

CONTRIBUTIONS ON NETWORKING TECHNIQUES FOR WIRELESS RELAY  
CHANNELS

Smrati Gupta  
Supervisor: Dr. M. A. Vázquez-Castro

PhD Programme in Telecommunications and Systems Engineering  
Department of Telecommunications and Systems Engineering  
Universitat Autònoma de Barcelona

May 2014

# Abstract

In the recent years, relaying has emerged as a powerful technique to improve the coverage and throughput of wireless networks. Consequently, the growing demands of the wireless relay networks based services has led to the development of novel and efficient networking techniques. These techniques can be used at different layers of the protocol stack and can be optimized to meet different objectives like throughput maximization, improving coverage etc. within existent networking framework. This thesis presents a series of contributions towards the networking techniques using a variety of tools in order to maximize the throughput of the network and satisfy the user demands. To make effective and concrete contributions, we have selected challenging problems in various aspects of advanced wireless networking techniques and presented neat solutions to these problems. In particular, we make use of the different tools like network coding, cognitive transmission techniques and game theory in order to design networking solutions for modern wireless relay networks. The main contributions of this thesis towards networking techniques at different layers of the protocol stack are as follows.

Firstly, at the physical layer, we maximize the throughput of the network using the tool of physical layer network coding (PNC) based on compute and forward (CF) in relay networks. It is known that the maximum achievable rates in CF-based transmission are limited due to the channel approximations at the relay. We propose the integer forcing precoder (IFP), which bypasses this maximum rate achievability limitation. With the help of IFP, we demonstrate a possible implementation of the promising scheme of CF thereby paving the way for an advanced precoder design to maximize network throughput.

Secondly, at the link-network layer, we maximize throughput with the use of two different tools: (a) network coding along with Quality of Experience (QoE) driven cross-layer optimization and (b) cognitive transmission techniques. For (a), we use network coding at link layer in coherence with cross-layer optimization and prove the existence of crucial trade-offs between throughput and achievable QoE. Moreover, it is proposed to use the realistic factors such as positioning of the end users in the relay network to optimize the service obtained in presence of such trade-offs. For (b), we use the cognitive transmission techniques to analyze the improvement in throughput of a particular wireless network, namely Dual Satellite systems (DSS). Moreover, an exhaustive taxonomic analysis of the different cognitive techniques in DSS is presented. With the help of this work, the possible designs for 'intelligent' networking techniques are proposed, which form a platform for maximizing the throughput performance of future wireless relay networks.

Thirdly, at the transport-application layer, we maximize not only the throughput but a joint utility comprised of throughput, QoE and cost of service, with the use of game theoretical tools. We consider a video application relayed over a wireless network and competing users trying to maximize their utilities. We model and predict the equilibriums achieved using repeated game formulations taking into account the realistic factors such as tolerance of the users and pareto optimality. With the help of this work, the potential of use of repeated game theoretical tools in wireless networks is proved which also promises to improve the existing system performance categorically.

Overall, this thesis presents effective and practical propositions along with holistic analysis towards different aspects of development of modern networking techniques for wireless relay networks.

# Acknowledgement

It was an absolute pleasure to be a graduate student of Dr. M. A. Vázquez-Castro. Apart from her guidance and support, she has been an inspiration at both personal and professional level. Thank you for letting me wander in the pool of research and steering me through at times when I lost track. Thank you for those brainstorming sessions, for your immense patience and for teaching me the power of identifying the essential key points within dense information. It has been a fortunate experience to work under your supervision and it is truly difficult to condense my gratitude in words.

I would like to thank Dr. E. V. Belmega, with whom I was given the opportunity to work on game theoretical tools. Every discussion with her, technical or otherwise, was enriching and encouraging.

I would never have been able to find my way through the field of research if it wasn't for the motivation from my professors at LNMIIT and Dr. S. Kota. Thank you for motivating me to pursue a PhD.

I was fortunate to be a part of the Wireless and Satellite Communications group and have exceptionally wonderful colleagues as Lei, Ricard, Alejandra and Paresh creating perfect office environment. I have spent an immense part of past years in company of my office mates who became great friends. Thank you Alejandra for the ever so refreshing evening coffee breaks and keeping me company in those late nights in the office chasing the deadlines.

A special thanks to Paresh for all the help, encouragement, laughs and 'count on me' friendship. Thanks for giving a good humor even with the erratic office hours, taking breaks to coffee machine, and white board discussions. Thanks for being my ultimate punching bag.

I am also grateful to all the colleagues at Department of Telecommunications and Systems Engineering in UAB. Also, I would like to thank the staff at the department to get through the long procedures of bureaucracy and technicalities.

I would like to thank Kalpana Srivastava (*didi*) for being a great support as a family away from family. Thank you for sharing and taking care of me in all the joyous and not so joyous times, for those delicious meals, wonderful weekends and giving the humorous one-liners when I needed the most.

I am blessed to have a brother like Manu Gupta whose unconditional love, care and affection kept me going through highs and lows. I can't thank you enough for all those motivational words and songs to cheer me up, for the much needed getaway weekends and believing in me all the time. It would not have been possible without you.

I thank my parents, Mr. S. K. Gupta and Mrs. Rani Gupta, for being the pillars of strength, inspiration and motivation. Thank you for everything and specially for giving me the gift of time. I am truly blessed in being your daughter. Thanks dad for giving me a perspective on bigger picture and thanks Mom, for all your prayers and Skype availability.

A special thanks to Late Mrs. Krishna Kota and her excellent book on 'Sri Lalita Sahasranama' which gave me the spiritual support when I needed it most.

Its not possible to really acknowledge everyone in this space, but I would like to thank all my professors, colleagues, friends and family members, who knowingly or unknowingly, have helped me through this study. It has been an experience to cherish through a lifetime.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgement</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation and Objectives . . . . .	1
1.2 Outline of the dissertation . . . . .	2
1.3 Research Contributions . . . . .	3
<b>2 Preliminaries</b>	<b>5</b>
2.1 Modeling . . . . .	5
2.1.1 Channel Model . . . . .	5
2.1.2 Layered Architecture Model . . . . .	6
2.2 Tools and Techniques . . . . .	7
2.2.1 Network Coding . . . . .	7
2.2.2 Physical Layer Network Coding with Compute and Forward . . . . .	8
2.2.3 Game Theoretical Tools . . . . .	9
2.3 Performance Metrics . . . . .	11
2.3.1 Offered Throughput . . . . .	11
2.3.2 Offered Quality of Experience . . . . .	11
2.4 Overall Methodology to networking solutions . . . . .	12
<b>3 Networking at Physical Layer</b>	<b>13</b>
3.1 Introduction . . . . .	13
3.1.1 Objective and technique in overall framework . . . . .	13
3.1.2 Outline of Chapter . . . . .	14
3.2 System Model . . . . .	14
3.2.1 Phase I: Source to Relay Multiple Access Channel . . . . .	14
3.2.2 Phase II: relay-to-destination point-to-point channel . . . . .	15
3.3 Phase I: Proposed Precoding-based Source to Relay Transmission . . . . .	16
3.3.1 Problem Formulation . . . . .	16
3.3.2 Proposed Integer Forcing Precoder . . . . .	17
3.3.3 Decoding at the relay . . . . .	19
3.4 Phase II: Proposed point-to-point based relay to destination transmission . . . . .	19
3.4.1 Problem Formulation . . . . .	20
3.4.2 Decoding linear combination at the destination . . . . .	20
3.4.3 Decoding original signals at the destination . . . . .	20

3.5	Performance Analysis . . . . .	21
3.5.1	Probability of error from source to relay node . . . . .	21
3.5.2	Probability of error from relay-to-destination node . . . . .	23
3.5.3	Overall end-to-end probability of error . . . . .	24
3.6	Numerical Results . . . . .	25
3.6.1	Probability of error . . . . .	25
3.6.2	Outage Probability . . . . .	26
3.6.3	Achievable Rates . . . . .	27
3.7	Concluding Remarks . . . . .	28
3.7.1	Contributions towards state-of-the-art . . . . .	28
3.7.2	Related Publications . . . . .	29
3.7.3	Possible Future lines of work . . . . .	29
<b>4</b>	<b>Networking at PHY-MAC layer</b>	<b>30</b>
4.1	Introduction . . . . .	30
4.1.1	Objective and technique in overall framework . . . . .	30
4.1.2	Outline of the Chapter . . . . .	33
4.2	Maximizing throughput using QoE-driven Cross-layer optimization for video transmission . . . . .	33
4.2.1	Cross-layer design preliminaries . . . . .	33
4.2.2	System Topologies . . . . .	34
4.2.3	Proposed Joint Network Coded Cross-layer Optimized Model . . . . .	36
4.2.4	Throughput-QoE Tradeoff . . . . .	39
4.2.5	Location Adaptive Joint Network Coded Optimized Model . . . . .	40
4.2.6	Performance Metrics . . . . .	41
4.2.7	Performance Analysis . . . . .	43
4.3	Maximizing throughput using cognition . . . . .	48
4.3.1	Concept of Cognition . . . . .	49
4.3.2	Dual Satellite Systems . . . . .	49
4.3.3	Cognition in DSS . . . . .	51
4.3.4	Taxonomic Analysis of Cognitive DSS . . . . .	53
4.4	Concluding remarks . . . . .	56
4.4.1	Contributions over state-of-the-art . . . . .	56
4.4.2	Publications . . . . .	57
4.4.3	Future lines of work . . . . .	57
<b>5</b>	<b>Networking at Transport -Application layer</b>	<b>59</b>
5.1	Introduction . . . . .	59
5.1.1	Objective and technique in overall framework . . . . .	59
5.1.2	Outline of the chapter . . . . .	61
5.2	System Model and QoE-driven Rate Adaptation . . . . .	61
5.2.1	System Model . . . . .	61
5.2.2	QoE-driven Rate Adaptation and Performance Metrics . . . . .	62
5.3	Game Theoretical Framework for QoE-driven Adaptive Video Transmission . . . . .	65
5.3.1	One-shot Non-cooperative Game . . . . .	65
5.3.2	Finite Horizon Repeated Game . . . . .	66
5.3.3	Infinite Horizon Repeated Game . . . . .	68
5.4	Selection of sustainable agreement profiles . . . . .	71
5.4.1	Tolerance Index . . . . .	71
5.4.2	Pareto Optimality . . . . .	72

