{N}_{max}$ find $m_K, m_{K+1}, \ldots, m_{\hat{N}_{max}}$ which are the solutions to the problem I1 for $N^* = K, K+1, \ldots, \hat{N}_{max}$.

ii The solution of O2 (i.e., the optimal (m, N) pair) is the pair (m_i, i), for $i = K, K+1, \ldots, \hat{N}_{max}$, minimizing the objective function (3.9).

3.1.1.2 CR-MNC Model

Let us consider the CR-MNC approach. In this case the value of m has a direct impact on the power associated to the transmission of each coded packet. Hence, also in this case the value of m has a practical upper bound (which is $\hat{m}_{max}$) depending on the system in use. In addition, we remark that we assumed $m \geq 1$.

Due to the fact that also the CR-MNC method aims at delivering each information message in such a way that it can be received by the considered MG at least with a certain delivery probability $\hat{\Phi}$, the optimization of the pair (m, N) takes places over the same kind of feasible set associated to O1. However, in this case we decided to minimize the overall transmission energy cost, i.e., we decided to minimize the overall power cost associated to the transmission of each information message. Hence, from (3.7), it is straightforward to note that, also in this case the model O1 can be efficiently use to achieve the aforementioned goals. Hence, the optimum pair (m, N) can be found by the two-steps procedure proposed in Section 3.1.1.1. Finally, in Section 3.1.2 we show that not only the overall transmission energy cost is minimized but also the delivery delay is significantly reduced if compared to the RLNC alternative.

Finally, from (3.7) and (3.8) it is straightforward to note that the overall transmission energy cost and delivery delay can not be jointly minimized. In particular, if only the delivery delay is minimized, the value of m will trivially be equal to $\hat{m}_{max}$, i.e., the power associated to the transmission of each coded packet is maximized.

3.1.2 Numerical Results

The performance of both CP-MNC and CR-MNC strategies optimized as proposed in the previous section are hereafter evaluated. We investigate the system performance in terms of the normalized overall transmission energy cost and delivery delay, defined as $\epsilon(m, N)/(K\,L\,E)$ and $\delta(m, N)/(K\,L\,T)$, respectively. From (3.7) and (3.8) we have

that, in the case of CP-MNC, the normalized overall transmission energy cost and delivery delay are equal.

In addition, we compare the MNC-based strategies to an RLNC approach where the value of N has been optimized by the flowing model (which is based on O1):

$$\text{(O3)} \qquad \min_{N} \; N \tag{3.17}$$

$$\text{subject to} \quad \Phi(m,N)\Big|_{m=1} \geq \hat{\Phi} \tag{3.18}$$

$$K \leq N \leq \hat{N}_{max}, \quad N \in \mathbb{N}. \tag{3.19}$$

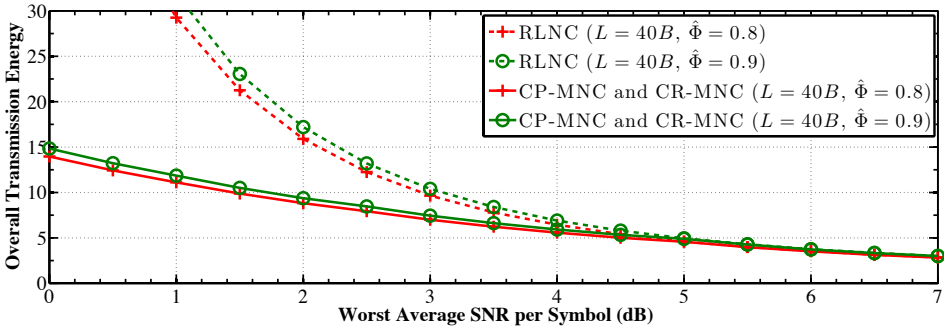
It is worth noting that in the case of RLNC, the transmission time duration, energy and power associated to each coded packet never change (i.e., the value of m is always equal to 1).

The scenarios herein considered refer to a MG composed by a variable number of UEs (namely, $M \in [2, 128]$). We assumed a finite field size equal to 2^8 for all the NC-based coding/decoding operations and a message length of $K = 20$ information packets. Two different packet lengths ($L = 20$ bytes and $L = 40$ bytes) have been considered. Finally, regardless of the considered MNC-based strategy, parameters $\hat{m}_{max}$ and $\hat{N}_{max}$ have been set equal to 10 and $50\,K$, respectively.

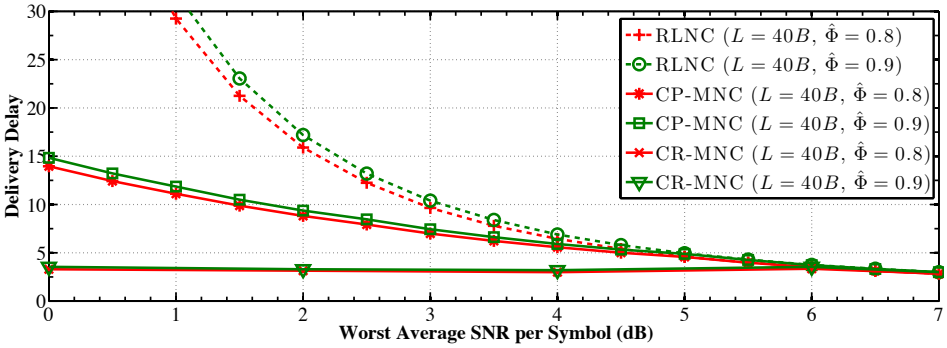
Figure 3.1a and Figure 3.1b show, for $M = 20$, the values of $\epsilon(m,N)/(K\,L\,E)$ and $\delta(m,N)/(K\,L\,T)$ as a function of $\overline{\gamma}_w$ which is the average SNR value of the UE experiencing the worst propagation conditions. In this scenario each UE is associated to an average SNR $\overline{\gamma}_u \in [0, 10]$ dB (for $u = 1, \ldots, M$). The figures show the performance metrics for two different values of the target delivery probability, namely $\hat{\Phi} = 0.8$ and $\hat{\Phi} = 0.9$.

As for the normalized overall transmission energy cost, Figure 3.1a clearly shows a maximum performance gain of at least three-fold for both the proposed CP-MNC and CR-MNC methods if compared to the optimized RLNC scheme. On the other hand, Figure 3.1b shows that the proposed MNC-based strategies outperform the optimized RLNC scheme in terms of the normalized delivery delay. In particular, it is worth noting that in this case the maximum normalized delivery delay which characterises the CP-MNC and CR-MNC methods is at least three-times and eight-times smaller than that associated to the optimized RLNC, respectively.

Even though the CR-MNC method outperforms the CP-MNC alternative in terms of the delivery delay, this result is achieved at the cost of an increase of the transmission power. However, as the overall power consumption at the BS side increases, the interference to other UEs or networks (which operates on the same frequency band)



(a) Normalized overall transmission energy cost



(b) Normalized overall delivery delay

Figure 3.1: Normalized overall transmission energy cost and delivery delay vs. average SNR of the UE experiencing the worst propagation conditions.

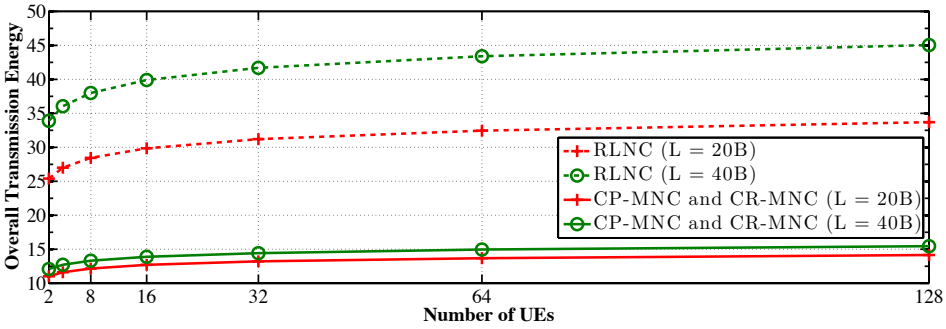


Figure 3.2: Normalized overall transmission energy cost vs. number of UEs.

augments. Of course, if the transmission power cost is minimized, also the impact of the interference on the overall system performance can be significantly reduced.

On the other hand, Figure 3.2 compares the performance of the considered strategies as a function of M. The figure refers to a network scenario where all the UEs are charactered by the same average SNR per symbol which is 1 dB. Figure 3.2 shows that the normalized overall transmission energy cost associated to the MNC-based strategies is up to three-times smaller than that associated to the optimized RLNC alternative.

Finally, the proposed MNC-based strategies minimise the overall transmission energy cost and significantly reduce the delivery delay in comparison to the RLNC alter-

native. In particular, a performance gain of at most three-fold in overall transmission energy cost and delivery delay is achieved by the CP-MNC method if compared to the optimized RLNC alternative. For what concerns the CR-MNC method, its maximum overall transmission energy cost and delivery delay are respectively at least three-times and eight-times smaller than those of the optimized RLNC.

3.2 General Allocation Model for Layered Video Broadcasting

In the rest of the section we focus on the video service delivery over 3GPP LTE-A networks. In particular, we address the challenge of optimizing the radio resource allocation process in eMBMS networks so that users, according to their propagation conditions, can receive video streams at the maximum achievable service level in a given cell.

Video content delivery over 4G mobile cellular networks, namely LTE and LTE-A, is estimated to grow exponentially due to the surge in demand for bandwidth-intensive applications based on video streaming [1].

This section deals with a SC-eMBMS deployment, where the eNB delivers broadcast video services to all UEs that belong to one cell. In particular, the main goal of the section is to define an efficient resource allocation strategy suitable for scalable video broadcasting. We consider video flows encoded by using the H.264/SVC [2] codec which provides video streams formed by multiple video layers, namely, the *base layer* and several *enhancement layers*. The base layer provides basic reconstruction quality which is gradually improved by decoding subsequent layers [3].

One of the key points of resource allocation strategies for PtM communications is the possibility of exploiting the user heterogeneity (in terms of propagation conditions) to maximize the level of satisfaction of each user. Multi-rate Transmission (MrT) schemes promise to overcome this problem by: (i) splitting the set of users targeted by the delivered PtM service into subsets, and (ii) differentiating the service delivery into subflows (one per subset) which are optimized according to the propagation conditions of each subset [74, 75]. Even though MrT schemes can better exploit the user heterogeneity, usually they assume that UEs provide feedback to the transmitting node reporting their propagation conditions [74, 11, 76] or positioning information [75]. In addition, these schemes do not address the resource allocation problem by taking into account the tight constraints imposed by 3GPP on the scheduling and structure of LTE radio frames containing eMBMS subframes (Chapter 13 [4]). Finally, it is worth

noting that, even though there are allocation strategies which aim to minimize the average/instantaneous user dissatisfaction [74, 11], none proposes a resource allocation framework ensuring a predefined service level to a certain group of users.

Reliable packet-loss resilient multimedia service broadcasting via eMBMS has been considered a challenging problem [77]. In particular, 3GPP has foreseen the adoption of AL-FEC scheme based on Raptor Codes to improve reliability of broadcast and multicast eMBMS communications [63]. However, a major concern about AL-FEC coding strategies is that they lead to a large communication delay [78]. In order to overcome that issue, link-level NC-based strategies have been recently proposed as a valuable and affordable (from the computational point of view) alternative to fountain code-based AL-FEC schemes [72, 78]. To this end, this section draws inspiration from [72] where authors propose to modify the standard LTE MAC by adding a coding sublayer, called MAC-Random linear NC (MAC-RNC), atop of that. In particular, it provides improved resilience to packet loss of delivered services by using RNC.

This section enhances the work presented in [72] by extending the MAC-RNC design to deliver H.264/SVC video streams as eMBMS broadcast traffic flows. In addition, the authors of [72] investigated the performance of the MAC-RNC-based delivery strategy by comparing it with 3GPP-standardized HARQ strategies. However, they did not try to optimize the system design under investigation. To this end, this section aims at jointly optimizing the MCS, the transmission rate and the NC scheme used to efficiently deliver each H.264/SVC video layer to heterogeneous set of user groups. We would like to highlight that unlike other published work ([74, 75, 11, 76]), the provided allocation strategy: (i) belongs to the MrT family, (ii) does not require any feedback from the UEs, and (iii) ensures that each service level is successfully received with a certain probability by the corresponding user group.

The section is organized as follows. Section 3.2.1 provides the necessary theoretical background on RNC and the H.264/SVC video compression standard. Section 3.2.2 describes the extension to the MAC layer we considered and the theoretical framework used to evaluate the service level of a H.264/SVC video service transmission. Moreover, Section 3.2.2 describes the proposed optimal resource allocation model. Numerical results are presented in Section 3.2.4.

3.2.1 H.264/SVC Broadcasting over LTE-A

Unlike traditional video compression standards, H.264/SVC (Annex G, [2]) defines procedures which encode a raw video bitstream into different video layers. In this section, by QoS we assume the received video quality in terms of the number of reconstructed

video layers. In particular, if a UE successfully receives the first l video layers, it actives a certain QoS profile; the perceived QoS level improves as l increases.

Modern 4G networks allow the eNB to broadcast (or multicast) data streams to target UEs at different MCSs, based on the users' channel conditions and their terminal capabilities. This means that users placed at different distances from the eNB would experience different QoS levels, and classical single-layer video encoding schemes may fail to deliver video services to users outside a specific coverage area in a given cell. For this reason, in this section we use the H.264/SVC that presents a suitable solution for video service delivery with distinct levels of video quality (e.g., time-, frequency-resolution, etc.), where users in worse channel conditions (i.e., close to the cell edge) would still be able to recover the video streams at the base layer quality, while UEs at a closer distance to the eNB can decode all the video layers (both base and enhancement layers).

In this section, we measure the video quality experienced at target users by means of the decoded video bitrate and the Peak Signal-to-Noise Ratio (PSNR) metrics. While the former indicates the maximum bitrate obtained at the output of the H.264/SVC decoder, the PSNR metric measures the quality of the decoded video at the application-layer [3].

3.2.2 System Model

Consider an H.264/SVC video stream which is delivered by an eNB as an SC-eMBMS flow to all the UEs in a cell. Moreover, assume that the service is composed by the layer set $\{v_0, v_1, \ldots, v_L\}$, where v_0 and $\{v_1, \ldots, v_L\}$ are base video layer and the $L-1$ enhancement layers, respectively.