5.5	Numerical Results . . . . .	72
5.5.1	Discount factor variation . . . . .	72
5.5.2	Tolerance Analysis and Pareto optimality . . . . .	74
5.5.3	Influence of weights of QoS , QoE and Cost . . . . .	76
5.5.4	QoS and QoE analysis . . . . .	78
5.6	Concluding remarks . . . . .	78
5.6.1	Contributions over state-of-the-art . . . . .	78
5.6.2	Publications . . . . .	79
5.6.3	Future lines of work . . . . .	80
<b>6</b>	<b>Overall Conclusions and Future work</b>	<b>81</b>
6.1	Conclusions . . . . .	81
6.2	Future Work . . . . .	83
<b>A</b>	<b>Proof of Proposition 16</b>	<b>84</b>
	<b>Bibliography</b>	<b>87</b>

# List of Figures

2.1	Two Way Wireless Relay Channel . . . . .	6
2.2	Internet Protocol Suite or TCP/IP Protocol Stack . . . . .	7
2.3	Relaying schemes in two way wireless relay channel . . . . .	9
2.4	Unified Methodology devised to obtain the networking contributions in the thesis . . . . .	12
3.1	System Model . . . . .	14
3.2	Comparison of probability of error at the relay for proposed scheme with varying SNR for $L = 2$ . The total signal power is adjusted for fair comparison with other schemes. . . . .	26
3.3	CF with IFP compared to TDM performance at the destination . . . . .	26
3.4	Outage probability curve. The horizontal axis shows the power available at the transmitter. The power available at the transmitter is assumed to be always greater than or equal to the signal power hence $\gamma_a > 1$ . With the increasing channel variance $\sigma_h$ , the outage decreases. . . . .	27
3.5	Comparison of achievable rates for various schemes for $L = 2$ . The average value is taken over 10000 different random channel realizations. The rates are measured in bits per channel use (bpcu). . . . .	28
4.1	System topologies . . . . .	35
4.2	Model NC and CL in bent-pipe topology . . . . .	36
4.3	Model NC+CL . . . . .	37
4.4	Model NC and CL with X-topology. The dotted line shows the overhearing links. . . . .	37
4.5	Model NC+CL. The source rate is adaptive and the relay network-codes the packets. . . . .	38
4.6	Flowchart to implement switching mechanism. The coordinates (x,y) represent the location of destination nodes. . . . .	41
4.7	Model NC+CL provides a fine balance in terms of throughput and flow continuity giving high joint utility . . . . .	44
4.8	The higher the drop, the better QoE for trade-off aware adaptation with smaller throughput compromise. . . . .	45
4.9	Comparison of throughput and flow continuity for varying channel capacity drops and varying probability of drop . . . . .	45
4.10	Performance Comparison in terms of QoS (Network Throughput) and QoE (Flow continuity). Note that in case there is no loss in overhearing, the switching mechanisms are not utilized . . . . .	46
4.11	QoS and QoE variation in Model NC+CL with NC-based switching for different geographical placements of destination nodes . . . . .	47
4.12	QoS and QoE variation in NC+CL with CL-based switching for different geographical placement of destination nodes . . . . .	47
4.13	Location-optimal Model Map: the optimal switching to be applied at different geographical placement of destination node. . . . .	48
4.14	Some typical DSS scenarios . . . . .	49

4.15	Proposed Block Diagram for Cognitive Cycle . . . . .	51
4.16	Translation of different steps of Cognitive Cycle into Engineering Design Steps . . . . .	52
5.1	Network Topology. The nodes A and B exchange information via R. The arrows indicate the links with positive channel capacity. . . . .	62
5.2	Rate adaptation curve. The red line indicates the rate $r_k(n)$ . The blue line indicates the channel capacity. Angle $\alpha$ remains constant for all time $n$ , whereas period $p$ depends on channel capacity. . . . .	63
5.3	Sustainable agreement region and corresponding discount factor region. . . . .	73
5.4	Variation of sustainable agreement region for varying tolerance index. As the tolerance index increases, the non-symmetric points become achievable. With the least tolerance, the players agree to the same adaptation rates where both of them get same quality video. . . . .	74
5.5	Comparison of the achievable payoffs with varying tolerance index. As the tolerance index increases, more payoffs are sustainable. The number of Pareto optimal payoffs reduce as the tolerance index decreases. . . . .	75
5.6	The region of all agreement points $(\alpha_1^*, \alpha_2^*)$ with weights assigned as $w_1 = w_2 = 0.4$ and $w_3 = 0.2$ . The sustainable agreement region is reduced with a higher weightage to the cost. Due to the high cost for faster adaptation to transmit the video, the users agree to slow adaptation points. . . . .	76
5.7	Comparison of the region of sustainable agreement points for different weights to QoS and QoE. The weight to the cost is fixed. The higher the weightage to QoS, the more is the achievable region and faster rate adaptation are sustainable. . . . .	77
5.8	Variation of QoS of the video at the pareto optimal outcomes of the game for different initial rates. Higher the discount factor, higher is the QoS. . . . .	77
5.9	Variation of Quality of Experience at different initial rates. Higher the discount factor and higher the initial rate, lower is the QoE. . . . .	79



# Chapter 1

## Introduction

### 1.1 Motivation and Objectives

Wireless networks have become an integral component of modern age communication systems. In the past decade, the demands of the wireless networks based services have tremendously increased, particularly with the advent of mobile and handheld devices. These demands are in terms of both higher amount of bandwidth for advanced and ubiquitous services and higher number of users required to be supported within the available resources and infrastructure. It ushers the need to maximize the network efficiency such that it is able to support the growing demand within the limited resources. This is achieved by development of efficient protocols which are able to transmit the traffic efficiently over the wireless networks and satisfy the target performance. The wireless network is composed of different nodes who want to transmit and/or receive data and these protocols are designed such that the data is transmitted across the network from the sending to the receiving node in an efficient manner. Such protocols form the basic components of wireless networking. The world of wireless research is constantly trying to find novel and efficient ways to improve the current networking techniques for next generation wireless demands.

Wireless relay networks form a natural choice of the network to be considered for the design of advanced networking techniques. This is because the wireless networks, in particular, have a powerful characteristic of a wide coverage of geographically distributed users. However, the physical and environmental channel conditions affect the wireless signal strength over large distances. This leads to the need of relaying of the signals across widely distributed areas using intermediate relay node(s) to allow onward transmission of the signals from source to destination. Hence, wireless relay networks are highly beneficial for wider coverage besides other benefits.

The development of efficient networking techniques further involves consideration of the network protocol stack which forms the backbone of the networking model. The well-known networking protocol stack shows the modularization of the networking protocols. The enhancement of networking techniques are associated with optimization of the technologies at all the levels of the stack namely at Physical (PHY) layer, Medium Access Control (MAC) or data link layer, Network (or IP) layer, Transport layer and Application layer (APP). In addition, the existent framework of wireless networking allows a number of cross-layer optimization techniques to enhance the overall network performance by taking advantages at different layers. It is imperative to assess the inter-play between these cross-layer technologies in existent frameworks and the advanced networking techniques 'purely' in a particular layer. This interplay is yet to be explored for many novel techniques. Besides, the optimization of the networking techniques based on different 'perceptive and random' factors like geographical distribution, user perception and behavior etc. have been found to play a

key role in the optimal data transmission over wireless networks. A concrete study of effect of such factors on the novel advanced networking techniques is yet to be analyzed.

Inspired by the need for the enhancement of throughput in wireless relay networks, this thesis makes a series of contributions towards the different aspects of networking techniques in wireless relay networks at various layers of the protocol stack. In this thesis, our goal is to study the wireless relay networks and propose contributions towards networking techniques in order to maximize the throughput using novel and intelligent algorithms. This goal is primarily achieved with the use of different advanced tools like physical layer network coding, packet level network coding, cognitive transmission tools and game theoretical tools. This thesis contributes towards improvement of the networking techniques which promises to form the basis of next generation wireless networks with an aim to optimize the overall network performance and satisfy the user demands.

## 1.2 Outline of the dissertation

This thesis largely focuses the design and assessment of advanced networking techniques for the improvement of the overall performance of the wireless relay networks. We cover several aspects of the wireless networking spanning across different layers of the protocol stack and using different analytical tools to quantize the performance improvement. To present these aspects with clarity, based on the layer of the protocol stack at which the technique has been applied, this thesis has been organized into the following chapters:

1. Chapter 1: This chapter introduces the overall motivation, objective and methodology of the research work carried out. It also summarizes the research contributions made during the course of preparation of this thesis.
2. Chapter 2: In this chapter, we cover the basic preliminaries from state-of-the-art required for the understanding of the contributions made in the rest of the work. The preliminaries address the models, tools and performance metrics which together comprise the methodology followed. This chapter presents the model of the wireless relay channel which is largely the channel model considered in the thesis. It also presents the protocol stack model based on which we classify the different contributions made towards networking techniques in rest of the chapters. The preliminaries of the tools including both performance enhancement tools (eg. network coding) and assessment tools (e.g. game theory) along with the performance metrics which are used to quantize the contributions are also presented in this chapter.
3. Chapter 3: This chapter introduces the contribution towards the networking techniques on the physical layer of the protocol stack. With the objective of improving the overall throughput of the relay wireless networks, we investigate the novel techniques of physical layer network coding with compute and forward and assess its limitations. Primarily, we focus on the lattice based implementation of CF which draws the limitation from the non-integral channel gains. We in turn propose solution(s) to this limitation using precoding and assess the performance using this technique. Both analytical and experimental propositions are made for end to end analysis of the relay networks at the physical layer. Finally, we justify the overall scheme implementation with both encoders and decoders designs.
4. Chapter 4: In this chapter, we consider the layer above the physical layer in the protocol stack by focussing on the MAC and network layer and designing networking schemes implemented at these layers. The objective is to improve the overall performance measured in terms of throughput of the wireless relay network. At this layer, we address the maximization of throughput in coherence with two other frameworks. Firstly, we use the tool of network coding along with cross-layer optimization framework to obtain improvement in the overall throughput. A particular focus is made on the video

applications and the associated quality of experience obtained, along with the primary objective of improved throughput. The related trade-offs in such scenarios are proposed in different topologies of relay networks. In addition, a novel optimization of trade-offs using the geographical location of the nodes proposed and analyzed. Secondly, we use the tool of cognition in wireless relay networks, particularly satellite framework, to maximize the overall throughput. To this end, this chapter presents a thorough taxonomical analysis of the different cognitive techniques in the literature. A special focus is made on the satellite networks and a concrete mapping is provided between the cognitive communication steps and the corresponding engineering design steps to accomplish such techniques in future generation networks.

5. Chapter 5: This chapter is focussed upon the investigation of networking techniques in the upper layers of the protocol stack. In this part of thesis, we model and optimize the affect of interplay of transport/application layer networking in wireless relay channel and the end-user assessment of the application performance. The objective of this work is to maximize the performance in terms of not only the throughput but also the performance measures which assess user satisfaction. We study a particular video exchange application in which two users exchange their video streams using Quality of Experience (QoE) driven rate adaptation at the transport/application layer. In such an interaction, both users tend to maximize the QoE of their received video while minimizing their individual cost incurred to transmit the video. Such an interaction is modeled via game theoretical tools as a non-cooperative game. The outcome of this game is studied when played once, and when played repeatedly via characterization of the Nash equilibrium region and the sub-game perfect equilibrium region respectively. It is shown that adaptive video exchange between selfish autonomous nodes for a limited time will not be sustained. However, if the video is exchanged over a longer period of time, the nodes have an incentive to cooperate and exchange video stream with QoE-driven rate adaptation based on the trust they build among themselves. Furthermore, the outcomes of the game that are most likely to be obtained at the end users are identified based on the pareto efficiency of the outcome of the repeated games and the tolerance level of the users.
6. Chapter 6: The final chapter of this dissertation summarizes the overall conclusions drawn from the different aspects studied in this work. We present the general implications of the contributions made via this thesis and also identify the future lines of work which can be pursued.

### 1.3 Research Contributions

The main goal of the this thesis was to find novel ways to improve the throughput of wireless relay networks by exploiting layered structure of the protocol stack with application of diverse tools. Consequently, the work in this thesis has resulted in different scientific publications and project reports. The contributions are listed as follows:

#### *Journals*

1. **S. Gupta**, M. A. Vázquez-Castro, "Physical Layer Network Coding Based on Integer Forcing Precoded Compute and Forward." *Future Internet* 5, no. 3: 439-459, 2013.
2. **S. Gupta**, E.V. Belmega, M. A. Vázquez-Castro, "A game-theoretical analysis of QoE-driven adaptive video transmission". *Submitted to Springer Journal on Multimedia Systems*, Nov. 2013.

#### *Book Chapter*

1. **S. Gupta**, M. A. Vázquez-Castro, R. Alegre-Godoy, "Dual Satellite Systems", Chapter in the book *Cooperative and Cognitive Satellite Systems*, Elsevier 2014.

**Conference Proceedings**

1. **S. Gupta**, M. A. Vazquez-Castro, "Physical-layer network coding based on integer-forcing pre-coded compute and forward," *Wireless and Mobile Computing, Networking and Communications (WiMob), 2012 IEEE 8th International Conference on*, vol., no., pp.600-605, 8-10 Oct. 2012
2. **S. Gupta**, M. A. Pimentel-Niño, M. A. Vázquez-Castro, "Joint network coded-cross layer optimized video streaming over relay satellite channel", In *3rd International Conference on Wireless Communications and Mobile Computing (MIC-WCMC)*, Valencia, June 2013.
3. **S. Gupta**, M. A. Vázquez-Castro, "Location-adaptive network-coded video transmission for improved Quality-of-Experience" In *31st AIAA International Communications Satellite Systems Conference (ICSSC)*, Florence, (2013).
4. **S. Gupta**, E.V. Belmega, M. A. Vázquez-Castro, "Game-theoretical Analysis of the tradeoff between QoE and QoS over satellite channels". *Submitted to ASMS/SPSC 2014*.

Moreover, this thesis has also been result of knowledge gained from participation in the following research projects:

**Projects**

1. Collaboration in the project "GEO-PICTURES", funded by the 7th Framework Programme of the European Commission.
2. Participation in European Union COST Action IC1104 on Random Network Coding and Designs over  $GF(q)$ .
3. Participation in SATellite Network of EXperts project (SATNEX-III).

The association of specific contribution towards each chapter will be pointed out at the end of each chapter in the thesis.

## Chapter 2

# Preliminaries

In this chapter, we introduce the preliminaries forming state-of-the-art in wireless relay networks. Using these preliminaries we will develop a methodology that we follow to achieve the objectives.

Our study covers a wide range of networking issues thereby addressing the overall throughput and/or user experience using wireless relay networks. We devise a simple methodology which we will follow for different contributions that we address in this thesis. Broadly speaking, the methodology we adopt comprises of three selection steps: Selection of Model, Selection of Tools and Selection of Performance metrics. The first step comprises of developing a clear and concrete model of the wireless relay network addressed in the assessment, the second step involves selection of an appropriate tool to be implemented to assess/optimize the performance and the third step involves selection of the exact performance metrics used to quantize the projected improvement. Before explaining the methodology in detail, we will first explain the preliminary components, which are derived from state-of-the-art, used to build up the methodology.

### 2.1 Modeling

We will now present some of the fundamental models which have been used in this thesis.

#### 2.1.1 Channel Model

The wireless networks have the fundamental advantage over wired network due to the ease in the ability to connect with a number of user without much physical resources but using the shared media of air. This also leads to a higher coverage of the wireless network in difficult terrains. Furthermore, the coverage can be enhanced (along with the throughput) using relay based strategies in wireless networks. Nodes in a relay network perform one or more of the three roles: Sources transmit information packets into the network, destinations are interested in recovering a set of information packets and relays help move information between source and destinations.

Wireless relay network holds several advantages. Information can travel long distances, even if the sender and receiver are far apart. It also speeds up data transmission by choosing the best path to travel between nodes to the receiver. If one node is too busy, the information is simply routed to a different one. Without relay networks, sending an information from one node to another would require the two nodes be connected directly together before it could work. However, the wireless relay networks face the challenge of interference when a number of users share the bandwidth and use the same node as relay. This interference limits the useful information transmitted across channel due to inability to decode the information affected by interference. This affects the overall throughput of the network.