Figure 3.3 shows the LTE protocol stack, proposed in [72], which we refer to. Assuming that each video layer is associated to an independent IP packet stream, the figure shows the stream composed of L video layers that enters the communication stack at the PDCP layer. The PDCP PDUs are concatenated/segmented in the RLC layer and then forwarded to the MAC layer [4]. Due to the fact that the MAC-RNC sublayer should improve the reliability of data broadcasting, we have that: (i) the stream of RLC PDUs related to a video layer is segmented into *information symbols* of L_S bits, (ii) information symbols are grouped into sets of K_l items, namely $\{p_1, \ldots, p_{K_l}\}$ the so-called *information messages* [67], and (iii) according to the RNC principle, the MAC-RNC sublayer produces a stream of *coded symbols* $\{c_1, c_2, \ldots\}$ from each information message. Finally, the i-th coded symbol is obtained as a linear combination of information symbols (forming an information message), i.e., $c_i = \sum_{j=1}^{K_l} g_j \cdot p_j$ where each

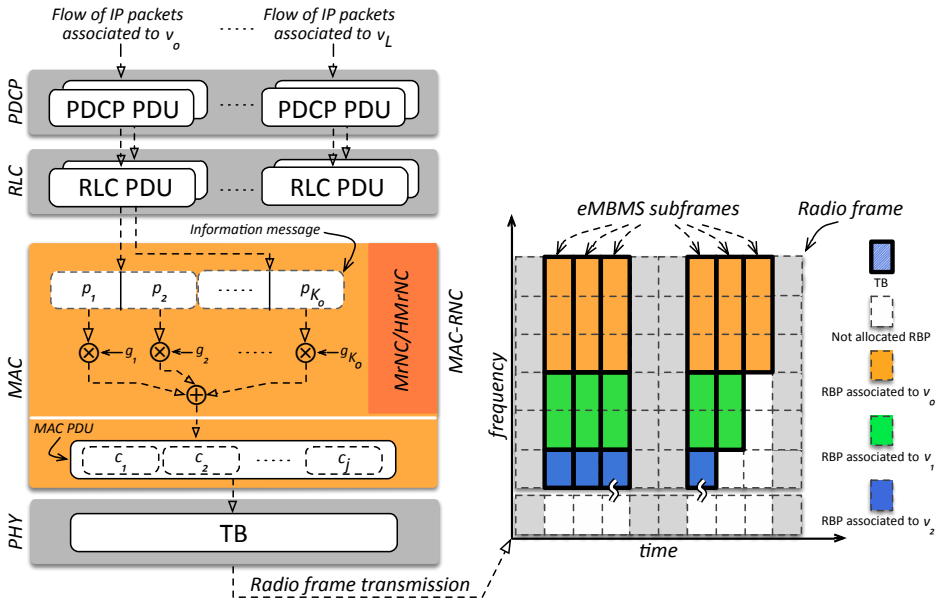


Figure 3.3: LTE/LTE-A protocol stack and (a part of) radio frame (for $L = 2$).

coding coefficient g_j is taken at random from an uniform distribution over a finite field of size q (see Section 2.2.1.1). A stream of coded symbols associated to one information message is mapped (by the standard LTE MAC layer) in N_l MAC PDUs. Finally, each MAC PDU is mapped on one TB and broadcast to the UEs. Hence, depending on the TB sizes and MCS in use[1], a TB holds a variable number of coded symbols. A UE recovers the delivered information message as soon as K_l linearly independent coded symbols are collected[2].

Table 3.1 lists the MCSs which are eligible for the TB transmission. In particular, in this section we considered the set of MCSs which corresponds to CQI values m_l that UEs feedback for Point-to-Point (PtP) services indicating their channel conditions [4]. Finally, we assume that all the TBs holding data associated to the l-th video layer are delivered by means of the same MCS m_l.

The transmission time duration of a TB is fixed and equal to Transmission Time Interval (TTI), namely 1 ms [4]. In addition, a TB may consist of $N_{\mathrm{RBP},l}$ RBPs[3]. Figure 3.3 shows the time-frequency structure of an LTE radio frame. It consists of 10 subframes (each of them has a transmission time duration of one TTI). The figure reports the (maximum) number of subframes that can be used for delivering eMBMS

[1]The LTE standard imposes that one MAC PDU has to be mapped in one TB. Hence, the MAC PDU size depends on the MCS used for the TB transmission. Furthermore, the standard MAC layer selects the MCS used in the TB transmission [4].

[2]For the sake of simplicity we assume that coding coefficients are known at the receiving UEs. However, it is worth noting that this assumption can be easily relaxed as suggested in [45].

[3]Which is a fixed frequency-time unit of resource allocation in LTE that consists of 12 OFDM subcarriers (180 kHz × 1 ms) [4].

Table 3.1: Number of coded symbols per RBP vs. m_l (for $L_S = 32$ bits) [72].

$\mathbf{m_l}$	**Mod.**	**Code Rate**	$\mathbf{n(m_l)}$	$\mathbf{m_l}$	**Mod.**	**Code Rate**	$\mathbf{n(m_l)}$
1-3	No Tx	-	-	10	64QAM	0.45	10
4	QPSK	0.3	2	11	64QAM	0.55	12
5	QPSK	0.44	3	12	64QAM	0.65	14
6	QPSK	0.59	4	13	64QAM	0.75	16
7	16QAM	0.37	5	14	64QAM	0.85	18
8	16QAM	0.48	6	15	64QAM	0.93	20
9	16QAM	0.6	8				

data flows, namely 6 out of 10 subframes per radio frame. Finally, as shown in the figure, we assume that during each eMBMS-capable subframe, the eNB can deliver (at most) one TB per video layer. Hence, a subframe holding eMBMS data, can deliver (at most) $L + 1$ TBs (namely, one base and L enhancement layers).

The TBs that hold coded symbols associated to the l-th video layer are delivered using the MCS m_l and contain $n(m_l)$ coded symbols per RBP. Hence, the total number of coded symbols that is to be placed in one TB is $C(m_l, N_{\mathrm{RBP},l}) = n(m_l) \cdot N_{\mathrm{RBP},l}$. Table 3.1 lists all the possible values of $n(\cdot)$, for $L_S = 32$ bits [72].

Moreover, define the l-th Multicast Area (MA) MA_l as the *fraction* of the cell area where all the UEs can recover the first $l + 1$ video layers with a certain probability. In this section we assume that the relation $m_0 \leq m_1 \leq \ldots \leq m_L$ holds, i.e., the MCS index of the l-th video layer cannot be smaller than that of the $(l-1)$-th one. Let us approximate MA_l (for $l = 0, \ldots, L$) as a circle of radius r_l equal to the maximum distance between the eNB and the farthest point where the TBLER Pe_{m_l} (characterizing the reception of TBs associated to the l-th video layer) is not greater than 10%[1]. For these reasons, we have that $r_l \leq r_{l-1}$. Assuming that UE distribution follows a Poisson point process of average density λ, the average (rounded up) number of UEs belonging to MA_l is given by $U_l = \lceil \lambda \, \pi \, r_l^2 \rceil$ [79]. Finally, it is useful to define the average number of UEs in the cell as $U_e = \lceil \lambda \, \pi \, r_e^2 \rceil$, where r_e is the maximum distance between the eNB and the cell-edge.

One H.264/SVC encoded video stream is divided into Group of Pictures (GoPs) that consist of g_{GoP} video frames. The video frame rate is given by f_{GoP} frames-per-second (fps), and the time duration of a GoP is $t_{\mathrm{GoP}} = g_{\mathrm{GoP}}/f_{\mathrm{GoP}}$. Moreover, we can express the time duration of a GoP in terms of the number of TTIs as: $d_{\mathrm{GoP}} = \left\lfloor t_{\mathrm{GoP}}/t_{\mathrm{TTI}} \right\rfloor$,

[1] In LTE/LTE-A systems, transmitting by using a given MCS is permitted as long as the TBLER experienced by a UE is equal to or smaller than 10^{-1} [4]. We assume that r_l can be estimated: (i) during the network deployment phase or (ii) by the eNB itself which uses CQI values reported by UEs for standard *Point-to-Point* services.

where a t_{TTI} is the LTE TTI (namely, 1 ms).

Since the decoding process of a H.264/SVC video takes place on a per-GoP basis, we define an RNC information message of the l-th SVC video layer as the set of information symbols forming the l-th layer of one GoP. Hence, K_l is defined as $K_l = \left\lceil (R_l \cdot t_{\mathrm{GoP}})/L_S \right\rceil$, where R_l is the bitrate of the l-th SVC video layer[1].

In this section, the term QoS refers to the received video quality in terms of the number of reconstructed video layers. For an information message of the l-th video layer, the probability that a UE recovers it (i.e., the probability that a UE collects at least K_l linearly independent coded symbols) after N_l TB transmissions as a function of N_l, m_l and $N_{\mathrm{RBP},l}$ can be expressed as follows (as presented in (3.4) and (3.5)):

$$\begin{aligned} P_{\mathrm{UE},l} &\doteq P_{\mathrm{UE}}(N_l, m_l, N_{\mathrm{RBP},l}) \\ &= \sum_{t=\chi_l}^{N_l} \left[\binom{N_l}{t} \mathrm{Pe}_{m_l}^{N_l - t} \left[1 - \mathrm{Pe}_{m_l}\right]^t \prod_{i=0}^{K_l - 1} \left(1 - \frac{1}{q^{t\, C(m_l, N_{\mathrm{RBP},l}) - i}}\right) \right] \end{aligned} \tag{3.20}$$

where $\chi_l = \left\lceil K_l / C(m_l, N_{\mathrm{RBP},l}) \right\rceil$ is the minimum (integer) number of TB transmissions needed to deliver at least K_l coded symbols. Let us assume that TB reception errors occur as statistically independent events among UEs of the same MA. From (3.20), the probability that U_l UEs recover the l-th SVC video layer of a GoP is $P_{\mathrm{UE},l}^{U_l}$. Hence, as presented in (3.6), the probability that U_l UEs belonging to MA_l recover the basic and the first l enhancement video layers is (at least) equal to

$$P_{\mathrm{MA},l} \doteq P_{\mathrm{MA}}\big(N_0, \ldots, N_l, m_0, \ldots, m_l, N_{\mathrm{RBP},0}, \ldots, N_{\mathrm{RBP},l}\big) = \prod_{i=0}^{l} P_{\mathrm{UE},i}^{U_i}. \tag{3.21}$$

3.2.3 Rate-Optimized and Coverage-Aware Model

The novel resource allocation strategy we propose, which we call "Multi-rate Network Coding" (MrNC), is embedded into the MAC-RNC sublayer (see Figure 3.3) implemented at the eNB side and does not rely on any information related to UEs in the given cell. The proposed strategy aims at allocating resources in order to ensure that heterogenous QoS levels are achieved for different MAs. That goal is achieved, for each video layer, by jointly optimizing (i) TB sizes (in terms of number of RBP per TB) $N_{\mathrm{RBP},l}$ (ii) the number of TB transmissions N_l, and (iii) selecting the MCS m_l of each MA. Hence, the proposed strategy aims at optimizing the number of transmitted coded

[1]It is worth mentioning that if the value of K_l is too large for the Gaussian Elimination decoder in use, the complexity of the decoding process can be reduced by referring to the systematic version of RNC [72].

symbols per video layer. Finally, the MrNC model can be stated as follows:

$$\text{(MrNC)} \quad \min_{\substack{m_0,\dots m_L \\ N_0,\dots,N_L \\ N_{\text{RBP},0},\dots,N_{\text{RBP},L}}} \quad \sum_{l=0}^{L} N_l \; N_{\text{RBP},l} \tag{3.22}$$

$$\text{subject to} \quad \frac{U_l}{U_e} \geq U_{\text{TH},l} \quad l = 0,\dots,L \tag{3.23}$$

$$m_l < m_{l+1} \quad l = 0,\dots,L-1 \tag{3.24}$$

$$P_{\text{MA},l} \geq \hat{P}_{\text{TH},l} \quad l = 0,\dots,L \tag{3.25}$$

$$N_{\text{RBP},l} \leq \hat{N}_{\text{TH}} \quad l = 0,\dots,L \tag{3.26}$$

$$N_l \leq \left\lfloor \text{TTI}_e \, d_{\text{GoP}} \right\rfloor \quad l = 0,\dots,L \tag{3.27}$$

where the constraint (3.23) ensures that the average number of UEs per MA is not smaller than a certain value, and (3.24) avoids the overlapping of any two MAs, since it would be pointless to deliver the same video service characterized by two different QoS levels across the same fraction of the cell area. Using the constraint (3.25), the $v_0,\dots,v_l$ video layers will be recovered with a probability which is at least equal to $\hat{P}_{\text{TH},l}$. It is worth mentioning that the value of $P_{\text{MA},l}$ in (3.25) has been evaluated by setting $\text{Pe}_{m_l} = 10^{-1}$ (i.e., we set Pe_{m_l} to the greatest TBLER value) in (3.20) and (3.21). As for (3.26), it ensures that the frequency span of each TB can not be greater than $\hat{N}_{\text{TH}}$. TB transmissions associated to each video layer have to be completed (at most) in d_{GoP} subframes. Due to the fact that only 60% ($\text{TTI}_e = 0.6$) of subframes per-frame are eMBMS-capable, the constraint (3.27) states that N_l cannot be greater than $0.6 \cdot d_{\text{GoP}}$. The objective function (3.22) will minimize the overall radio resource footprint associated with the transmission of all the video layers of a GoP conditioned that the QoS constraints as defined in (3.23)-(3.27) are met. In particular, (3.22) minimizes the overall number of RBPs ($N_l \cdot N_{\text{RBP},l}$) associated with each video layer.

Unfortunately, the MrNC model is a nonlinear integer optimization problem which is hard to solve in closed form. Hence, the rest of this section addresses these issues by proposing an heuristic strategy to solve the MrNC model: the Heuristic MrNC (HMrNC) strategy. HMrNC comprises three sequential steps which aim to: (i) optimize the MCSs of each MA, (ii) choose the TB sizes, and (iii) optimize the number of TB transmissions.

Considering Procedure 3.1, it is in charge of the first step, namely: (i) it iterates over the MCS values (starting from 15, see Table 3.1), and (ii) for each video layer, it identifies the smallest MA such that the constraints (3.23) and (3.24) hold. For the second step of HMrNC, we decided to set the values of $N_{\text{RBP},l}$ equal to the maximum value

Procedure 3.1 Definition of the MAs.

```
t ← 15
for l = 0 → L do
    repeat
        m_l ← t
        t ← t − 1
    until U_l/U_e ≥ U_TH,l or t < 4
end for
```

($\hat{N}_{\mathrm{TH}}$) and then optimize the number of TBs transmitted to each MA[1] (the third step). Let us define[2] $\tilde{P}_{\mathrm{MA},l}(N_0,\ldots,N_l) \doteq P_{\mathrm{MA}}\big(N_0,\ldots,N_l\big|m_0,\ldots,m_l,N_{\mathrm{RBP},0},\ldots,N_{\mathrm{RBP},l}\big)$. Since $m_0,\ldots,m_L$ and $N_{\mathrm{RBP},0},\ \ldots,N_{\mathrm{RBP},L}$ are given, the MrNC problem can be restated as follows

$$\text{(H1)} \qquad \min_{N_0,\ldots N_L} \sum_{l=0}^{L} N_l \tag{3.28}$$

$$\text{subject to} \qquad N_l \leq \left\lfloor \mathrm{TTI}_e\, d_{\mathrm{GoP}} \right\rfloor \quad l=0,\ldots,L \tag{3.29}$$

$$\tilde{P}_{\mathrm{MA},l}(N_0,\ldots,N_l) \geq \hat{P}_{\mathrm{TH},l} \quad l=0,\ldots,L. \tag{3.30}$$

Once again, H1 is a noninteger and nonlinear problem but in this case it can be efficiently solved. To this end, considering $\tilde{P}_{\mathrm{MA},l}(N_0,\ldots,N_l)$, from (3.21) we can see that the probability value cannot decrease when N_l increases and the remaining variables are kept constant. Furthermore, let N_l^* (for $l=0,\ldots,L$) be the smallest value of N_l such that $\tilde{P}_{\mathrm{MA},l}(N_0,\ldots,N_l) \geq \hat{P}_{\mathrm{TH},l}$ (for $l=0,\ldots,L$) holds. Likewise, the approach presented in [80] and starting from $l=0$, the value of N_l^* can be efficiently found by testing all the possible values of N_l from χ_l until $\tilde{P}_{\mathrm{MA},l}(N_0,\ldots,N_l) \geq \hat{P}_{\mathrm{TH},l}$ holds. In particular, Proposition 3.1 proves that the objective function (3.28) is minimized by $\left\{N_0^*,\ldots,N_L^*\right\}$. Finally, Procedure 3.2 proposes a possible implementation of the proposed strategy.