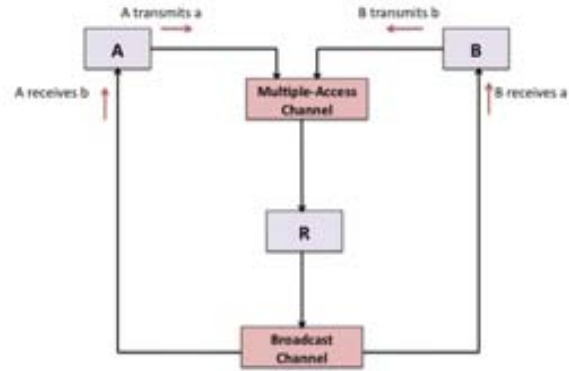


Figure 2.1: Two Way Wireless Relay Channel

Inspired by the advantages of relay networks, we focus on a particular relay channel in this thesis namely, the two way relay channel. This example first appeared in the paper by Wu. et. al. in 2004 [1]. As shown in figure. 2.1, two nodes A and B want to exchange the information. However, since they cannot hear each other, they communicate with the help of a relay node R. All the links between the nodes are entitled to channel fading at the physical layer, transpired as channel erasures at the higher layers. The users share the same frequency band and each node operates in half duplex mode (i.e., it can either send or receive during a single time slot but not both). If both users transmit simultaneously, the relay node will hear a superposition of the two signals scaled by channel gains and corrupted by noise. Therefore, this effect can be modeled by a multiple-access channel with inputs from A and B to R. Next, whatever the relay sends can be heard by both the nodes A and B and therefore, this channel can be modeled as a broadcast channel with input from relay sent to A and B. Both the multiple-access channels and broadcast channels have been well studied in the literature. See the book [2] for details.

### 2.1.2 Layered Architecture Model

We will now describe the network layered architecture design which forms the basis of our signal models in the thesis.

Today's wireless networks use the layered architecture (commonly addressed as network protocol stack) to decouple the different functionalities involved in the networking process like wireless signaling schemes, flow control, scheduling etc. Each layer in the protocol stack hides the complexity of the layer below and provides a service to the layer above. The advantages of layered architecture are provision of a network design that is scalable, evolvable and implementable [3]. A widely accepted protocol stack is given by Internet Protocol Suite and is commonly called TCP/IP which is shown in figure. 2.2. A brief description of the component layers of this architecture is as follows (for details, refer [4]):

- **Physical Layer:** The physical layer is concerned with transmitting raw bits over the physical links of a communication channel. It provides the electrical, mechanical and procedural interface to the transmission medium.
- **Medium Access Layer:** The link layer converts the stream of bits into packets/frames. The data link layer is traditionally responsible for issues of error control mechanism, physical addressing, frame synchronization etc.

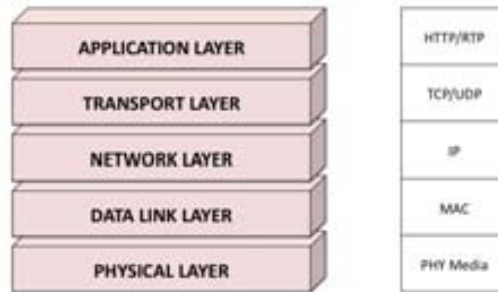


Figure 2.2: Internet Protocol Suite or TCP/IP Protocol Stack

- Network Layer: The network layer is responsible for packet forwarding through the routers using the Internet Protocol (IP) which defines the IP addresses. This layer defines the addressing and routing structures.
- Transport Layer: The transport layer accepts data from the application layer and splits it into smaller units, if needed. It is an end-to-end layer all the way from source to destination and can provide congestion control mechanisms depending on the protocol being used. The Transport Control Protocol (TCP) and User datagram Protocol(UDP) are the main transport layer protocols used.
- Application Layer: The application layer contains all the higher-level protocols, like File transfer Protocol (FTP), electronic-mail (SMTP) etc. The application layer deals with process to process communication.

The networking techniques can be enriched and optimized at every layer of the protocol stack to obtain the desired gains in the transmission. Therefore, in this thesis, we will look into each of these layers separately and contribute towards enhancement of networking paradigms.

## 2.2 Tools and Techniques

The next step after developing an appropriate model and thereby formulating the problem, is to use the appropriate tools and techniques to realize the performance improvements. We now illustrate the main tools and techniques used in the thesis.

### 2.2.1 Network Coding

Network coding is a novel technique which promises to change the paradigm of future networking. It was first proposed in 2001 in the seminal paper of Ahlswede [5] in which it was shown that network coding allows communication networks to achieve multicast capacity based on Shannon's theory [6]. Thereafter, it has been widely studied by the research community. A number of surveys [7, 8, 9] and books [10, 11] outline the prime developments in the field and its application. The concept of network coding was initially introduced for the wireline network, but later it was extended to wireless networks [12] primarily because they

serve a natural platform to avail the advantages of network coding. Network coding employs the concept of intelligent mixing of signals from multiple sources at the intermediate nodes such that they can be decoded at the destination. Let us explain the idea underlying the network coding using the famous Alice-Bob example.

Consider the example in figure 2.3 where Alice (A) and Bob (B) want to exchange their packets (a fixed length stream of bits) via the relay node R. By the traditional routing approach (figure 2.3a), A sends the packet to R which forwards it to B and B sends the packet to R which forwards it to A. Overall 4 transmissions are required for the exchange. Now consider the network coding approach (figure 2.3b). Both A and B send their packets to R. The node R instead of forwarding each packet to the respective destination, XORs the two packets and broadcasts the XOR-ed version of the packets to both A and B. When A and B received the XOR-ed packet, they XOR the received packet again with their own packet which was sent, in order to obtain the packet of each other. This process takes 3 transmissions, thereby increasing the overall throughput.

As can be seen, the concept of Network Coding entails intelligent mixing of signals at the intermediate node. The mixing of the signals can be done in a variety of ways, however, it was proved in [13] that linear encoding and decoding is sufficient to achieve the capacity of a network for multicasting. Network Coding has been practically implemented with testbeds at the MAC layer using the architecture COPE proposed in [14]. This architecture was one of the first and widely accepted attempt to implement network coding at the MAC layer.

The advantages provided by network coding range from increasing the throughput, increasing the robustness to packet losses and link failures, reducing complexity as compared to routing in complex networks, providing for more secure communication [15]. The advantage of network coding can be visible at different levels depending on the network considered.

## 2.2.2 Physical Layer Network Coding with Compute and Forward

The concept of mixing signals at MAC layer led to the birth of Network Coding. Soon enough, the idea was extended to the physical layer with the work in 2006 from Zhang et. al. [16] and a myriad of research work was done in physical layer network coding (PNC). The basic idea of physical layer network coding is to exploit the network coding operation that occurs naturally when electromagnetic waves (EM) are superimposed on one another. The PNC is shown in figure. 2.3c.

The idea of PNC stems from two key observations: firstly, the relay node, in NC framework, only requires a combination (like XOR) of the bits and secondly, the EM waves naturally added up on the wireless channel. In PNC framework, the two nodes A and B transmit the packets simultaneously to the node R. The node observes a superposition of the two waves, and it decodes an XORed packet from the observed signal. This packet is then broadcasted to both the nodes, thereby taking 2 time slots in overall transmission. The key issue in case of PNC is how the relay R can deduce the XORed packet from the superimposed EM waves. This issue is solved by the use of PNC mapping schemes which accurately map the observed superposition to the XOR (or other possible combination) as proposed in [17].

However, the PNC mappings for the superimposed EM waves which have been faded by the physical wireless channel are more complicated. This is because physical layer signal faces the random fading based on the channel conditions and added noise at the receiver, which should be constructively dealt with in order to extract a linear combination like XOR of the bit level original signals. To deal with this issue, in [18], Nazer and Gastpar proposed a novel strategy of generalized relaying, called compute and forward (CF), which enables the relays in any Gaussian wireless network to decode linear equations of the transmitted symbols with integer coefficients, using the noisy linear combinations provided by the channel. The linear equations are transmitted to the destination, and upon receiving sufficient linear equations, the destination can decode the desired symbols. The key point in this strategy is the use of nested lattice codes for encoding the original messages. Nested lattice codes satisfy the property that a linear combination of codewords gives another codeword. Further, information theoretical tools are used in [18] to obtain the achievable rate regions.



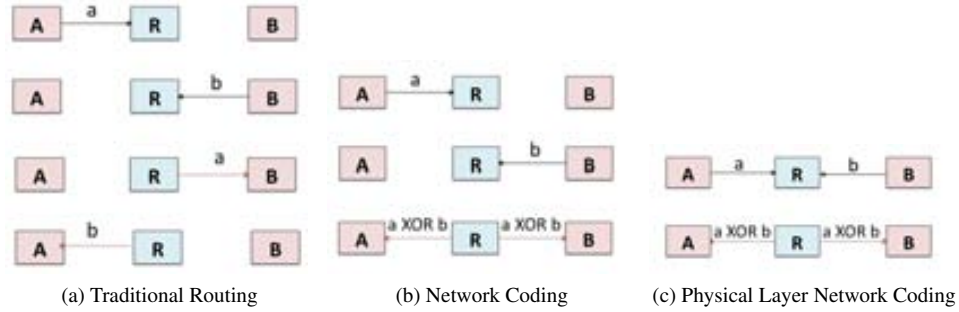


Figure 2.3: Relaying schemes in two way wireless relay channel

An algebraic approach to implement CF has been introduced in [19]. In this work, the authors have related the approach introduced by Nazer and Gastpar to the fundamental theorem of finitely generated modules over the principal ideal domain (PID). Consequently, the isomorphism between the message space and the physical signal space is identified using module theory. Furthermore, the authors identify the lattices and lattice partitions, which can be utilized to implement CF in finite dimensions. The authors show that the union bound estimate for the probability of decoding a linear equation at the relay node is limited by the gap between the channel gain and the integer approximation.

In [20], Niesen and Whiting have analyzed the asymptotic behavior of CF and pointed to a fundamental limitation. They have shown that the degrees of freedom (DoF) achievable by the lattice-based implementation of CF is at most two, when the channel gains are irrational. This is due to the gap between the natural channel gains and the integer coefficient of the linear combination in coherence to the observation in [19]. Further, it is proved that this limitation is not inherent to the fundamental concept of CF, but is due to the lattice-based implementation of CF. In addition, it has also been shown in [20] that the maximum DoF can be achieved by CF when channel state information (CSI) is available at the transmitter.

Another recent work in [21] studies the practical aspects of CF. Specifically, the decoding techniques are studied, and it is shown that the additional noise created due to the non-integer channel coefficients makes the effective noise non-Gaussian and increases the complexity of Maximum likelihood decoding in CF.

The study of CF has attracted the attention of the research community, and a wide body of work is available in the literature. Interested readers can refer to an extensive survey of the recent developments in CF in [22]. Different implementations of CF have also been proposed. The selection of integer coefficients of linear combination at the relay has been studied in [23], where the authors provide a scheme to choose sub-optimal integer coefficients without using CSI at the relay. Further, in [24, 25], precoding-postcoding techniques have been studied in order to minimize the effects of the non-integer channel penalty in some scenarios, particularly, using the spatial orientation of signals.

### 2.2.3 Game Theoretical Tools

The previous two techniques described are primarily used for enhancement of network performance in this thesis. We will now describe the preliminaries of the analytical tool of game theory which we will use in the thesis to assess the realistic performance of the systems taking into account not only the technical aspects of networking but also the competitive and perceptive aspects of the nodes involved in the network. Game theory is a description of strategic interaction, such that the behavior of entities can be mathematically captured particularly in situations in which individual's success in making choices depends on choices of others.

Strategy	Defect	Cooperate
Defect	(5,5)	(15,0)
Cooperate	(0,15)	(1,1)

Table 2.1: Prisonner's Dilemma

Game theory is an extensive tool and there are several books and other literature which has been developed since the first work in the field by John Von Neumann and Oskar Morganstern in 1944 called "Theory of Games and Economic behavior". Interested readers can refer [26, 27] for an overview of Game theory and also its implementation in wireless networks. We will now describe some basic definitions and examples which are useful for game theoretical analysis.

**Definition 1.** A non-cooperative static game is defined using three elements given by:

- Player set  $\mathcal{P}$ : The player is an individual or a group of individuals making a decision in a game which have conflicting interests and they play the game to maximize their interests. The set of  $n$  players as  $\mathcal{P} = 1, 2, 3, \dots, n$ .
- Action set  $\mathcal{A}$ : The action set defines a set of moves or actions a player will follow in a given game. An action set must be complete, defining an action in every contingency, including those that may not be attainable. For each player  $i \in \mathcal{P}$ , the set of possible actions that player  $i$  can take is given by  $A_i = \{1, 2, \dots, n\}$ , and we let  $\mathcal{A} = \{A_1 \times A_2 \dots A_n\}$  denote the space of all action profiles.
- Payoff/Utility  $\mathcal{U}$ : Utility is described as the utility function which attains a particular value when operated with a certain action. Hence it describes which action will the player prefer when she is faced with a decision making situation. For each player  $i \in \mathcal{P}$ ,  $u_i : \mathcal{A} \rightarrow \mathcal{U}$  denote player  $i$ 's utility function.

Together, these three components describe a game as a tuple given by  $\mathcal{G} = \{\mathcal{P}, \mathcal{A}, \mathcal{U}\}$ . We now define an important equilibrium concept in game theory called Nash Equilibrium.

**Definition 2.** Nash equilibrium is a set of actions, one for each player, such that no player has incentive to unilaterally changing its action. So Nash equilibrium is a stable operating point because no user has incentive to change. More formally, a Nash Equilibrium is an action profile  $a_i \in \mathcal{A}_i$  such that for all  $a_i \in \mathcal{A}_i$ ,  $u_i(a_i, a_{-i}) \geq u_i(a_i', a_{-i})$ .

We explain the concept of Nash Equilibrium using the famous Prisonner's Dilemma example.

**Example 3.** Prisonner's dilemma

In this game, two suspects are arrested by the police. The police have insufficient evidence for a conviction, and having separated both prisoners, visit each of them to offer the same deal. If one testifies (defects from the other) for the prosecution against the other and the other remains silent (cooperates with the other), the betrayer goes free and the silent accomplice receives the full 15 - year sentence. If both remain silent, both prisoners are sentenced to 1 year in jail for a minor charge. If each betrays the other, each receives a five - year sentence. Each prisoner must choose to betray the other or to remain silent. Each one is assured that the other would not know about the betrayal before the end of the investigation. How should the prisoners act? Now we denote this situation using a payoff matrix table (2.1) usually used in game theory. The rows represent the action for player 1 and columns represent the action for player 2. The first entry denotes the row player choice and second entry denotes the column player choice. The less the number of years of prison the more is the payoff number.

Clearly, here the best equilibrium for both the players should be (cooperate, cooperate). However, according to Nash equilibrium both players would choose to play defect and equilibrium occurs at (defect, defect)

because in case if any of the player changes its strategy believing other will not change, it will receive a worse payoff so the player would rather prefer to defect than cooperate. Hence it clearly indicates that there should be some modifications in this result which at least help us to arrive at an optimum equilibrium with at least some probability.

We will now describe another optimality criteria which is used to characterize the outcomes of the games and determine their efficiency.

**Definition 4.** Pareto Optimality: An action set  $A^* = (a_1, a_2, \dots, a_n)$  is said to be pareto dominant if there is another action set  $A$  such that

- no player gets a worse payoff with  $A^*$  than  $A$ . i.e.,  $u_i(A^*) \geq u_i(A)$  for all  $i$
- at least one player gets a better payoff with  $A^*$  than with  $A$ , i.e.,  $u_i(A^*) > u_i(A)$

The action profile  $A^*$  is pareto optimal if there is no other action profile which pareto dominates it.

It can be observed that in case of Prisoner's dilemma, although Nash Equilibrium is (defect,defect), the pareto optimal profile is (cooperate,cooperate). In general, there are number of scenarios, in which the nash equilibrium is obtained out of competition among the players, but it is not pareto optimal and there is a possibility that one or more players can obtain higher payoff without reducing other player's payoff.

Game theoretical tools have already been used to study various aspects of transmission over wireless networks [28, 27]. To be more precise, the game theoretical framework has been applied at physical layer to design power allocation games [29, 30], and at application layer to design rate allocation games [31, 32]. Further, different heterogeneous games such as network topology selection games [33], pricing games between service providers and users for network congestion control [34] have been widely studied. Moreover, recent works also study games which are played repeatedly among the same players, at both physical [35] and application layer [36]. In such cases, the previous outcomes of the games, form an additional information using which the players decide their actions.

## 2.3 Performance Metrics

We now describe the main performance metrics that we will consider in this thesis. However, for in-depth assessment and analysis, we have considered other metrics as well, which will be described within the particular sections in the thesis.

### 2.3.1 Offered Throughput

The primary basis of comparison in our contributions is the offered throughput. Throughput is basically the successful message delivery over a communication channel. It is usually measured in bits per second (bps) or packets per second (pps). The throughput could be also be defined as system throughput (which adds up the successful data rates of all the terminals). We use this performance metric in order to quantize the increase in the communication with use of a proposed scheme.

### 2.3.2 Offered Quality of Experience

The throughput forms an absolute metric which can be used for any type of application. Besides throughput, another performance metric that we will use is more precisely connected with the application used. The metric is called Quality of Experience (QoE). The quality of experience is a subjective measure of a user's experiences with a service ( like web browsing, phone call or video conference). It measures the user's

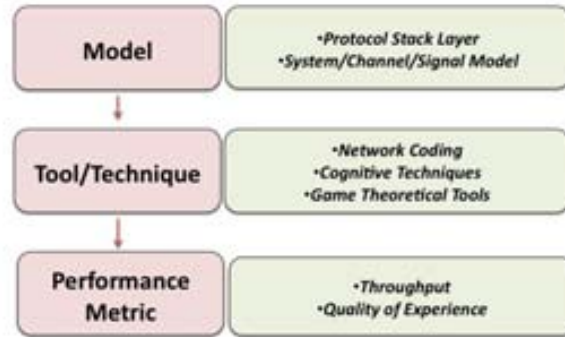


Figure 2.4: Unified Methodology devised to obtain the networking contributions in the thesis

perception of the service and his/her satisfaction with it. QoE is an important metric from the end user's perspective whereas throughput is an important metric from the service provider's perspective. This is because, a network might have a high overall throughput, but application wise, the required packets might not be available at individual user leading to degraded service experience. The mathematical quantization of QoE depends on the application under study. In this thesis, we will be considering video applications due to their increasing demands. In such case, we will define QoE as the probability that the video application will run without freezing. When a video is viewed by the user, a higher QoE will relate to continuous video playing without any freezing whereas a lower QoE signifies that the user is experiencing non-continuous video which degrades its experience.