Proposition 3.1. Considering $\left\{N_0^*,\ldots,N_L^*\right\}$, it is an optimum solution of H1.

Proof. The probability $\tilde{P}_{\mathrm{MA},l}(N_0,\ldots,N_l)$ is a nondecreasing function with respect to the variable N_l (for any $l=0,\ldots,L$). Considering Procedure 3.2, it starts from $l=0$ and minimizes the functions $\tilde{P}_{\mathrm{MA},l}(N_0),\ldots,\tilde{P}_{\mathrm{MA},l}(N_0,\ldots,N_l)\big|_{N_0=N_0^*,\ldots,N_{l-1}=N_{l-1}^*}$, etc. Let us assume the existence of another solution $\left\{N_0',\ldots,N_L'\right\}$ of H1 such that

[1] This method will tend to reduce the transmission time duration of a GoP rather than optimize the TB sizes. In addition, the latter aspect can be indirectly addressed during the service deployment phase by choosing the value of $\hat{N}_{\mathrm{TH}}$.

[2] In this section we define $f(x|t_0,\ldots,t_w)$ as the parametric function where x is the variable and $t_0,\ldots,t_w$ are parameters.

Procedure 3.2 Minimization of the time duration of the process.

for $l = 0 \rightarrow L$ **do**
  $N_l^* \leftarrow \chi_l$
  while $\tilde{P}_{\mathrm{MA},l}(N_0, \ldots, N_l)\big|_{N_0=N_0^*,\ldots,N_{l-1}=N_{l-1}^*} < \hat{P}_{\mathrm{TH},l}$ **do**
    $N_l^* \leftarrow N_l^* + 1$
  end while
end for

$\sum_{l=0}^{L} N_l' < \sum_{l=0}^{L} N_l^*$. Hence, there is at least one term N_l' such that $N_l' < N_l^*$. However, because of the definition of N_l^*, the constraint (3.30) would not hold. This completes the proof by reductio ad absurdum. □

As for Procedure 3.2, it can solve H1 in a finite number of steps. In particular, we can note that N_l^* belongs to the interval $I = \left[\chi_l, \left\lfloor \mathrm{TTI}_e \cdot d_{\mathrm{GoP}} \right\rfloor\right]$. During one iteration, the procedure tests just one value of I. Hence, N_l^* is found in a number of iterations such that are less than or equal to the number of items in I. For this reason, Procedure 3.2 returns (at most) after Q iterations such that:

$$Q \leq \sum_{l=0}^{L} \left(\left\lfloor \mathrm{TTI}_e \, d_{\mathrm{GoP}} \right\rfloor - \chi_l + 1 \right) = \left(1 + \left\lfloor \mathrm{TTI}_e \, d_{\mathrm{GoP}} \right\rfloor \right) \left(L + 1 \right) - \sum_{l=0}^{L} \chi_l. \quad (3.31)$$

3.2.4 Numerical Results

This section investigates the system performance in terms of the resource load index η defined as:

$$\eta = \frac{1}{\left\lfloor \mathrm{TTI}_e \, d_{\mathrm{GoP}} \right\rfloor \hat{N}_{\mathrm{TH}}} \sum_{l=0}^{L} N_{\mathrm{RBP},l} \, N_l \quad (3.32)$$

where $\sum_{l=0}^{L} N_{\mathrm{RBP},l} \cdot N_l$ represents the *radio resource footprint* of the allocation strategy. In addition, we consider the probabilities[1] $P_{\mathrm{MA},l}$ that a reference group of 10 UEs can recover each service QoS level (see (3.21)) and hence the maximum achievable PSNR defined as:

$$\overline{p} = \max_{l=0,\ldots,L} \left\{ \hat{p}_l \, P_{\mathrm{MA},l} \right\} \quad (3.33)$$

where $\hat{p}_l$ is the PSNR obtained after recovery of the video layers $v_0, \ldots, v_l$.

We provide performance comparisons between the proposed allocation strategies based on the MrNC model and HMrNC heuristic approach. We also consider the allocation model proposed in [11] named hereafter as the *Video Rate Allocation* (VRA)

[1] In this this sections we referred to Pe_{m_l} (for $l = 0, \ldots, L$) values computed by averaging TBLER values obtained by 10^3 iterations of the datalink simulation framework presented in [72].

Table 3.2: Simulation parameters considered.

Paramter		**Value**
Inter-Site-Distance (ISD)		500 m
System Bandwidth		20 MHz
Transmission Scheme		SISO
Duplexing Mode		FDD
Carrier Frequency		2.0 GHz
Transmission Power		40 W per-sector
eNB and UE Antenna Gains		see Table A.2.1.1-2 [81]
Pathloss and Penetration Loss		see Table A.2.1.1.5-1 [81]
$\hat{P}_{\mathrm{TH},l}$		0.9, for $l \in 0, \ldots, L$
$\hat{N}_{\mathrm{TH}}$		$[4, \ldots, 12]$ RBPs
Stream A [11]	$\{\hat{r}_0, \ldots, \hat{r}_2\}$ [kbps]	$\{117.1, 402.5, 1506.3\}$
	$\{\hat{p}_0, \ldots, \hat{p}_2\}$ [dB]	$\{29.94, 34.78, 40.73\}$
	$\{U_{\mathrm{TH},0}, \ldots, U_{\mathrm{TH},2}\}$	$\{0.4, 0.5, 0.9\}$
Stream B	$\{\hat{r}_0, \ldots, \hat{r}_3\}$ [kbps]	$\{160.0, 300.0, 560.0, 1150.0\}$
	$\{\hat{p}_0, \ldots, \hat{p}_3\}$ [dB]	$\{29.45, 32.30, 34.52, 38.41\}$
	$\{U_{\mathrm{TH},0}, \ldots, U_{\mathrm{TH},3}\}$	$\{0.4, 0.55, 0.75, 0.9\}$
Stream A, B	g_{GoP}	16 frames
	f_{GoP}	30 fps
	q	2^8

strategy which tries to maximize the sum of the video quality perceived by each UE. In order to make a fair comparison among the MrNC, HMrNC and VRA methods, we impose that the eNB cannot skip the transmission of any video layer, hence, we restate the VRA objective function as follows[1]:

$$\max_{m_0,\ldots,m_L} \sum_{l=0}^{L} U_l \, \hat{p}_l. \tag{3.34}$$

Furthermore, we compare both MrNC and HMrNC with a MrT-based strategy [6]. For the latter strategy, we draw inspiration from [76], where UEs are split into multiple MGs, the transmission rate used to deliver data to one MG is constrained to the UE experiencing the worst propagation conditions (in the MG). This means that the MrT optimization problem can be restated in the following equivalent form:

$$\arg\min_{m_l \in [4,\ldots,15]} \left\{ m_l \middle| \frac{U_l}{U_e} \geq U_{\mathrm{TH},l} \right\} \quad \text{for any } l = 0, \ldots, L. \tag{3.35}$$

It tries to deliver the l-th video layer over a MA_l by using the minimum MCS such

[1]The original formulation of the VRA model aims at jointly optimizing the set of delivered layers and MCSs used in the transmissions [11].

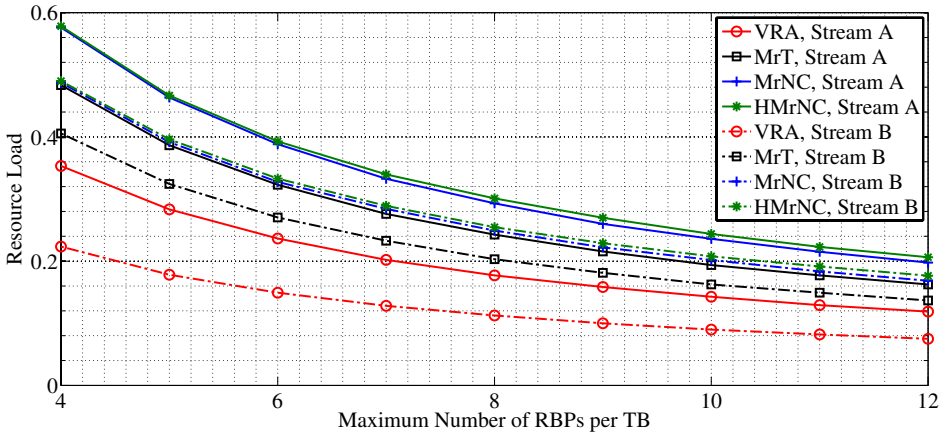


Figure 3.4: Resource load index as a function of $\hat{N}_{\mathrm{TH}}$.

that the relation (3.23) holds (namely, a LCG-based approach is used within a MG).

Due to the fact that neither the VRA, nor MrT strategies explicitly address the TB sizing problem, we assume that each TB consists of $\hat{N}_{\mathrm{TH}}$ RBPs. In addition, both the VRA and LGC strategies assume that UEs can report to the eNB CQI feedback but one of the key points of MrNC and HMrNC methods is that they do not rely on any UE feedback. Hence, for the sake of comparison, we assume that the actual number of UEs which on average report the CQI value m_l is equal to U_l. Finally, in the case of both the VRA- and MrT-based video delivery, transmissions take place through the standard LTE MAC layer (namely, a communication stack without the MAC-RNC sublayer).

We consider a network of 19 macro-cell eNBs, each managing three hexagonal sectors. eNBs are organized in two concentric circles centred on the target eNB. In addition, TBLER values experienced by a UE, as a function of a given MCS and distance from the eNB, are estimated by the finite-state Markov model approach presented in [82] and extended in [72]. Finally, Table 3.2 summarizes both the main simulation parameters and the couple of H.264/SVC *Foreman* video streams [83] which we considered.

Results reported in this section will clearly show that the proposed MrNC and HMrNC strategies provide resource allocation solutions which meet predefined service constraints (see (3.23) and (3.25)) with the minimum resource footprint (see (3.22)). Furthermore, in spite of the fact that the radio resource footprint of VRA and MrT (required to achieve their respective goals) is smaller than those associated to the MrNC and HMrNC strategies, they cannot ensure that a predefined video QoS levels is maintained over the targeted fractions of the cell area.

In Figure 3.4, we compare the value of η, as a function of $\hat{N}_{\mathrm{TH}}$, characterizing all the considered resource allocation strategies, in the case of video stream A and B. From (3.32), we have that the overall number of RBPs used to deliver one stream

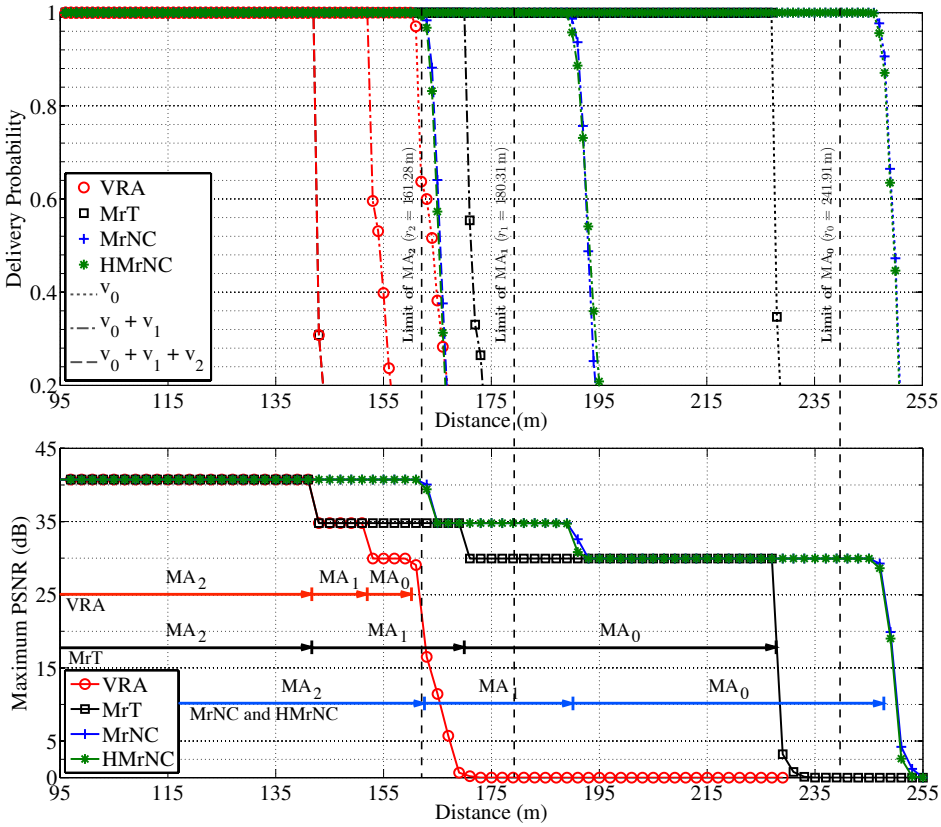


Figure 3.5: Video delivery probabilities and maximum PSNR of the stream A.

increases as the value of η rises. Considering the figure, the performance gap between the MrNC and HMrNC models is negligible (at most it is less than 0.01). It is worth noting that it is caused by the fact that, in the latter case, both the MCS selection and TB sizing processes are separate from the optimization of the number of TB transmissions. On the other hand, both the VRA and MrT strategies deliver the video stream A (B) by using a proportion of bandwidth resources which is smaller than that associated to the MrNC strategy (and the HMrNC as well) of at most 1.63 and 1.19 (2.18 and 1.20) times, respectively.

In spite of the larger radio resource footprint for MrNC and HMrNC, it is worth noting that the proposed models *can deliver a service with the desired QoS level over a given fraction of the cell-area*. Considering the stream A, Figure 3.5 compares (for $\hat{N}_{\mathrm{TH}} = 6$) the $P_{\mathrm{MA},l}$ values of each QoS level and $\overline{p}$ as a function of the distance of the considered reference group from the eNB. For each MA, the figures report the value of r_l (the dashed lines). Unlike MrNC and HMrNC, both the VRA and MrT strategies cannot deliver the service over the desired fractions of the cell-area. For instance, MA_0 (MA_2) defined by the VRA and MrT strategies extends up to a distance which is 81.9m and 14.9m (20.2m for both the strategies) smaller than the (minimum) required one,

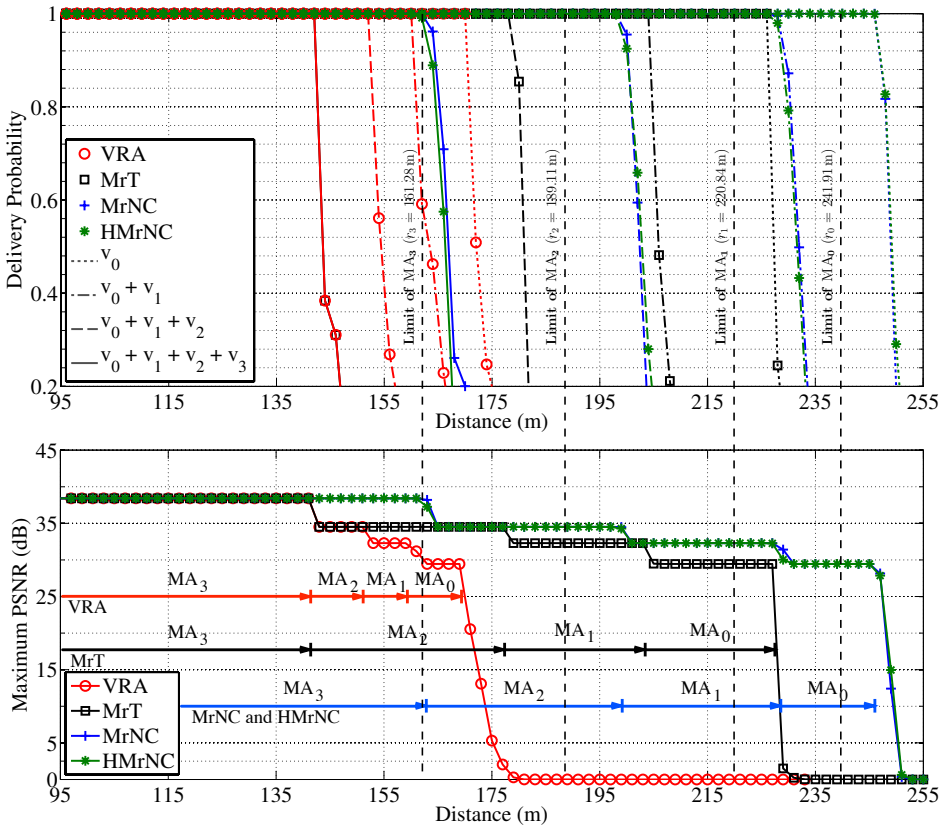


Figure 3.6: Video delivery probabilities and maximum PSNR of the stream B.

respectively. Finally, Figure 3.6 shows (for $\hat{N}_{\mathrm{TH}} = 6$) similar behaviour for video stream B. In this case we note that MA_0 (MA_3) provided by the VRA and MrT allocation strategies spans up to a distance that negatively diverges from the (minimum) required one of 71.9m and 15.9m (20.2m for both the strategies), respectively.

Considering Figures 3.5 and 3.6, the HMrNC allocation model can deliver one video stream, at a certain QoS level, over a MA which can be slightly greater than that defined by the MrNC strategy. In particular, this effect can be noted in Figure 3.5 (Figure 3.6) by considering the delivery probability values associated to the reception of v_0 and v_1 (v_0, v_1 and v_2). That is not surprising because the resource allocation solution derived by the HMrNC model is: (i) a feasible solution of MrNC problem but (ii) suboptimal in terms of the number of RBPs used to deliver all the video layers. As a consequence, the HMrNC model provides a resource allocation that leads to deliver more coded symbols per-video layer than the MrNC alternative. Hence, the $P_{\mathrm{MA},l}$ (see (3.21)) values can be slightly greater than those associated to the MrNC model.

Finally, we propose an optimal (the MrNC model) and heuristic resource allocation strategy (the HMrNC procedure) suitable for SC-eMBMS broadcast communications delivered through the MAC-RNC sublayer. We demonstrated that HMrNC can effi-

ciently derive feasible solutions of MrNC with a reduced computational load. Unlike VRA and MrT strategies, both MrNC and HMrNC ensure the desired service coverage. In particular, the VRA and MrT strategies can deliver the considered video streams at the maximum (minimum) QoS level over MAs which, at least, are 22% (50% and 12%, respectively) smaller than the desired ones.

Chapter 4

Unacknowledged Power Allocation Strategies

In this chapter, we propose an energy efficient resource allocation model for delay sensitive service delivery by means of the RLNC approach. The proposed optimization model aims at minimizing the overall transmission energy by jointly optimizing the transmission power and the RLNC scheme. We considered a scenario where several multicast services are delivered to a set of users. The service delivery is constrained both in terms of delivery delay and recovery probability. We develop an efficient heuristic strategy which provides, in a finite number of steps, a good quality feasible resource allocation such that service constraints are met and the overall transmission energy is significantly reduced.

4.1 Power Allocation for Heterogeneous Broadcasting

This section deals with a communication system where a BS delivers multiple downlink PtM data flows to a group of UEs belonging to the same MG. Each flow has to be recovered by the MG with a certain probability and within a given time interval. Namely, a data flow has to be successfully recovered by all the UEs of the MG with a certain probability and within a certain time. Several error control strategies for PtM communications have been proposed [22]. However, all the proposals are not capable of controlling the QoS level on a per-MG basis (in terms of recovery probability and delivery delay), as one PtM service should be delivered providing to *all* the UEs of the MG the desired QoS level. Because of the complexity of managing multicast/broadcast services through a retransmission process, ARQ or HARQ schemes can be efficiently

replaced with a NC-based principle [22].

In addition to PtM services reliability issues, there is another factor of paramount importance for both the network provider and environment, namely the energy footprint of service delivery. In fact, modern wireless communication networks are responsible for more than the 0.2% of the total carbon emissions [61]. In spite of the huge amount of resource allocation strategies aiming at minimizing the transmission power [14], a little attention has been paid to reduce the transmission energy of PtM communications.

This section proposes an efficient resource allocation model, which aims at jointly optimizing both the BS transmission power and the NC scheme used to deliver each PtM flow. In this way, the proposed optimization model can minimize the overall transmission energy. Finally, we propose an efficient resource allocation procedure which can efficiently find (in a finite number of steps) a good quality feasible solutions of the proposed optimization model.

The rest of the chapter is organized as follows. Section 4.2 describes the considered system model and provides the necessary background. The proposed optimization strategy and heuristic procedure are presented in Section 4.3. Section 4.4 inspects the performance of the proposed allocation model.

4.2 System Model

Let us assume that the BS delivers S different multicast services to M UEs forming a MG. In addition, let us consider that all the services are delivered at the same time through statistically independent communication channels. To this end, without loss of generality, we assume that: (i) communications are organized in frames, and (ii) the delivered services are multiplexed by means of the OFDMA scheme.

Considering the s-th service, it can be modelled as a stream of information messages. Each message $\mathbf{E}_s$ consists of K_s information elements (for instance, generic network Packet Data Units). For each delivered service, the BS linearly combines all the information elements belonging to the same information message to produce a stream of $N_s \geq K_s$ (for $s \in \{1, \ldots, S\}$) coded elements (see Section 2.2.1.1). The BS transmits, at the same time, to the MG S streams of coded elements (one stream per-delivered service). We assume that a coded element is always L bits long, regardless of the service. Finally, it is worth noting that each service is delivered according to the RLNC principle (see Section 2.2.1.1).

We assume a framed communication system where a frame is a frequency $\times$ time structure of fixed size elements (see Figure 4.1). One frame element spans one *subchannel* and *time slot* in the frequency and time domain, respectively. A frame location can

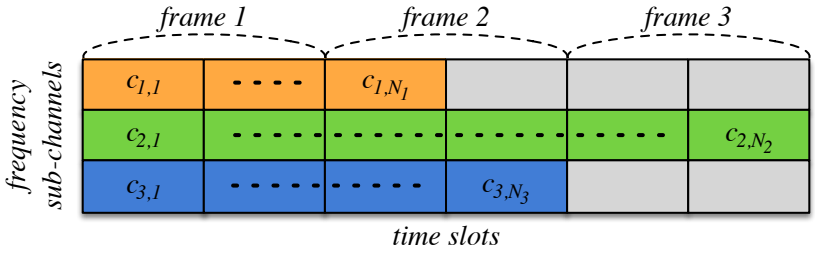


Figure 4.1: Example of the considered framed communication delivery for $S = 3$.

hold just one coded element and a service is conveyed to the MG by frame elements belonging to the same subchannel.

Let P_s be the power associated to the transmission of one coded element belonging to the s-th service. We assume that the transmission power of the BS during a time slot cannot be greater than the overall power budget $\hat{P}$, i.e., $\sum_{s=1}^{S} P_s \leq \hat{P}$. In addition, let us define the term $\overline{P} \doteq \hat{P}/S$. The value of P_s can be expressed as $P_s = m_s \overline{P}$, where the term m_s belongs to the set[1] $\mathbb{R}^+$. Assuming that the impact of the interferences on the considered system is negligible, it is straightforward to note that the following relation holds

$$\sum_{t=1}^{S} P_s \leq \hat{P} \;\Leftrightarrow\; \sum_{s=1}^{S} m_s \overline{P} \leq S\,\overline{P} \;\Leftrightarrow\; \sum_{s=1}^{S} m_s \leq S. \tag{4.1}$$

Assuming that communications adopt the BPSK[2], the transmission energy of one coded element is $m_s \cdot E \cdot L$, where E is the energy associated to the transmission of one bit. In the rest of the section, we refer to the the energy associated to the transmission of N_s coded elements, normalized to $L \cdot E$, namely $m_s \cdot N_s$. Finally, the overall transmission energy (normalized to $L \cdot E$) needed to deliver the coded elements associated to one information message of service s is $\sum_{s=1}^{S} m_s\, N_s$.

Let γ_u and $\overline{\gamma}_u$ be the instantaneous and the average SNR associated to a coded element reception by the u-th UE of the MG, respectively. Considering the s-th service, it is straightforward to note that, relations $\gamma_u = m_s\, \gamma_{o,u}$ and $\overline{\gamma}_u = m_s\, \overline{\gamma}_{o,u}$ hold, where $\gamma_{o,u}$ and $\overline{\gamma}_{o,u}$ are the instantaneous and the average SNR (associated to the reception of a coded element) experienced by the u-th UE for $m_s = 1$, respectively. Assuming that communications occurs over a Rayleigh communication channel, the element error probability $P_u(m_s)$ can be expressed as reported in (3.3).

[1] In this section we refer with $\mathbb{R}^+$ and $\mathbb{N}$ to the set of real positive and natural numbers, respectively.

[2] The theoretical derivation we propose is quite general and can be extended to other modulation schemes.

4.3 Optimization Model

Let us consider the delivered set of services. The resource allocation we propose aims at optimizing each couple (m_s, N_s) such that: (i) the overall transmission energy of the BS is minimized, and (ii) each information message can be recovered within a certain time by any UE of the MG with a probability which is not smaller than $\overline{\Phi}$.

Before going into details of the proposed optimization strategy, it is worth defining the probability that one UE recovers an information message (belonging to the s-th service), as a function of m_s and N_s (as presented in (3.4) and (3.5)):

$$F_u(m_s, N_s) = \sum_{j=K_s}^{N_s} \binom{N_s}{j} P_u^{N_s-j}(m_s)\Big[1 - P_u(m_s)\Big]^j g(j) \tag{4.2}$$

where $g(j) = \prod_{h=0}^{K_s-1}\left[1 - \frac{1}{q^{j-h}}\right]$ is the probability that at least K_s over j (for $j \geq K_s$) coded elements are linearly independent. As presented in (3.6), all the UEs of the MG can recover an information message with a probability which is

$$\Phi(m_s, N_s) = \prod_{u=1}^{M} F_u(m_s, N_s). \tag{4.3}$$

As a result, the proposed green resource allocation model can be expressed as follows:

$$\text{(M1)} \quad \min_{\substack{m_1,\ldots,m_S \\ N_1,\ldots,N_S}} \sum_{s=1}^{S} m_s N_s \tag{4.4}$$

$$\text{subject to} \quad \Phi(m_s, N_s) \geq \hat{\Phi}, \qquad s \in \{1, \ldots, S\} \tag{4.5}$$

$$\sum_{s=1}^{S} m_s \leq S \tag{4.6}$$

$$K_s \leq N_s \leq \hat{N}_s, \qquad s \in \{1, \ldots, S\} \tag{4.7}$$

$$m_s \in \mathbb{R}^+, \; N_s \in \mathbb{N}, \; s \in \{1, \ldots, S\} \tag{4.8}$$

where the constraint (4.5) ensures that the MG recovers services with a probability which is not smaller than $\hat{\Phi}$, we assume that $\hat{\Phi}$ is constant for all the transmitted services. From (4.1), the constraint (4.6) ensures that the instantaneous transmission power of the BS is not greater than $\hat{P}$. The constraint (4.7) imposes an upper-bound to the coded element transmissions for each service (i.e., the value of N_s cannot be greater than $\hat{N}_s$). Hence, the constraint (4.7) upper-bounds the transmission time duration of each service. Finally, it is straightforward to note that M1 is a mixed integer non-linear

optimization problem. Hence, it is hard to efficiently solve it. To this end, we show that Procedure 4.1 can find a feasible solution to M1 in a finite number of steps.

Considering M1, let us relax the constraint (4.6). In the rest of the section, the resulting optimization model is called Unbounded Transmission Power (UTP) model. It is worth noting that the UTP allocation model is introduced for the sake of analysis. It represents an "unsafe" allocation model due to the fact that it aims at minimizing the overall transmission energy but it could provide a resource allocation solution which requires an overall transmission power that exceeds $\hat{P}$.

The UTP model is equivalent to a set of S independent problems, where the s-th problem can be expressed as follows

$$\text{(M2)} \quad \min_{m_s, N_s} \; m_s \, N_s \tag{4.9}$$

$$\text{subject to} \quad \Phi(m_s, N_s) \geq \hat{\Phi} \tag{4.10}$$

$$K_s \leq N_s \leq \hat{N}_s \tag{4.11}$$

$$m_s \in \mathbb{R}^+, \; N_s \in \mathbb{N}. \tag{4.12}$$

It has been already shown that the solution of M2 can be efficiently found as follows [80]:

(i) For any value of N_s (where $K_s \leq N_s \leq \hat{N}_s$) find that value of m_s such that $\Phi(m_s, N_s)$ is equal to $\hat{\Phi}$.

(ii) Choose the (m_s, N_s) pair, among the computed ones, which minimises the objective function (4.9).