## 2.4 Overall Methodology to networking solutions

Our overall methodology adopted in this dissertation is described in figure. 2.4. We have classified our contributions on the basis of the layer of the protocol stack at which they are implemented. Our approach begins with firstly development of an appropriate model to assess the networking limitations in the existing system. This constitutes building of a concrete wireless relay network based on the layer of the protocol stack that we are dealing with. The next step is to identify the appropriate tools and techniques which can be implemented in order to enhance the system performance. Lastly, we will assess the overall benefits attained using these tools and techniques will be measured using the chosen performance metrics like throughput , QoE etc.

## Chapter 3

# Networking at Physical Layer

### 3.1 Introduction

The physical layer of the communications protocol stack plays a key role in determining the overall system performance. The networking techniques to optimize the performance at this layer are applied at the level of electromagnetic waves and are hence quite complicated but highly effective in obtaining overall performance variations. It is therefore imperative that we begin our study towards the contributions to different networking techniques with the physical layer.

#### 3.1.1 Objective and technique in overall framework

**Objective:** *Maximizing throughput using Physical layer Network Coding*

Our objective is to develop framework to maximize the throughput of the wireless relay network. Such a network is limited by the interference of co-transmitting nodes with the help of interference-embracing technique namely physical layer network coding. The basic motivation to pursue this technique is that it provides an elaborate insight into the potential of physical layer to improve the throughput of the current wireless networks. It embraces the natural structure of the wireless networks and increases the overall throughput, thereby promising to be an exceptionally productive area for future generation wireless networks.

**Technique:** *Physical layer network coding*

We use physical layer network coding with the help of Compute and forward framework introduced in [18]. This scheme enables the relays in any Gaussian wireless network to decode linear equations of the transmitted symbols with integer coefficients, using the noisy linear combinations provided by the channel. The linear equations are transmitted to the destination, and upon receiving sufficient linear equations, the destination can decode the desired symbols. The key point in this strategy is the use of nested lattice codes for encoding the original messages. Nested lattice codes satisfy the property that a linear combination of codewords gives another codeword.

#### Problem Statement

As discussed in Chapter 2, the recently developed Compute and Forward (CF) strategy to implement PNC is highly promising, particularly due to its generalized approach which is valid for any gaussian network. Motivated by the benefits of CF in order to maximize the throughput, in this chapter, we aim to contribute towards the networking techniques using CF by removing the decoding limitations of the previous proposals,

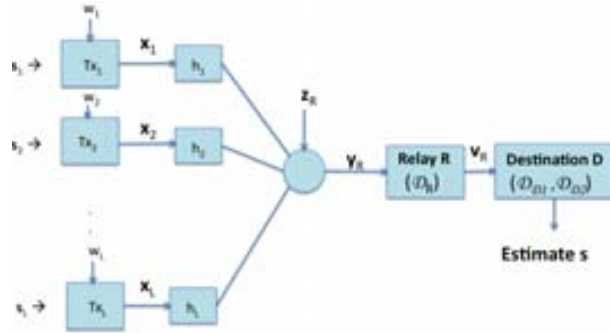


Figure 3.1: System Model

focusing on the physical layer. Particularly, the existing limitation in performance of CF is two fold. Firstly, the approximation of the channel by an integer contributing to additional self-noise apart from receiver noise at the physical layer. Secondly, the absence of full-rank matrix from the integral approximation of channel at the decoder. For the second problem, it has been shown that with some compromise on the rate of computation, the decoder can obtain linear combination of the signals with full rank integral approximation with high probability [18].

To tackle the first problem, in this chapter, we propose a precoder, which removes the self-noise completely and, therefore, improves the performance of CF. Further, in most of the existent works, the performance of CF is studied at the relay node, assuming the transmission from relay-to-destination to be perfect point-to-point transmission. We consider a noisy transmission from relay-to-destination and analyze the end-to-end performance of CF.

### 3.1.2 Outline of Chapter

This chapter is organized as follows: Section II describes the system model and the relevant assumptions. In Section III, we present the source to relay transmission. To this end, we propose the design of an integer forcing precoder. We also present the corresponding decoding process at the relay. Section IV describes the relay-to-destination transmission modeled as a point-to-point transmission. The theoretical performance analysis of our proposed scheme is presented in Section V. The numerical results are presented in Section VI. The conclusions are presented in Section VII.

## 3.2 System Model

We will consider an end-to-end system model for Compute and Forward as shown in figure 3.1. There are  $L$  sources and a single destination node  $D$ . The  $L$  sources communicate with the destination via a single relay node  $R$ . For proposed precoding based CF, transmission occurs in two phases: phase I during which the sources transmit simultaneously to  $R$  followed by phase II during which  $R$  transmits to  $D$ .

### 3.2.1 Phase I: Source to Relay Multiple Access Channel

Let  $\mathbf{s}_l \in \Lambda$  be the message vector to be transmitted by  $l$ -th source ( $l = 1, \dots, L$ ) where  $\Lambda \subset \mathcal{L}$  and  $\mathcal{L} = \{\lambda = \mathbf{G}\mathbf{v} \mid \mathbf{G} \in \mathbb{R}^{n \times n}, \mathbf{v} \in \mathbb{Z}^n\}$  is an  $n$ -dimensional lattice with generator  $\mathbf{G}$  having components over time. The

message vector satisfies the average power constraint

$$\frac{1}{n}E[\|\mathbf{s}_l\|^2] = 1 \quad (3.1)$$

Each transmitter precodes the message signal with precoder  $w_l \in \mathbb{R}$  to obtain the precoded signal  $\mathbf{x}_l \in \mathbb{R}^n$  as

$$\mathbf{x}_l = w_l \mathbf{s}_l \quad (3.2)$$

The average power of the precoder satisfies the constraint  $E[|w_l|^2] \leq \gamma_w$  where  $\gamma_w > 0$ . The precoded signal satisfies the power constraint of

$$\frac{1}{n}E[\|\mathbf{x}_l\|^2] \leq \gamma_w \quad (3.3)$$

If  $\gamma_w = 1$ , the original signal power is preserved after precoding.

The channel output  $\mathbf{y}_R \in \mathbb{R}^n$  observed at the relay is given by

$$\mathbf{y}_R = \sum_{l=1}^L h_l \mathbf{x}_l + \mathbf{z}_R \quad (3.4)$$

where  $h_l \in \mathbb{R}$  is the channel coefficient between transmitter  $l$  and the relay node,  $\mathbf{z}_R$  is i.i.d Gaussian noise vector given by  $\mathbf{z}_R = [z_{R1} \ z_{R2} \ \dots \ z_{Rn}]^T$ ,  $z_{Ri} \sim \mathcal{N}(0, \sigma_1^2)$ . The channel coefficient vector is given by  $\mathbf{h} = [h_1 \ \dots \ h_L]$  and the channel is assumed to be quasi-static. The aim of the relay is to compute a linear combination of source signals given by

$$\mathbf{v}_R = \sum_{l=1}^L a_l \mathbf{s}_l \quad (3.5)$$

where  $a_l \in \mathbb{Z}$  are integer coefficients chosen on the basis of  $h_l$ . The linear coefficient vector is given by  $\mathbf{a} = [a_1 \ \dots \ a_L]$ . The relay is free to choose these linear coefficients. Since the linear combination  $\mathbf{v}_R$  is computed over  $\mathbb{Z}$  while the channel output is obtained over  $\mathbb{R}$ , the linear coefficients  $a_l$  are chosen in a way to efficiently exploit this channel output for this computation. Here,  $\mathbf{v}_R \in \Lambda'$  where

$$\Lambda' = \{\lambda' \mid \lambda' = a_1 \lambda_1 + a_2 \lambda_2 + \dots + a_L \lambda_L, a_i \in \mathbb{Z}, \lambda_i \in \Lambda\} \quad (3.6)$$

Hence,  $\Lambda'$  is also  $n$ -dimensional lattice. The estimate of  $\mathbf{v}_R$  obtained at the relay using the decoder  $\mathcal{D}_R : \mathbb{R}^n \rightarrow \Lambda'$  is given by

$$\hat{\mathbf{v}}_R = \mathcal{D}_R(\mathbf{y}_R) \quad (3.7)$$

It is assumed that each transmitter has the CSI of its own channel only, i.e., the channel between the transmitter itself and the relay.