In addition, it can be proved that [80]: (i) the solution of M2 can be efficiently derived in a finite number of steps and belongs to the set $\mathcal{L}_s \doteq \left\{ (m_s, N_s) \in \mathbb{R}^+ \times \mathbb{N} \;\middle|\; K_s \leq N_s \leq \hat{N}_s \;\wedge\; \Phi(m_s, N_s) = \hat{\Phi} \right\}$ which is the frontier of the feasible set of M2,[1] and (ii) $\Phi(m_s, N_s)$ is a monotonically increasing function both with respect to m_s and N_s. In particular, if the relation $\Phi(m, n) \geq \hat{\Phi}$ holds, then $\Phi(m', n+1) \geq \hat{\Phi}$ holds as well, where $m' \leq m$.

Let us consider also a special case of UTP, where the transmission power P_s is fixed to $\overline{P}$ (for any $s \in \{1, \ldots, S\}$), i.e., $m_s = 1$ for each PtM service. In the rest of the section, the aforementioned model is called Constant Transmission Power (CTP) model

[1] For the sake of simplicity and with a little of notation abuse, in this section $\mathcal{L}_s(N_s)$ represents the value of m_s such that $(m_s, N_s) \in \mathcal{L}_s$.

and can be expressed as follows:

$$\text{(CTP)} \quad \min_{N_1,\ldots,N_S} \sum_{s=1}^{S} N_s \tag{4.13}$$

$$\text{subject to} \quad \Phi(1, N_s) \geq \hat{\Phi}, \qquad s \in \{1, \ldots, S\} \tag{4.14}$$

$$K_s \leq N_s \leq \hat{N}_s, \quad s \in \{1, \ldots, S\} \tag{4.15}$$

$$N_s \in \mathbb{N}, \qquad s \in \{1, \ldots, S\}. \tag{4.16}$$

It is worth proving the proposition below.

Proposition 4.1. Let (m_s^*, N_s^*), (m_s', N_s') and (m_s'', N_s''), for any $s \in \{1, \ldots, S\}$, be the optimum solutions of M1, UTP and CTP models, respectively. The following relation holds

$$\sum_{s=1}^{S} m_s' \, N_s' \leq \sum_{s=1}^{S} m_s^* \, N_s^* \leq \sum_{s=1}^{S} m_s'' \, N_s''. \tag{4.17}$$

Proof. The solution of CTP meets the constraints of M1 (i.e., any solution to CTP is, at least, a suboptimal solution of M1). In addition, the M1 model represents a spacial case of the UTP model. Hence, the proof follows from the fact that $\mathcal{M}'' \subseteq \mathcal{M}^* \subseteq \bigcup_{s=1}^{S} \mathcal{M}_s'$, where $\mathcal{M}^*$, $\mathcal{M}_s'$ and $\mathcal{M}''$ are the feasible sets of models M1, UTP and CTP, respectively. □

For these reasons, if $\sum_{s=1}^{S} m_s' \leq S$ then the optimum solution of M1 can be found by solving M2 for any service $s \in \{1, \ldots, S\}$. On the other hand, Proposition 4.1 states that the optimal solution of M1 cannot be worse than that of CTP. In the case of $\sum_{s=1}^{S} m_s' > S$, it is still possible to find a good quality suboptimal solution of M1. To this end, let (m_s^{**}, N_s^{**}) (for $s \in \{1, \ldots, S\}$) be the solution which is returned by Procedure 4.1. In brief, the procedure comprises the following steps:

(i) $(m_1^{**}, N_1^{**}), \ldots, (m_S^{**}, N_S^{**})$ are set equal to the solutions of M2 for any $s \in \{1, \ldots, S\}$, if the constraint (4.6) is met then the procedure gives (at least) a feasible solution to M1.

(ii) Otherwise, the while-loop body [lines 2-18] aims at computing the product $\tilde{o}_s = (N_s^{**} + 1)\, \mathcal{L}_s(N_s^{**} + 1)$ for any service ([lines 3-10]) and finding the service index associated to the smallest $\tilde{o}_s - o_s$ value ([line 15]).

(iii) The while-loop iterates until the constraint (4.6) is met [line 2] or, if, at any loop step, there is no $\tilde{m}_s$ such that $(\tilde{m}_s, N_s^{**} + 1) \in \mathcal{L}_s$ [lines 11-14]. Considering the latter case, the procedure selects the optimal solution of CTP as suboptimal solution of M1, see [line 13].

Procedure 4.1 Heuristic Solution of M1

1: Initialize $m_s^{**} \leftarrow m'_s$, $N_s^{**} \leftarrow N'_s$ and $o_s \leftarrow m'_s \cdot N'_s$, for $s \in \{1, \ldots, S\}$
2: **while** $\sum_{s=1}^{S} m_s^{**} > S$ **do**
3: **for** $s \leftarrow 1, \ldots, S$ **do**
4: **if** $N_s^{**} + 1 \leq \hat{N}_s$ **then**
5: $\tilde{m}_s \leftarrow \mathcal{L}_s(N_s^{**} + 1)$
6: $\tilde{o}_s \leftarrow \tilde{m}_s \cdot (N_s^{**} + 1)$
7: **else**
8: $\tilde{o}_s \leftarrow \infty$
9: **end if**
10: **end for**
11: **if** $\tilde{o}_s = \infty, \forall s \in \{1, \ldots, S\}$ **then**
12: $m_s^{**} \leftarrow 1$ and $N_s^{**} \leftarrow N''_s$, $\forall s \in \{1, \ldots, S\}$
13: **return** (m_s^{**}, N_s^{**}), $\forall s \in \{1, \ldots, S\}$
14: **end if**
15: $i \leftarrow \arg\min\{\tilde{o}_1 - o_1, \ldots, \tilde{o}_S - o_S\}$
16: $N_i^{**} \leftarrow N_i^{**} + 1$
17: $m_i^{**} \leftarrow \tilde{m}_i$
18: **end while**
19: **if** $\sum_{s=1}^{S} m_s^{**} N_s^{**} > \sum_{s=1}^{S} N''_s$ **then**
20: $m_s^{**} \leftarrow 1$ and $N_s^{**} \leftarrow N''_s$, $\forall s \in \{1, \ldots, S\}$
21: **end if**
22: **return** (m_s^{**}, N_s^{**}), $\forall s \in \{1, \ldots, S\}$

(iv) Finally, an if-then statement [lines 19-21] checks the quality of the solution found by the procedure. Considering the condition reported at [line 19], if it holds then, also in this case, the procedure chooses the optimal solution of CTP as suboptimal solution of M1.

It is straightforward to note that $(m_1^{**}, N_1^{**}), \ldots, (m_S^{**}, N_S^{**})$ is (at least) a feasible solution of M1. In addition, Procedure 4.1 returns a solutions that cannot be worse than that of the CTP model.

Unfortunately, Procedure 4.1 cannot always derive the optimal solution of M1. In fact, it is possible to find some instances of M1 such that $(m_1^{**}, N_1^{**}), \ldots, (m_S^{**}, N_S^{**})$ is a suboptimal solution. However, in all the problem instances we considered, the gap between the optimal solution and the heuristic one (namely, the solution derived by Procedure 4.1) is negligible. Finally, due to the fact that values of N_s^{**} (for each $s \in \{1, \ldots, S\}$) meet the constraint (4.7), hence, Procedure 4.1 returns in a finite number of steps which is equal to or less than $\sum_{s=1}^{S}(\hat{N}_s - K_s)$.

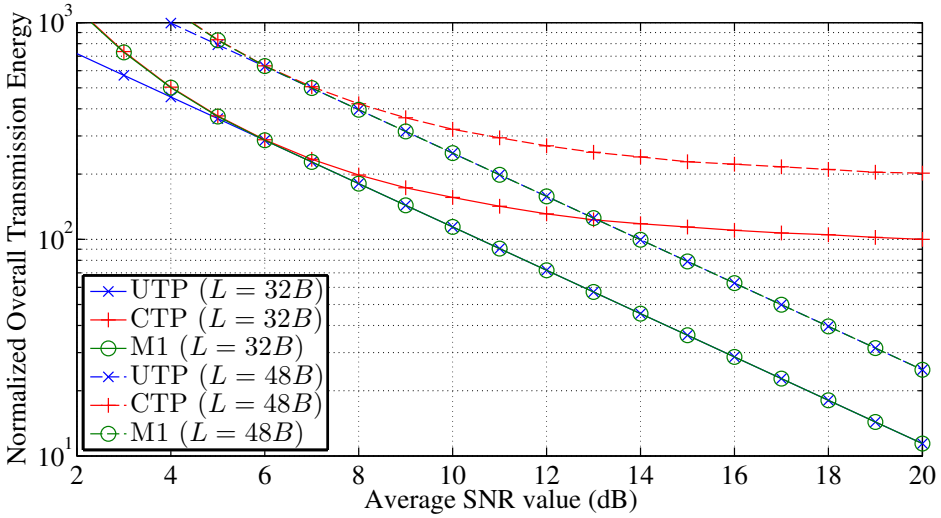


Figure 4.2: Normalized overall transmission energy vs. minimum value of $\overline{\gamma}_{o,h}$.

4.4 Numerical Results

This section compares the performance of the proposed resource allocation model (presented in M1) to the UTP and CTP alternatives. For what concerns the allocation strategy based on M1, this section shows performance results obtained by solving the problem M1 by means of Procedure 4.1.

In this section, we consider a scenario where the BS delivers $S = 3$ PtM services. Numerical results have been derived by considering a variable number of UEs belonging to the MG, namely $M \in [2, 128]$ UEs. The value of $\overline{\gamma}_{o,u}$ spans the interval $[0, 20]$ dB. Each PtM service is delivered according to the RLNC scheme which refers to a finite field of size $q = 2^8$. In addition, we consider two different information element lengths, namely L is equal to 32 or 48 bytes. The number of elements associated to an information message of each service is: $K_1 = 20$, $K_2 = 30$ and $K_3 = 40$. We assume that $\hat{N}_s = 20 \cdot N_s$ (for $s \in \{1, \ldots, S\}$). Finally, each information message has to be recovered by the MG (at least) with a probability $\hat{\Phi}$ equal to 0.8 or 0.9. Let $\overline{L}$ be the minimum value of L among the considered ones (i.e., $\overline{L} = 32$ bytes). The performance evaluation refers to the normalized overall transmission energy associated to the delivery of one information message per-PtM service, defined as $\epsilon \doteq \frac{L}{\overline{L}} \sum_{s=1}^{S} m_s N_s$. Finally, we also consider the maximum value of m_s, namely $\tilde{m} \doteq \max\{m_1, \ldots, m_S\}$, as performance index.

Let u_h be the h-th UE of the MG that experiences the worst propagation conditions, i.e., $\overline{\gamma}_{o,h} = \min\{\overline{\gamma}_{o,1}, \ldots, \overline{\gamma}_{o,M}\}$. Figures 4.2 and 4.3 show the overall transmission energy and the $\tilde{m}$ value as a function of $\overline{\gamma}_{o,h}$ (for $\hat{\Phi} = 0.9$), respectively. The figures compare both the UTP and CTP strategies to the proposed one, for different informa-

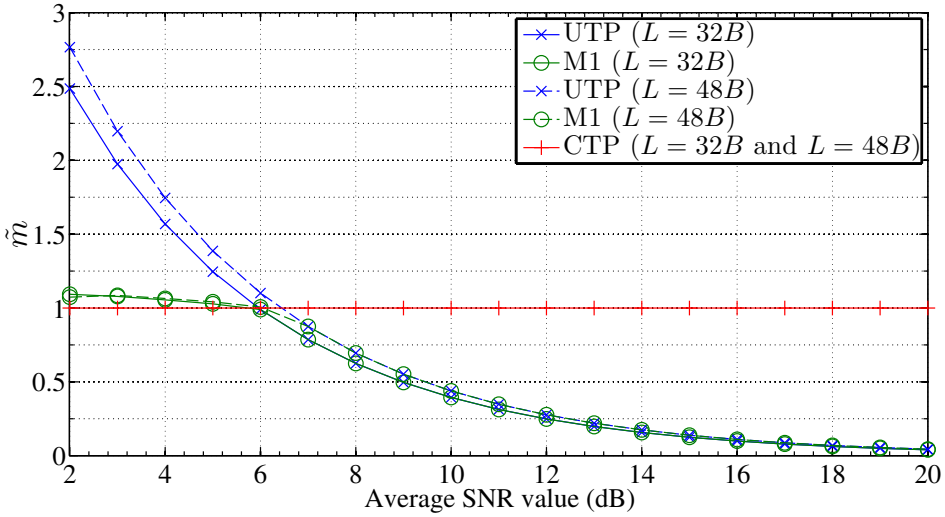


Figure 4.3: Maximum value of m_s among S PtM services vs. minimum value of $\overline{\gamma}_{o,h}$.

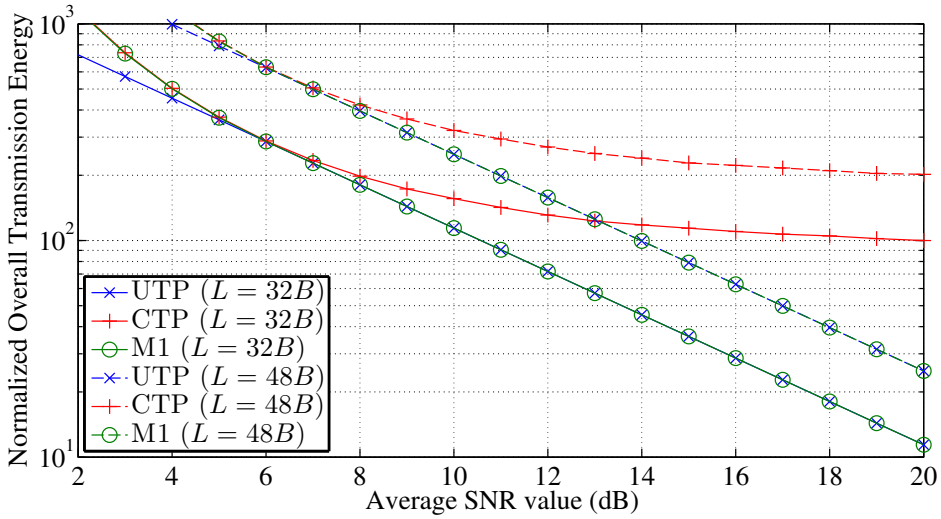


Figure 4.4: Normalized overall transmission energy vs. number of UEs M.

tion element lengths. We can note that the values of ϵ and $\tilde{m}$ of the solution derived by Procedure 4.1 overlap those associated to the UTP strategy, for $\overline{\gamma}_{o,h} > 6$ dB. In that case (i) Procedure 4.1 can derive the optimal solution of M1, (ii) the problem M1 is equivalent to the UTP one, and (iii) the performance of M1 (and hence of UTP strategy) diverges from that of the CTP strategy. For instance, if $\overline{\gamma}_{o,h} = 20$ dB, values of ϵ and $\tilde{m}$ associated to the M1 strategy are ∼8 and ∼22 times smaller than those of the CTP one, respectively.

Figure 4.4 compares the performance (in terms of ϵ) of all the considered allocation strategies as a function of the number of UEs M, for $\overline{\gamma}_{o,h} = 10$ dB, $L = 48$ bytes and different values of $\hat{\Phi}$. It is worth noting that ϵ values associated to the UTP and M1 strategies overlap, regardless of the value of M. Also in this case, the proposed strategy

(as well as the UTP one) outperforms the CTP approach.

Chapter 5

Conclusions

In this thesis, we investigated the resource allocation issues of PtM communication flows. In particular, we focused on fully reliable and delay sensitive services. Once more, we would like to remark that the considered service classes are of paramount importance in modern communication networks based on LTE and LTE-A.

This thesis deals with three main topics: (i) definition of resource allocation models suitable for error control strategies based on the HARQ-CC and NC approaches (suitable for fully reliable service delivery), (ii) development of allocation models for delay sensitive services, and (iii) definition of standard agnostic power allocation models for delay sensitive service delivery.

Chapter 2 described the MHARQ-SC approach where multiple copies of the same packet are consecutively transmitted. Of course, each receiving node combines all the received copies as stated by the SC principle. In this way, reliability of fully reliable communications is improved. The chapter proposed an efficient optimization framework which can be used to select the optimum amount of copies minimizing either the transmission energy or the delivery delay. We would like to remark that performance of the MHARQ-SC scheme has been efficiently inspected by means of the AMC theory. The performance of MHARQ-SC solution has been validated under AWGN and frequency-non-selective slow Rayleigh faded propagation conditions. The MHARQ-SC approach has been compared to multiple alternatives, such as: the classical HARQ-SC and optimized HARQ-SC schemes. Finally, the effectiveness of the proposed solution has been clearly shown.

Furthermore, Chapter 2 proposes an error control strategy based on the NC and SC principle (namely, the SDI-NC principle). Unlike MHARQ-SC, SDI-NC proposes to transmit each coded packet such that the duration of each symbol is increased by a factor m. The goal of the proposed optimization is to find the value of m which minimises the mean delivery delay and energy consumption (needed to successfully recover an

information packet). Performance of SDI-NC has been inspected both in a broadcast and butterfly network models. In particular, we focused on AWGN and slow faded propagation conditions. Results clearly show that the SDI-NC scheme significantly outperforms the existing RLNC-based alternative. Finally, at the end of the chapter, we extended the SDI-NC approach to satellite and SC-eMBMS networks.

In Chapter 3, we proposed a completely unacknowledged version of the MNC strategy suitable for delay sensitive service delivery. In particular, we developed a couple of multicast communication strategies which aim at improving system performance both in terms of transmission energy and delivery delay associated to the delivery of the whole data flow. These goals are achieved in two ways: (i) by optimally selecting the transmission data rate while the power associated to transmission of each packet is kept constant (the CP-MNC), or (ii) by optimizing the transmission power cost and keeping constant the transmission rate (the CR-MNC). Numerical results clearly showed that the proposed strategies minimise the overall transmission energy cost and significantly reduce the delivery delay in comparison to the RLNC alternative.

Chapter 3 also focused on the video service delivery over LTE-A networks. In particular, it addressed the challenge of optimizing the radio resource allocation process in eMBMS networks so that users, according to their propagation conditions, can receive video streams at the maximum achievable service level in a given cell. The main goal of the proposed optimization framework (namely, the MrNC model) is to define an efficient resource allocation strategy suitable for scalable video broadcasting (encoded by using the H.264/SVC codec). To this end, we proposed a model that can jointly optimize MCS, transmission rate and NC scheme used to efficiently deliver each H.264/SVC video layer to heterogeneous set of user groups. Finally, we demonstrated that the MrNC model we proposed can efficiently deliver layered video services over SC-eMBMS networks.

In Chapter 4, we proposed a energy efficient resource allocation model for delay sensitive service delivery based on the RLNC approach. Numerical results clearly showed that the proposed optimization model can minimize the overall transmission energy of heterogeneous PtM data flows by jointly optimizing the transmission power and the NC scheme.

Appendix A

PEP of SC Communications

This Appendix provides the expression of the Packet Error Probability (PEP) $P_B(m)$ (experienced by a generic node of a MG) used in the theoretical derivation of Section 2.1.

In this appendix we considered both the AWGN and frequency non-selective slowly-faded Rayleigh multipath fading channels. In addition, we refer with γ (AWGN regime) and $\bar{\gamma}$ (frequency non-selective, slow Rayleigh multipath fading regime) to the SNR and mean SNR characterizing the reception of a generic node, respectively.

A.1 AWGN Propagation Conditions

Let us assume that the receiving node performs an ideal coherent detection. The decision variable for each of the L bits forming a j-th received copy (of the same packet) is [50]:

$$z_i(j) = d_i\sqrt{E_b} + n_i(j) \quad i = 1, \ldots, L \tag{A.1}$$

where d_i is a random variable which can be equal to -1 or $+1$ with the same probability. If the i-th bit is 0 then $d_i = -1$, $d_i = 1$ otherwise. The term E_b is the energy associated to each transmitted bit and, $n_i(j)$ is a random variable which follows a Gaussian distribution with zero mean and a variance equal to $N_0/2$ (where N_0 is the one-side power spectral density of the white Gaussian noise affecting the communication link).

According to the SC principle [84], we can define the following decision variable:

$$Z_i = \sum_{j=1}^{m} z_i(j) = m\, d_i\sqrt{E_b} + n_i \quad i = 1, \ldots, L \tag{A.2}$$

where the term n_i results to be a Gaussian random variable with zero mean and a variance equal to $m\, N_0/2$.

According to the standard decision criterion for equiprobable symbols, we have that

the bit error probability related to the BPSK modulation is [50]:

$$P_e(m) = Q\Big(\sqrt{2\,m\,\gamma}\Big) \tag{A.3}$$

where $\gamma = E_b/N_0$. Hence, under the assumption of an ideal detecting code, the error probability related to an information packet (as function of m) is:

$$P_B(m) = 1 - \Big[1 - P_e(m)\Big]^L . \tag{A.4}$$

A.2 Slow Rayleigh Fading Conditions

Let us consider a frequency non-selective, slow Rayleigh multipath faded channel. Moreover, let us assume that all the channel parameters are known at the AP side. Hence, according to the optimal maximal ratio combiner, the overall decision variable for the i-th bit received by the generic node can be defined as follows [84]:

$$\tilde{Z}_i = \sqrt{E_b} \sum_{j=1}^{m} \alpha_j^2 + \sum_{j=1}^{m} \alpha_j\, n_j \quad i = 1, \ldots, L \tag{A.5}$$

where n_j (for $j = 1, \ldots, m$) are m *iid* Gaussian random variables with zero mean and a variance equal to $N_0/2$. Terms α_j (for $j = 1, \ldots, m$) are m *iid* Rayleigh random variables and constant for all the bits forming each copy of the same information packet. In addition, they are independent on a packet copy basis. Hence, the decision variable $\tilde{Z}_i$ follows a Gaussian distribution. Moreover, the bit error probability of the BPSK modulation is:

$$P_e(m) = Q\Big(\sqrt{2\,\psi(m)}\Big) \tag{A.6}$$

where

$$\psi(m) \doteq \frac{E_b}{N_0} \sum_{j=1}^{m} \alpha_j^2 \tag{A.7}$$

From [50], $\gamma(m)$ is a χ^2-distributed random variable with $2m$ degrees of freedom. Hence, we have:

$$p\Big(\psi(m)\Big) = \frac{1}{(m-1)!\;\bar{\gamma}^m} \gamma(m)^{m-1} e^{-\frac{\gamma(m)}{\bar{\gamma}}} \tag{A.8}$$

where

$$\bar{\gamma} \doteq \frac{E_b}{N_0} E(\alpha_i^2) \tag{A.9}$$

$E(\alpha_i^2)$ is the mean value of α_i. For these reasons, the packet error probability of an information packet can be expressed as

$$P_B(m) = 1 - \int_0^\infty \Big[1 - Q\Big(\sqrt{2\,\gamma(m)}\Big)\Big]^L \; p\Big(\gamma(m)\Big)\, d\gamma(m) \; . \tag{A.10}$$

Appendix B

Introduction to the AMC Modeling

This appendix aims at giving the necessarily theoretical background on the AMC by considering a classical telecommunication problem: the evaluation of the resequencing delay in a TDD system which adopts the Selective Repeat Automatic Repeat-reQuest (SR-ARQ) error control strategy. The reported analytical derivation can be used as an instant primer on the performance model considered in Section 2.1.1.

B.1 Resequencial Delay Model

We refer here to a wireless communication system characterized by an access scheme where the time is arranged in *frames*. In particular, we assume that each frame can hold up to k information packets. We assume that the transmitting nodes always has information packets to transmit to the receiving end.

Considering a stream of packets, each of them is labelled by a Sequence Number (SN) and they are progressively transmitted according to their own SNs (starting from the lower one). In addition, we assumed that: (i) the acknowledgement process occurs over a fully reliable feedback channel and, (ii) packet transmissions are always acknowledged within the end of a transmitted frame.

For the sake of the analysis, we assumed that packet errors occur as statically independent events.

For any erroneously received packet a negative acknowledged message (NACK) is sent to the transmitting node to request a packet retransmission (during the next frame). In what follows, we will denote as P_e the packet error probability. Finally, we assume that the capacity of the transmission buffer as well as the capacity of the resequencing one is infinite.

As soon as a packet is correctly received, the following procedure is performed: (i) the related SN is read, (ii) if there is at least *one* packet in the resequencing buffer with

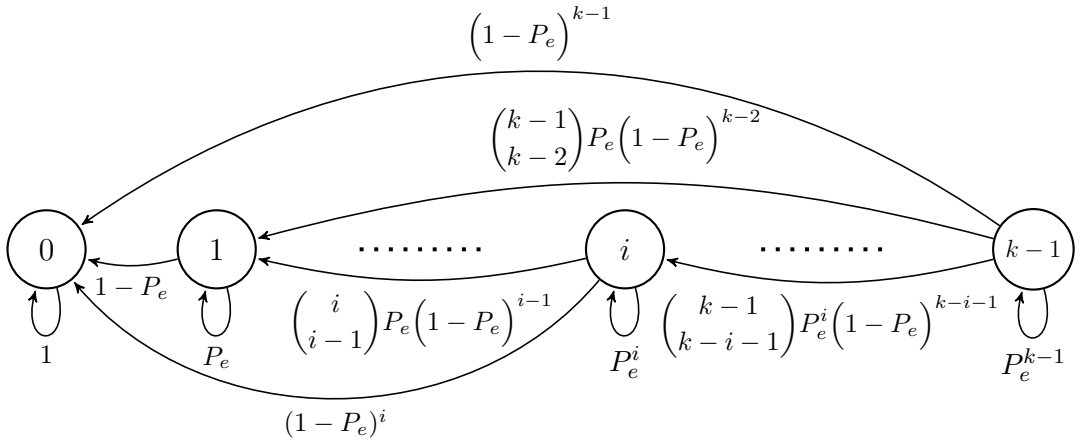


Figure B.1: State transition diagram.

a lower SN, it is stored in that buffer or, (iii) it is passed to the upper layers.

Therefore, we define the normalized resequencing delay as the time (expressed in terms of number of frames) elapsed between the correct reception of a packet and its delivery to the higher layers. Hence, the resequencing delay of the i-th information packet is null if it is directly passed to the upper layers (i.e., when it is correctly received and all the packets having a lower SN have been already passed to the upper layers).

In order to properly model the resequencing delay process of any correctly received packet, we define the i-th state s_i of that process as follows:

Definition B.1. The state s_i is equal to the number of packets in the resequencing buffer with a SN which is less than i.

On the basis of our assumptions, it is straightforward to note that the resequencing delay process can be modelled as a Markov chain with states $\{s_0, \ldots, s_{k-1}\}$. In particular, it is worth noting that the aforementioned model is an Absorbing Markov Chain (AMC) [24] because: (i) the state s_0 once entered, cannot be left (i.e., s_0 is the final state or *absorbing state* of the process) and, (ii) any state s_i (for $i = 1, \ldots, k-1$) results to be a *transient state* (i.e., once left it is never reached again).

In order to complete the definition of the AMC model we have to derive the state transition probabilities. Let us consider a packet which is correctly received and stored into the resequencing buffer in the i-position, i.e., the resequencing process associated to it starts from the state s_i. The transition probability $p_{i,j}$ from the state s_i to s_j can

be defined as follows:

$$p_{i,j} = \begin{cases} \binom{i}{i-j} P_e^j \left(1-P_e\right)^{i-j} & \text{if } i \geq j \\ \\ 0 & \text{otherwise.} \end{cases} \tag{B.1}$$

We remark that a packet is dequeued from the resequencing buffer only if *all* the packets previously transmitted (before the considered one) have been correctly received. The state transition diagram associated to the adopted AMC model is sketched in Figure B.1.

B.2 Absorbing Markov Chain Analysis

This Section provides the analysis of the resequencing delay model defined in Section B.1. In particular, by means of the AMC theory [24], we will define the *fundamental matrix* $\mathbf{N}$ associated to the AMC of interest. Finally, from the expression of $\mathbf{N}$ and the initial state probabilities of the resequencing delay process we derive the expression of the mean resequencing delay of a correctly received packet.

Let us start our analysis by providing the definition of the *fundamental matrix* $\mathbf{N}$. From (B.1), the corresponding $k \times k$ transition matrix $\mathbf{P}$ of the AMC process can be expressed as

$$\mathbf{P} \doteq \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1-P_e & P_e & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \left(1-P_e\right)^{k-2}\binom{k-2}{k-4}P_e^2\left(1-P_e\right)^{k-4} & \cdots & 0 \\ \left(1-P_e\right)^{k-1}\binom{k-1}{k-3}P_e^2\left(1-P_e\right)^{k-3} & \cdots & P_e^{k-1} \end{bmatrix}, \tag{B.2}$$

From (B.2) we note that the matrix $\mathbf{P}$ is expressed in its *canonical* form [24]. This means that $\mathbf{P}$ can be given as

$$\mathbf{P} \doteq \left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{R} & \mathbf{Q} \end{array}\right], \tag{B.3}$$

where $\mathbf{Q}$ is a $(k-1) \times (k-1)$ transition matrix which models the behaviour of the

AMC process as long as it involves only transient states; in particular, it is defined as:

$$\mathbf{Q} \doteq \begin{bmatrix} & P_e & & \cdots & 0 \\ & \vdots & & \vdots & \vdots \\ \binom{k-2}{k-4} P_e^2 \left(1-P_e\right)^{k-4} & & & \cdots & 0 \\ \binom{k-1}{k-3} P_e^2 \left(1-P_e\right)^{k-3} & & & \cdots & P_e^{k-1} \end{bmatrix} . \tag{B.4}$$

The term $\mathbf{R}$ is a $k-1$ dimensional column vector which lists the transition probabilities originating from a transient state and directed to the absorbing one. It is given by

$$\mathbf{R} \doteq \begin{bmatrix} 1-P_e \\ \vdots \\ \left(1-P_e\right)^{k-2} \\ \left(1-P_e\right)^{k-1} \end{bmatrix} . \tag{B.5}$$

Finally, $\mathbf{0}$ is a $k-1$ dimensional row vector composed by null elements.