### 3.2.2 Phase II: relay-to-destination point-to-point channel

In phase II, the relay transmits the linear combination estimated in (3.7) to the destination  $D$ . The channel output observed at the destination  $\mathbf{y}_D \in \mathbb{R}^n$  is given by

$$\mathbf{y}_D = h_{RD} \hat{\mathbf{v}}_R + \mathbf{z}_D \quad (3.8)$$

where  $h_{RD} \in \mathbb{R}$  is the channel coefficient between the relay node and the destination. We assume that  $h_{RD}$  is uniformly distributed Gaussian variable with mean  $\mu_{h_{RD}}$  and variance  $\sigma_{h_{RD}}^2$  such that  $h_{RD} \sim \mathcal{N}(\mu_{h_{RD}}, \sigma_{h_{RD}}^2)$ .

The additive i.i.d Gaussian noise vector is given by  $\mathbf{z}_D = [z_{D1} \ z_{D2} \ \dots \ z_{Dn}]^T$ ,  $z_{Di} \sim \mathcal{N}(0, \sigma_2^2)$ . Consequently, the estimate of  $\hat{\mathbf{v}}_R$  obtained at the destination using the decoder  $\mathcal{D}_{D1} : \mathbb{R}^n \rightarrow \Lambda'$  is given by

$$\hat{\mathbf{v}}_D = \mathcal{D}_{D1}(\mathbf{y}_D) \quad (3.9)$$

The destination obtains a linear combination of the original source signals at the end of two-phase transmission. In order to decode the original source signals, the destination node collects  $L$  such linear combinations. After  $L$  repeated transmissions of their respective message from the  $L$  sources, the destination obtains  $L$  linear combinations of  $L$  source messages. Using these linear combinations, the destination decodes the original source messages using the decoder

$$\hat{\mathbf{s}} = \mathcal{D}_{D2}(\mathbf{v}) \quad (3.10)$$

where  $\mathbf{v} = [\hat{\mathbf{v}}_D^1 \ \dots \ \hat{\mathbf{v}}_D^L]$  is the matrix containing  $L$  linear combinations obtained at the destination.  $\hat{\mathbf{v}}_D^l$  denotes the  $l$ th linear combination obtained at the destination. The original message matrix estimated at the destination is given by  $\hat{\mathbf{s}} = [\hat{\mathbf{s}}_1 \ \dots \ \hat{\mathbf{s}}_L]$ . Note that in this work, we focussed on real-valued system model for simplicity, however, we assert that the work can be extended to complex-valued system.

### 3.3 Phase I: Proposed Precoding-based Source to Relay Transmission

In this section, we focus upon the transmission phase I from the  $L$  sources to the relay node. Each of the  $L$  sources transmits a lattice point from the same lattice towards the relay nodes. The received signal at the relay is a noisy linear combination of the lattice points with linear coefficients given by the channel gains. As explained in Chapter 2, the concept of CF relies on the lattice property that an integral linear combination of lattice points is another lattice point. However, the received signal at the relay does not necessarily have integral coefficients as the natural channel gains are not necessarily integers. Therefore, the closest lattice point to the received signal may or may not be an integral linear combination of the original lattice points depending on the additive noise and the 'self-noise' contributed by the non-integral nature of the channel. Since additive noise is inherent to the system, the self-noise due to the non-integral nature of the channel should be minimized. In this section, the problem of obtaining an integer linear combination of original signals from the received signal is formulated and subsequently, an integer forcing precoder design is proposed.

#### 3.3.1 Problem Formulation

The aim of a precoding based implementation of CF is to obtain the linear combination of original signals at the relay without incurring the errors due to approximation of channel by integers. To this end, the precoder is used to shift the channel coefficients to the closest possible integer using the CSI at the transmitter. Therefore, consider a channel decomposition as

$$h_l = h_l^z + h_l^{\bar{z}} \quad (3.11)$$

where  $h_l^z \in \mathbb{Z}$  is an integer approximation of  $h_l$  and consequently  $h_l^{\bar{z}} \in \mathbb{R}$ . Using (3.2) and (3.11), we can rewrite (3.4) as,

$$\mathbf{y}_R = \sum_{l=1}^L h_l w_l \mathbf{s}_l + \mathbf{z}_R + \sum_{l=1}^L h_l^z \mathbf{s}_l - \sum_{l=1}^L h_l^{\bar{z}} \mathbf{s}_l$$

which can be simplified as,



$$\mathbf{y}_R = \underbrace{\sum_{l=1}^L h_l^z \mathbf{s}_l}_{\mathbf{v}_{eff}} + \underbrace{\sum_{l=1}^L (h_l w_l - h_l^z) \mathbf{s}_l}_{\mathbf{z}_{eff}} + \mathbf{z}_R \quad (3.12)$$

The relay aims to decode  $\mathbf{v}_{eff}$  from  $\mathbf{y}_R$ . The effective noise,  $\mathbf{z}_{eff}$ , is comprised of the receiver noise ( $\mathbf{z}_R$ ) and the additional “self-noise” arising due to difference between the real channel and its integral approximation. Using the rate of computation derived in [18], we can write the rate of computation of linear combination of original signals as

$$R(\mathbf{h}, \mathbf{w}) = \log^+ \left( \frac{P}{1 + P \sum_{l=1}^L (h_l w_l - h_l^z)^2} \right)$$

where  $P$  is the signal to noise ratio which is  $P = 1/\sigma_1^2$  in this case and  $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_L]$ . Therefore,

$$R(\mathbf{h}, \mathbf{w}) = \log^+ \left( \frac{1}{\sigma_1^2 + \sum_{l=1}^L (h_l w_l - h_l^z)^2} \right)$$

where  $\log^+(x) = \max(0, \log(x))$ . Using the power constraints, the rate can be rewritten as

$$R(\mathbf{h}, \mathbf{w}) = \log^+ \left( \frac{1}{\sigma_1^2 + \|\mathbf{h}\mathbf{W} - \mathbf{h}^z\|^2} \right)$$

where  $\mathbf{h}^z = [h_1^z \ \dots \ h_L^z]$  and  $\mathbf{W} = \text{diag}(\mathbf{w})$ . It is assumed that  $h_l^z \neq 0$ , to ensure linear combination has non-zero coefficient of each of the  $L$  messages, therefore,  $\|\mathbf{h}^z\| \geq 1$ . Hence, we obtain,

$$R(\mathbf{h}, \mathbf{w}) \leq \log^+ \left( \frac{\|\mathbf{h}^z\|^2}{\sigma_1^2 + \|\mathbf{h}\mathbf{W} - \mathbf{h}^z\|^2} \right) \quad (3.13)$$

Hence, the precoder  $w_l$  should be designed such that it is able to minimize  $\mathbf{z}_{eff}$  in (3.12) and maximize the achievable rate in (3.13). The design of such a precoder is proposed in next subsection.

### 3.3.2 Proposed Integer Forcing Precoder

We aim to precode the signals such that a linear combination of original signals can be obtained at the relay at maximum rate. It is well known that Zero Forcing (ZF) precoding is a standard suboptimal approach for precoding which is known to provide a promising trade-off between complexity and performance [37]. The traditional zero-forcing precoder is designed to eliminate the effect of the channel fading while satisfying constraints of power optimally as per certain performance measure.

The precoder required in implementing CF, has the similar design requirements as ZF precoding but with an important difference. In CF, the precoder is required to eliminate only the non-integral part of the channel while retaining the integral part in the precoded signal. In this subsection, we show the design of a naive precoder to implement CF.

A sub-optimal integer forcing precoder (IFP) that maximizes the rate of computation while eliminating the non-integral part of channel is given by the solution to the following optimization problem

$$\max_{\mathbf{h}^z \in \mathbb{Z}^L, \mathbf{h}^z \neq \mathbf{0}} R(\mathbf{h}, \mathbf{w}) \quad (3.14)$$

$$\text{subject to } E[|w_l|^2] \leq \gamma_w \quad (3.15)$$

In order to solve this problem, we relaxed condition (3.15) making it an unconstrained optimization problem. A direct differentiation of  $R(\mathbf{h}, \mathbf{w})$  with respect to  $\mathbf{W}$  gives

$$\mathbf{W}_{opt} = \text{diag}(\mathbf{h}^{-1} \mathbf{h}_{opt}^z)$$

where  $\mathbf{h}_{opt}^z$  is the solution to

$$\max_{\mathbf{h}^z \in \mathbb{Z}^L, \mathbf{h}^z \neq 0} \log^+ \left( \frac{\|\mathbf{h}^z\|^2}{\sigma_1^2} \right) \quad (3.16)$$

$$\text{subject to } E[|h_l^z h_l^{-1}|^2] < \gamma_w \quad \text{for } l = 1 \dots L \quad (3.17)$$

Hence, the integer part of the channel  $\mathbf{h}^z$  should be chosen such that, within the constraint of maximum precoding power  $\gamma_w$ , the computational rate is maximized. A possible solution is given by (rounding off using Babai estimate [38])

$$\mathbf{h}_{opt}^z = [\gamma_w \mathbf{h}] \quad (3.18)$$

where  $[\cdot]$  represents for the closest integer. Without loss of generality, one of the solutions can be taken as  $\mathbf{h}_{opt}^z = [\mathbf{h}]$ . The resultant precoder, which also imposes the constraint  $\mathbf{h}^z \neq 0$ , is given by

$$w_l = \begin{cases} \frac{h_l^z}{h_l}, & |h_l| > 0.5 \\ \frac{1}{h_l}, & |h_l| \leq 0.5 \end{cases} \quad \text{for } l = 1 \dots L \quad (3.19)$$

This solution maximizes the rate (3.13) because  $h_l^z$  is the closest integer to the channel  $h_l$ . Note that here each transmitter precodes the signal by imposing  $h_l^z \neq 0$ , thereby making the probability of the condition  $\mathbf{h}^z = 0$  to occur strictly zero. Since we have assumed that the  $l$ th transmitter is aware of only  $h_l$  channel coefficient, therefore, simply imposing  $h_l^z \neq 0$  ensures that  $\mathbf{h}^z \neq 0$ .