In order to derive the average resequencing delay, it is worth referring to the Proposition B.1 which is a classical result in the AMC theory [24] (see Cap. III).

Proposition B.1. Let $\mathbf{I}$ be the $(k-1) \times (k-1)$ identity matrix. Since the matrix $\mathbf{Q}^l$ tends to $\mathbf{O}$ (which is the $k-1 \times k-1$ zero matrix) as l goes to infinity[1], the following relation holds

$$\mathbf{N} \doteq \sum_{l=0}^{\infty} \mathbf{Q}^l = \left[\mathbf{I} - \mathbf{Q}\right]^{-1} . \tag{B.6}$$

Proof. From (B.4) the (i,j)-th entry of the matrix $\mathbf{Q}^l$ is the probability of entering the j-th transient state coming from the i-th one after that l frames have been transmitted. It can be proved that the following relation hold [24]:

$$\mathbf{I} - \mathbf{Q}^l = \left(\mathbf{I} - \mathbf{Q}\right) \cdot \left(\mathbf{I} + \mathbf{Q} + \ldots + \mathbf{Q}^{l-1}\right) \tag{B.7}$$

Due to the fact that each element of the matrix $\mathbf{Q}^l$ tends to zero as l tends to infinity, $(\mathbf{I} - \mathbf{Q}^l)$ tends to $\mathbf{I}$. Hence, we have that the determinant of $(\mathbf{I} - \mathbf{Q}^l)$ is different than zero for sufficiently large values of l. For this reason we have that: (i) the determinant of the rightmost member of (B.7) is different than zero and, (ii) the determinant of

[1] In the section with $\mathbf{Q}^l$ we will refer to the l-th power of the matrix $\mathbf{Q}$.

$(\mathbf{I}-\mathbf{Q})$ is non-null. Thus, (B.7) can be rewritten as follows:

$$\left(\mathbf{I}-\mathbf{Q}\right)^{-1}\cdot\left(\mathbf{I}-\mathbf{Q}^{l}\right)=\mathbf{I}+\mathbf{Q}+\ldots+\mathbf{Q}^{l-1}=\sum_{l=0}^{\infty}\mathbf{Q}^{l} \tag{B.8}$$

Since the term $\mathbf{Q}^l$ tends to $\mathbf{O}$ as l tends to infinity, the relation (B.6) holds. □

In addition, the following proposition holds:

Proposition B.2. Let $N(i,j)$ be a generic element of $\mathbf{N}$. $N(i,j)$ results to be the mean value of the total number of times that the process, started in the state s_i, enters the state s_j (where both s_i and s_j are transient). Moreover, the mean number of consecutive frames $\gamma(i)$ after that the process (started from the state s_i) enters into the absorbing state s_0 (i.e., the mean value of the resequencing delay for a given packet) can be expressed as follows

$$\gamma(i)=\sum_{j=1}^{k-1}\mathbf{N}(i,j),\quad i=1,\ldots,k-1\ . \tag{B.9}$$

Proof. Let $Y_{i,j}$ be the random variable representing the total number of frame transmissions needed by the process, started from s_i, to reach the state s_j (where both s_i and s_j are transient). Hence, the random variable defining the total number of frame transmissions needed by the process, started from the state s_i, to reach the absorbing state s_0 is

$$T_i\doteq\sum_{j=1}^{k-1}Y_{i,j},\quad i=1,\ldots,k-1. \tag{B.10}$$

In addition, the mean value of T_i can be expressed as follows

$$\gamma(i)\doteq\sum_{j=1}^{k-1}\mathrm{E}[Y_{i,j}],\quad i=1,\ldots,k-1. \tag{B.11}$$

Let $Y_n^{(i,j)}$ be a random variable which is 1 if the process, started from s_i reaches s_j after n frame transmissions and 0, otherwise. It is straightforward to note that the following relation holds

$$Y_{i,j}=\sum_{n=0}^{\infty}Y_n^{(i,j)},\quad i,j=1,\ldots,k-1. \tag{B.12}$$

In addition from (B.4), the mean value of $Y_n^{(i,j)}$ results to be

$$\mathrm{E}[Y_n^{(i,j)}]=\mathbf{Q}^n(i,j),\quad i,j=1,\ldots,k-1 \tag{B.13}$$

where $Q^n(i,j)$ is the (i,j)-th item of the matrix $\mathbf{Q}^n$. Hence, from (B.12) and (B.13) we have that

$$\mathrm{E}[Y_{i,j}] = \sum_{n=0}^{\infty} \mathrm{E}[Y_n^{(i,j)}] = \sum_{n=0}^{\infty} \mathbf{Q}^n(i,j), \quad i,j = 1,\ldots,k-1. \tag{B.14}$$

From (B.6) and (B.14), we have that the relation $\mathrm{E}[Y_{i,j}] = N(i,j)$ holds. As a consequence, the relation (B.11) can be rewritten as follows

$$\gamma(i) = \sum_{j=1}^{k-1} N(i,j), \quad i = 1,\ldots,k-1. \tag{B.15}$$

This completes the proof. □

In order to define the mean value of the resequencing delay for any correctly received packet, we assume that when a packet is delivered for the first time, it can be transmitted into any of the k positions of the frame with a probability which is $1/k$. Hence, considering a packet (transmitted in the j-th position of the frame) which is successfully received, it starts the resequencing process from the *initial* state s_i if and only if i over $j-1$ packets (transmitted before it within the current frame) have been received with errors. Hence, the probability $\pi(i,j)$ that a packet transmitted in the j-th position of the frame enters into the resequencing buffer in the i-position (i.e., the packet resequencing process starts from the s_i) is defined as

$$\pi(i,j) \doteq \begin{cases} \binom{j-1}{i} P_e^i \left(1-P_e\right)^{j-i-1} & \text{if } i \le j \\ 0 & \text{otherwise.} \end{cases} \tag{B.16}$$

Moreover, from (B.16) we have that the probability that a packet starts its resequencing process from one of the possible states s_i (for $i = 0,\ldots,k-1$) is given by

$$\overline{\pi}(i) \doteq \sum_{j=1}^{k} \frac{1}{k}\,\pi(i,j), \quad i = 0,\ldots,k-1\ . \tag{B.17}$$

Let δ be the random variable representing the value of the resequencing delay of a packet. It follows from (B.9) and (B.17) that the mean resequencing delay $\overline{\delta}$ of a successfully received packet is

$$\overline{\delta} \doteq \sum_{i=1}^{k-1} \gamma(i)\,\overline{\pi}(i)\ . \tag{B.18}$$

Finally, it is useful to define the Complementary Cumulative Density Function

(CCDF), $\psi(t)$, of the random variable δ namely, i.e., the probability that the resequencing delay of a packet is greater than t (for any value of t which is integer and greater than 1). Let $\mathbf{p}_t$ be a row vector of k items where i-th element $p_t(i)$ is the probability that the resequencing process of a packet is in the state s_i after t frame transmissions. The vector $\mathbf{p}_t$ can be expressed as follows

$$\mathbf{p}_t \doteq \Big[p_t(0), p_t(1), \ldots, p_t(k-1)\Big] = \overline{\pi} \cdot \mathbf{P}^t \ . \tag{B.19}$$

where $\overline{\pi}$ is a k-dimensional row vector where the i-th component is equal to $\overline{\pi}(i)$ (for $i = 0, \ldots, k-1$), given by (B.17).

From (B.19) the probability that a packet is passed to the higher layers after t frame transmissions (for $t \geq 1$) is the first component of the vector $\mathbf{p}_t$ (namely, $p_t(0)$). Hence, the CCDF of the resequencing process is

$$\psi(t) \doteq 1 - p_t(0), \quad t \geq 1 \tag{B.20}$$

Appendix C

Proofs of Proposition 2.1 and 2.2

C.1 Proof of Proposition 2.1

C.1.1 AWGN Propagation Conditions

Proof. Towards this end, we rewrite (2.37), for $i = 1, \ldots, M$, as:

$$\frac{d}{d\hat{m}}\Big(\hat{\Gamma}_i(\hat{m})\Big) = \frac{1 - g_i(\hat{m}) - h_i(\hat{m})}{[1 - g_i(\hat{m})]^{L+1}} \ , \tag{C.1}$$

where

$$g_i(\hat{m}) := \mathbf{Q}\Big(\sqrt{\hat{m}\gamma_i}\Big) \ , \tag{C.2}$$

$$h_i(\hat{m}) := L\sqrt{\frac{\gamma_i \hat{m}}{2\pi}} e^{-\frac{\hat{m}\gamma_i}{2}} \ . \tag{C.3}$$

Moreover, due to the fact that both $g_i(\hat{m})$ and $h_i(\hat{m})$ decrease, the following relation holds:

$$\begin{aligned} \frac{d^2}{d\hat{m}^2}\Big(\hat{\Gamma}_i(\hat{m})\Big) \geq 0 \Leftrightarrow \Big[-\frac{d}{d\hat{m}}\Big(g_i(\hat{m})\Big) - \frac{d}{d\hat{m}}\Big(h_i(\hat{m})\Big)\Big]\Big[1 - g_i(\hat{m})\Big] \\ -(L+1)\Big[1 - g_i(\hat{m}) - h_i(\hat{m})\Big]\frac{d}{d\hat{m}}\Big(g_i(\hat{m})\Big) \geq 0 \ . \end{aligned} \tag{C.4}$$

As a result, $\frac{d\hat{\Gamma}_i(\hat{m})}{d\hat{m}}$ increases, for these reasons $\hat{\Gamma}_i(\hat{m})$ is convex in $\mathbb{R}^+/\{0\}$. In addition, for all practical operative conditions (namely for $L \geq 8$ bits and $\gamma_i \geq -1$ dB) the following relations hold:

$$\frac{d}{d\hat{m}}\Big(\hat{\Gamma}_i(\hat{m})\Big)\Big|_{\hat{m}=1} = \frac{1 - \mathbf{Q}\Big(\sqrt{\gamma_i}\Big) - L\sqrt{\frac{\gamma_i}{2\pi}}e^{-\frac{\gamma_i}{2}}}{[1 - \mathbf{Q}\left(\sqrt{\gamma_i}\right)]^{L+1}} < 0, \tag{C.5}$$

and

$$\lim_{\hat{m}\to\infty} \frac{d}{d\hat{m}}\Big(\hat{\Gamma}_i(\hat{m})\Big) > 0 \ . \tag{C.6}$$

Thus, $\hat{\Gamma}_i(\hat{m})$ has an unique minimum $\hat{m}_o \geq 1$ [85]. □

C.1.2 Slow Fading Propagation Conditions

Proof. In the case of slow fading propagation conditions, let us consider the following definitions:

$$l_i(\hat{m}) \doteq \int_0^{\infty} \Big[1 - Q(\sqrt{\hat{m}\,\gamma_i})\Big]^L e^{-\frac{1}{\overline{\gamma}_i}\gamma_i} d\gamma_i \ , \tag{C.7}$$

$$s_i(\hat{m}) \doteq \frac{L\sqrt{\hat{m}}}{2\sqrt{2\pi}} \int_0^{\infty} \sqrt{\gamma_i}\Big[1 - Q(\sqrt{\hat{m}\,\gamma_i})\Big]^{L-1} e^{-\frac{2+\hat{m}\overline{\gamma}_i}{2\overline{\gamma}_i}\gamma_i} d\gamma_i \ . \tag{C.8}$$

where $l_i(\hat{m}) : \mathbb{R}^+/\{0\} \longrightarrow \mathbb{R}^+$ and $s_i(\hat{m}) : \mathbb{R}^+/\{0\} \longrightarrow \mathbb{R}^+$ (for $i = 1, 2, \ldots, M$) are continuously differentiable in $\mathbb{R}^+/\{0\}$. The first and the second-order derivative of $\hat{\Gamma}_i(\hat{m}) : \mathbb{R}^+/0 \longrightarrow \mathbb{R}^+$ can be expressed by the following relations:

$$\frac{d}{d\,\hat{m}}\Big(\hat{\Lambda}_i\Big) = \overline{\gamma}_i \frac{l_i(\hat{m}) - s_i(\hat{m})}{l_i^2(\hat{m})}, \tag{C.9}$$

and

$$\frac{d^2}{d\,\hat{m}^2}\Big(\hat{\Lambda}_i\Big) \geq 0 \Leftrightarrow 2\frac{d}{d\,\hat{m}}\Big(l_i(\hat{m})\Big)s_i(\hat{m}) - l_i(\hat{m})\left[\frac{d}{d\,\hat{m}}\Big(l_i(\hat{m})\Big) - \frac{d}{d\,\hat{m}}\Big(s_i(\hat{m})\Big)\right] \geq 0 \ . \tag{C.10}$$

Since (C.10) is verified in any operative conditions (namely for $L \geq 8$ bits and $\overline{\gamma}_i \geq -1$ dB), we have that $\frac{d}{d\,\hat{m}}\Big(\hat{\Lambda}_i\Big)$ increases. For these reasons, the $\hat{\Lambda}_i(\hat{m})$ function is convex in $\mathbb{R}^+/\{0\}$ [58]. □

C.2 Proof of Proposition 2.2

C.2.1 Rician Propagation Conditions

Proof. Because of their definitions (see (2.61) and (2.62)), the functions $w(\hat{m})$ and $t(\hat{m})$ are continuously differentiable in $\mathcal{P}$. Moreover, the first-order derivative of $\hat{\Lambda}_{L2L}(\hat{m})$, can be expressed as follows:

$$\frac{d}{d\,\hat{m}}\Big(\hat{\Lambda}_{L2L}\Big) = \frac{w(\hat{m}) - t(\hat{m})}{w^2(\hat{m})} \ . \tag{C.11}$$

Let us consider the following parameters: $L \geq 8$ bits, a mean SNR value $\overline{\gamma} \geq 0$ dB, and a Rician factor $V \geq 1$ dB. The relation $\frac{d^2}{d\,\hat{m}^2}\left(\hat{\Lambda}_{L2L}\right) \geq 0$ holds. Hence, $\frac{d}{d\,\hat{m}}\left(\hat{\Lambda}_{L2L}\right)$ increases. For these reasons $\hat{\Lambda}_{L2L}(\hat{m})$ is convex in $\mathcal{P}$ [58]. □

Bibliography

[1] Cisco Systems, "Cisco Visual Networking Index - Forecast and Methodology, 2012-2017," Tech. Rep., 2013.

[2] ITU-T H.264, "Advanced Video Coding for Generic Audiovisual Services," Tech. Rep., Nov. 2007.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, 2007.

[4] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons, 2011.

[5] Z. Han and K. Liu, *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications*, ser. Resource Allocation for Wireless Networks: Basics, Techniques, and Applications. Cambridge University Press, 2008.

[6] R. Afolabi, A. Dadlani, and K. Kim, "Multicast Scheduling and Resource Allocation Algorithms for OFDMA-Based Systems: A Survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 240–254, 2013.

[7] P. Agashe, R. Rezaiifar, and P. Bender, "cdma2000 high rate broadcast packet data air interface design," *IEEE Commun. Mag.*, vol. 42, no. 2, pp. 83–89, 2004.

[8] K. Bakanoğlu, W. Mingquan, L. Hang, and M. Saurabh, "Adaptive Resource Allocation in Multicast OFDMA Systems," in *Proc. of IEEE WCNC 2010*, Sydney, Australia, AUS, 2010, pp. 1–6.

[9] H. Won, H. Cai, D. Y. Eun, K. Guo, A. Netravali, I. Rhee, and K. Sabnani, "Multicast scheduling in cellular data networks," in *Proc. of INFOCOM 2007*, Anchorage, US-AK, USA, 2007, pp. 1172–1180.

[10] B. Li and J. Liu, "Multirate Video Multicast Over the Internet: An Overview," *IEEE Network*, vol. 17, no. 1, pp. 24–29, 2003.

[11] D. Munaretto, D. Jurca, and J. Widmer, "Scalable video broadcast in cellular networks: impact on QoS and network resources," in *Proc. of IEEE ISCC 2010*, Riccione, Italy, IT, 2010, pp. 969–974.

[12] A. Tassi, C. Khirallah, D. Vukobratović, F. Chiti, J. Thompson, and R. Fantacci, "Reliable Rate-Optimized Video Multicasting Services over LTE/LTE-A," in *Proc. of IEEE ICC 2013*, Budapest, Hungary, HU, Jun. 2013, pp. 1–5.

[13] D. Wang, L. Toni, P. Cosman, and L. Milstein, "Uplink Resource Management for Multiuser OFDM Video Transmission Systems: Analysis and Algorithm Design," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 2060–2073, 2013.

[14] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio Resource Allocation Algorithms for the Downlink of Multiuser OFDM Communication Systems," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 3, pp. 92–106, 2009.

[15] D. Yuan, J. Joung, C. K. Ho, and S. Sun, "On Tractability Aspects of Optimal Resource Allocation in OFDMA Systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 863–873, 2013.

[16] D. Kim, T. Fujii, and K. Lee, "A Power Allocation Algorithm for Maximizing Total Utility over an MBSFN," *IEEE Wireless Communications Letters*, vol. 2, no. 3, pp. 283–286, 2013.

[17] C. Khirallah, D. Vukobratović, and J. Thompson, "Bandwidth and Energy Efficiency of Video Broadcasting Services over LTE/LTE-A," in *Proc. of IEEE WCNC 2013*, Shanghai, China, CN, 2013.

[18] G. Papadopoulos, G. Koltsidas, and F.-N. Pavlidou, "Two hybrid ARQ algorithms for reliable multicast communications in UMTS networks," *IEEE Commun. Lett.*, vol. 10, no. 4, pp. 260–262, Apr. 2006.

[19] J. Wang, S. Y. Park, D. Love, and M. Zoltowski, "Throughput Delay Tradeoff for Wireless Multicast Using Hybrid-ARQ Protocols," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2741–2751, Sep. 2010.

[20] T.-Y. Lin, S.-K. Lee, H.-H. Tang, and M.-C. Lin, "An Adaptive Hybrid ARQ Scheme with Constant Packet Lengths," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2829–2840, Oct. 2012.

[21] R. Fantacci, "Performance evaluation of efficient continuous ARQ protocols," *IEEE Trans. Commun.*, vol. 38, no. 6, pp. 773–781, Jun. 1990.

[22] J. Kim, H. Jin, D. K. Sung, and R. Schober, "Optimization of Wireless Multicast Systems Employing Hybrid-ARQ with Chase Combining," *IEEE Trans. Veh. Technol.*, vol. 59, no. 7, pp. 3342–3355, Sep. 2010.

[23] D. Chase, "A Combined Coding and Modulation Approach for Communication over Dispersive Channels," *IEEE Trans. Commun.*, vol. 21, no. 3, pp. 159–174, Mar. 1973.

[24] J. Kemény and J. Snell, *Finite Markov Chains*, ser. University series in undergraduate mathematics. Van Nostrand, 1960.

[25] N. El Heni, X. Lagrange, and P. Maule, "Optimization of link adaptation and HARQ schemes for multicast in high speed cellular networks," in *Proc. of ISWCS 2009*, Siena, Italy, IT, Sep. 2009, pp. 131–135.

[26] S. Le Digabel, "Algorithm 909: NOMAD: Nonlinear Optimization with the MADS Algorithm," *ACM Trans. Math. Softw.*, vol. 37, no. 4, Feb. 2011.

[27] M. A. Abramson, C. Audet, G. Couture, J. J. E. Dennis, S. Le Digabel, and C. Tribes, "The NOMAD project." [Online]. Available: http://www.gerad.ca/nomad

[28] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[29] D. Vukobratović, Č. Stefanović, V. Crnojević, F. Chiti, and R. Fantacci, "Rateless packet approach for data gathering in wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1169–1179, Sep. 2010.

[30] Č. Stefanović, D. Vukobratović, F. Chiti, L. Niccolai, V. Crnojević, and R. Fantacci, "Urban Infrastructure-to-Vehicle Traffic Data Dissemination Using UEP Rateless Codes," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 1, pp. 94–102, Jan. 2011.

[31] C.-C. Wang, "Pruning Network Coding Traffic by Network Coding - A New Class of Max-Flow Algorithms," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1909–1929, Apr. 2010.

[32] M. Ghaderi, D. Towsley, and J. Kurose, "Reliability gain of network coding in lossy wireless networks," in *Proc. of IEEE INFOCOM 2008*, Phoenix, US-AZ, USA, Apr. 2008, pp. 2171–2179.

[33] C. Fragouli, "Network coding: Beyond throughput benefits," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 461–475, Mar. 2011.

[34] J. Kumar Sundararajan, D. Shah, and M. Médard, "ARQ for network coding," in *Proc. IEEE ISIT 2008*, Toronto, Ontario, CA, Jul. 2008, pp. 1651–1655.

[35] P. Larsson, B. Smida, T. Koike-Akino, and V. Tarokh, "Analysis of Network Coded HARQ for Multiple Unicast Flows," in *Proc. of IEEE ICC 2010*, Cape Town, South Africa, ZA, May 2010, pp. 1–6.

[36] X. Li, T. Jiang, Q. Zhang, and L. Wang, "Binary linear multicast network coding on acyclic networks: principles and applications in wireless communication networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 738–748, Jun. 2009.

[37] F. Chiti, R. Fantacci, and D. Vukobratović, "Joint Discrete Power-Level and Delay Optimization for Network Coded Wireless Communications," in *Proc. of IEEE ICC 2010*, Cape Town, South Africa, ZA, May 2010, pp. 1–5.

[38] D. Chase, "Code Combining - A Maximum Likelihood Decoding Approach for Combining an Arbitrary Number of Noisy Packets," *IEEE Trans. Comm.*, vol. 33, pp. 385–393, May 1985.

[39] C. Fragouli and E. Soljanin, *Network Coding Fundamentals*. Hanover, US-MA, USA: Now Publishers Inc., 2007.

[40] O. Trullols-Cruces, J. Barcelo-Ordinas, and M. Fiore, "Exact Decoding Probability Under Random Linear Network Coding," *IEEE Commun. Lett.*, vol. 15, no. 1, pp. 67–69, Jan. 2011.

[41] X. Zhao, "Notes on "exact decoding probability under random linear network coding"," *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 720–721, May 2012.

[42] P. Chou and Y. Wu, "Network coding for the internet and wireless networks," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 77–85, Sep. 2007.

[43] M. Tan, R. Yeung, S.-T. Ho, and N. Cai, "A unified framework for linear network coding," *IEEE Trans. on Inf. Theory*, vol. 57, no. 1, pp. 416–423, Jan. 2011.

[44] J. Heide, M. Pedersen, F. Fitzek, and M. Médard, "On Code Parameters and Coding Vector Representation for Practical RLNC," in *Proc. of IEEE ICC 2011*, Kyoto, Japan, J, Jun. 2011, pp. 1–5.

[45] C.-C. Chao, C.-C. Chou, and H.-Y. Wei, "Pseudo Random Network Coding Design for IEEE 802.16m Enhanced Multicast and Broadcast Service," in *Proc. of IEEE VTC 2010-Spring*, Taipei, Taiwan, RC, May 2010, pp. 1–5.

[46] M. Gruber and D. Zeller, "Multimedia broadcast multicast service: new transmission schemes and related challenges," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 176–181, Dec. 2011.

[47] A. Eryilmaz, A. Ozdaglar, and M. Médard, "On delay performance gains from network coding," in *Proc. of 40th Annual Conference on Information Sciences and Systems*, Princeton, US-NJ, USA, Mar. 2006, pp. 864–870.

[48] D. Nguyen, T. Nguyen, and X. Yang, "Joint Network Coding and Scheduling for Media Streaming Over Multiuser Wireless Networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1086–1098, Mar. 2011.

[49] D. Nguyen, T. Tran, T. Nguyen, and B. Bose, "Wireless Broadcast Using Network Coding," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 914–925, Feb. 2009.

[50] J. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, Aug. 2000.

[51] A. Mood, F. Graybill, and D. Boes, *Introduction to the theory of statistics*, ser. McGraw-Hill series in probability and statistics. McGraw-Hill, 1974.

[52] F. Ohrtman, *WiMax Handbook: Building 802.16 Wireless Networks*, ser. McGraw-Hill Communications Series. McGraw-Hill, 2005.

[53] M. Médard and A. Sprintson, *Network Coding: Fundamentals and Applications*. Elsevier Science, 2011.

[54] A. Lewandowski, B. Niehoefer, and C. Wietfeld, "Concept and Performance Evaluation of a Galileo-Based Emergency Short Message Service," in *Proc. of IEEE 69th VTC Spring 2009*, Barcellona, Spain, E, Apr. 2009, pp. 1–5.

[55] F. Vieira, S. Shintre, and J. Barros, "How feasible is network coding in current satellite systems?" in *Proc. of ASMA 2010*, Cagliari, Italy, IT, Sep. 2010, pp. 31–37.

[56] J. Sun and I. Reed, "Linear diversity analyses for M-PSK in Rician fading channels," *IEEE Trans. on Comm.*, vol. 51, no. 11, pp. 1749–1753, Nov. 2003.

[57] F. Chiti, R. Fantacci, F. Schoen, and A. Tassi, "Optimized Random Network Coding for Reliable Multicast Communications," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1624–1627, 2013.

[58] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, US-NY, USA: Cambridge University Press, 2004.

[59] I. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0," http://cvxr.com/cvx, Aug. 2012.

[60] L. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gódor, G. Auer, and L. Van Der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66–72, Nov. 2010.

[61] Z. Hasan, H. Boostanimehr, and V. Bhargava, "Green Cellular Networks: A Survey, Some Research Issues and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, 2011.

[62] E. Yaacoub, "Performance study of the implementation of green communications in LTE networks," in *Proc. of ICT 2012*, Aalborg, Denmark, DK, Apr. 2012, pp. 1–5.

[63] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: overview and Rel-11 enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 68–74, Nov. 2012.

[64] P. Vingelmann, F. Fitzek, M. Pedersen, J. Heide, and H. Charaf, "Synchronized multimedia streaming on the iphone platform with network coding," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 126–132, Jun. 2011.

[65] C. Khirallah, D. Vukobratović, and J. Thompson, "Performance evaluation and energy efficiency of random network coding in LTE-Advanced," in *Proc. of IEEE ICC*, Ottawa, Ontario, CA, Jun. 2012, pp. 4574–4578.

[66] D. Vukobratović, C. Khirallah, Č. Stanković, and J. Thompson, "Random Network Coding for Multimedia Delivery over LTE-Advanced," in *Proc. of IEEE ICME 2012*, Melbourne, Australia, AUS, Jul. 2012, pp. 200–205.

[67] C. Fragouli, "Network coding: beyond throughput benefits," *Special Issue on Network Coding at the Proceedings of the IEEE*, 2011.

[68] A. Melikov, Ed., *Cellular Networks. Positioning, Performance Analysis, Reliability*. InTech, 2011.

[69] A. Ghosh and R. Ratasuk, *Essentials of LTE and LTE-A*. Cambridge University Press, 2011.

[70] N. Bisnik, "Protocol design for wireless ad hoc networks: the cross-layer paradigm," *Teknik Rapor, Rensselaer Polytechnic Institute*, 2005.

[71] S. Nazir, D. Vukobratović, and V. Stanković, "Performance evaluation of Raptor and Random Linear Codes for H.264/AVC video transmission over DVB-H networks," in *Proc. of IEEE ICASSP 2011*, Prague, Czech Republic, CZ, May 2011, pp. 2328–2331.

[72] C. Khirallah, D. Vukobratović, and J. Thompson, "Performance Analysis and Energy Efficiency of Random Network Coding in LTE-Advanced," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4275–4285, Dec. 2012.

[73] R. Fantacci, "Performance Evaluation of Efficient Continuous ARQ Protocols," *IEEE Trans. on Comm.*, vol. 38, no. 6, pp. 773–781, Jun. 1990.

[74] D. Munaretto and M. Zorzi, "Robust opportunistic broadcast scheduling for scalable video streaming," in *Proc. of IEEE WCNC 2012*, Paris, France, FR, 2012, pp. 2134–2139.

[75] R. Radhakrishnan and A. Nayak, "An Efficient Video Adaptation Scheme for SVC Transport over LTE Networks," in *Proc. of IEEE ICPADS 2011*, Tainan, Taiwan, RC, 2011, pp. 127–133.

[76] C. Tan, T. Chuah, and S. Tan, "Adaptive multicast scheme for OFDMA based multicast wireless systems," *Electronics Letters*, vol. 47, no. 9, pp. 570–572, 2011.

[77] J. Vella and S. Zammit, "A Survey of Multicasting over Wireless Access Networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 718–753, 2013.

[78] E. Magli, M. Wang, P. Frossard, and A. Markopoulou, "Network Coding Meets Multimedia: A Review," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1195–1212, 2013.

[79] J. Andrews, R. Ganti, M. Haenggi, N. Jindal, and S. Weber, "A primer on spatial modeling and analysis in wireless networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 156–163, 2010.

[80] F. Chiti, R. Fantacci, F. Schoen, and A. Tassi, "Optimized random network coding for reliable multicast communications," *IEEE Commun. Lett.*, vol. PP, no. 99, pp. 1–4, 2013.

[81] 3GPP TR 36.814 v9.0.0 (Release 9), "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects," 2010.

[82] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, 1999.

[83] "YUV Video Sequences." [Online]. Available: http://trace.eas.asu.edu/yuv/

[84] D. Brennan, "Linear diversity combining techniques," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 331–356, Feb. 2003.

[85] D. P. Bertsekas, *Convex Analysis and Optimization.* Athena Scientific, 2003.