In the next theorem, it is shown that the proposed IFP eliminates the additional self-noise in CF.

**Theorem 5.** *The IFP can completely eliminate the additional self-noise in CF scheme with an upper bounded power penalty for channel with reasonable gains.*

*Proof.* Firstly, to prove the elimination of self-noise by the use of IFP, the received signal in (3.4) is rewritten, using (3.19), as

$$\mathbf{y}_R = \sum_{l=1}^L h_l w_l \mathbf{s}_l + \mathbf{z}_R = \sum_{l=1}^L h_l^z \mathbf{s}_l + \mathbf{z}_R$$

Assuming  $\mathbf{v}_R = \sum_{l=1}^L h_l^z \mathbf{s}_l$ , the received signal is given by

$$\mathbf{y}_R = \mathbf{v}_R + \mathbf{z}_R \quad (3.20)$$

The effective noise is clearly only Gaussian noise and the self-noise is completely eliminated. However, the use of IFP requires an extra power penalty at the transmitter.

Next, to prove that this additional power required is within finite range for any channel, the proposed precoder in (3.19) can be written as

$$w_l = \frac{1}{1+r}$$

where  $r = h_l^z / h_l$ . Since  $h_l^z \in \mathbb{Z}$  and  $h_l^z \in (-0.5, 0.5)$  for  $|h_l| > 0.5$ , therefore, the range of  $r$  is given by  $r \in (-0.5, 0.5)$ . Putting  $r$  in expression of  $w_l$  above, we get the range of  $w_l$  as  $w_l \in (\frac{2}{3}, 2)$ . Consequently, the range of  $w_l^2$  is given by  $w_l^2 \in (\frac{4}{9}, 4)$ . In the limiting condition, the upper bound of average precoder power

$E[|w_l|^2] = \gamma_w$  (using the power limitation from (3)), can be given as

$$\frac{4}{9} < \gamma_w < 4 \quad (3.21)$$

Hence, the precoded signal power using proposed IFP is bounded for channel realization with reasonably good gains. This concludes our proof.  $\square$

*Remark 6.* From theorem 5, it can be seen that the use of IFP to implement CF effectively reduces the channel to a reliable AWGN channel. Therefore, the achievable rate using the above IFP is given by

$$R_{IFP}(\mathbf{h}) \leq \log^+ \left( \frac{\|\mathbf{h}^z\|^2}{\sigma_1^2} \right) \quad (3.22)$$

From (3.2) and (3.21), it is clear that when  $\gamma_w = 1$ , the power consumed by the precoded signal is equal to the power consumed by the original signal under limiting conditions. However, when  $\gamma_w < 1$  (or equivalently,  $h_i^z < h_l$ ), the precoded signal consumes lesser power than the original signal. On the other hand, when  $\gamma_w > 1$ , the precoded signal consumes additional power. If this additional power is not available at the transmitter, the precoding may fail. We study this case in detail under outage formulation Section 3.6.2.

With a finite power penalty, the proposed IFP can convert the signal obtained at the relay into a reliable linear combination of source messages. The decoder at the relay employs the Maximum Likelihood (ML) decoding to obtain a linear combination of source messages as explained in the next subsection.

### 3.3.3 Decoding at the relay

We aim to apply ML decoding on the received signal at the relay. According to Theorem 5, the received signal is given by  $\mathbf{y}_R = \mathbf{v}_R + \mathbf{z}_R$  where  $\mathbf{v}_R = \sum_{l=1}^L h_l^z \mathbf{s}_l$  and  $\mathbf{z}_R \sim \mathcal{N}(0, \sigma_1^2 \mathbf{I}_n)$ . The linear combination of signals  $\mathbf{v}_R \in \Lambda'$  such that

$$\Lambda' = \{\lambda' \mid \lambda' = h_1^z \lambda_1 + h_2^z \lambda_2 + \dots + h_L^z \lambda_L, h_i^z \in \mathbb{Z}, \lambda_i \in \Lambda\} \quad (3.23)$$

By the properties of lattices [39], a linear combination of lattice points gives another lattice point, therefore,  $\Lambda' \subseteq \mathcal{L}$ . The ML decoder  $\mathcal{D}_R(\cdot)$  in (3.7) gives the estimate of  $\mathbf{v}_R$  as

$$\hat{\mathbf{v}}_R = \mathcal{D}_R(\mathbf{y}_R) = \arg \min_{\mathbf{t} \in \Lambda'} \|\mathbf{y}_R - \mathbf{t}\|^2$$

A low complexity sphere decoder [38] can be used to perform ML decoding. Note that the decoder architecture reflects a point-to-point AWGN channel with input  $\mathbf{v}_R$  and additive channel noise. This shows that the complexity of the decoder architecture is equivalent to the decoder complexity of point-to-point channel.

## 3.4 Phase II: Proposed point-to-point based relay to destination transmission

In this section, we describe the phase II of the transmission of proposed precoding based CF scheme between the relay node and the destination node.

### 3.4.1 Problem Formulation

The computation of original signals at the destination is a two-fold problem : firstly, the destination node decodes the linear combination obtained from the relay node and secondly, after obtaining sufficient linear combinations, the destination node decodes the original source signals. Therefore, we formulate the two problems at the decoder and propose the respective solutions using two decoders, namely  $\mathcal{D}_{D1}$  and  $\mathcal{D}_{D2}$  respectively, in subsequent subsections.

The first aim of the destination is to obtain the linear combination estimated at the relay from the noisy signal obtained from the channel. We recall that the received signal at the destination is given by (3.8)

$$\mathbf{y}_D = h_{RD}\hat{\mathbf{v}}_R + \mathbf{z}_D$$

The destination node aims to obtain the linear combination  $\hat{\mathbf{v}}_R$ . To this end, the received signal is equalized using an equalizer  $\alpha$ , and a suitable decoder is used to obtain the linear combination. Therefore, the first decoder in (3.9) is given as

$$\hat{\mathbf{v}}_D = \mathcal{D}_{D1}(\alpha, \mathbf{y}_D)$$

Therefore, the primary problem at the destination is to obtain a decoder suitable for the above decoding.

The next aim of the destination is to obtain the original source signals from  $L$  linear combinations obtained. Therefore, the matrix of the linear combinations after  $L$  transmissions given by  $\mathbf{v} = [\hat{\mathbf{v}}_D^1 \dots \hat{\mathbf{v}}_D^L]^T$  is supplied to the second decoder at the destination which obtains the estimate of original source signal matrix given by  $\mathbf{s} = [s_1 \dots s_L]^T$  such that

$$\hat{\mathbf{s}} = \mathcal{D}_{D2}(\mathbf{v})$$

The second problem at the decoder is to obtain the design of a decoder to obtain the original source signals.

### 3.4.2 Decoding linear combination at the destination

In order to obtain the decoder at the destination to obtain the linear combination of original signals from the noisy channel output, we propose to use a zero forcing equalizer. Therefore, for this decoding we set  $\alpha$  such that  $\alpha = \frac{1}{h_{RD}}$  which leads to the equalized signal being

$$\alpha \mathbf{y}_D = \hat{\mathbf{v}}_R + \frac{\mathbf{z}_D}{h_{RD}} \quad (3.24)$$

Consequently, an ML decoder is used to obtain  $\hat{\mathbf{v}}_R$ . A low complexity ML decoder, namely sphere decoder, can be used to perform the decoding operation. The ML decoder is designed identical to the sphere decoder used at the relay as the lattice to which the target decoded signal  $\mathbf{v}_R$  belongs is the same at both the relay and the destination node. Therefore,

$$\hat{\mathbf{v}}_D = \mathcal{D}_{D1}(\alpha, \mathbf{y}_D) = \arg \min_{\mathbf{t} \in \Lambda'} \|\alpha \mathbf{y}_D - \mathbf{t}\|^2$$

where  $\Lambda'$  is given by (3.23). Note that the decoding is without error as long as  $\frac{\mathbf{z}_D}{h_{RD}}$  in (3.24) is within the Voronoi Region of the lattice  $\Lambda'$ . We explore the probability of error in detection of linear combination in section 3.5.

### 3.4.3 Decoding original signals at the destination

In order to detect the original source signals, the destination node requires to collect  $L$  independent linear combinations. After  $L$  transmissions, the destination obtains  $L$  linear combinations of  $L$  messages from

$L$  users. Using these linear combinations, the destination decodes the original source messages using the decoder given in (3.10). In order to design the decoder, we rewrite the  $L$  linear combinations as (for each element of the linear combination)

$$\underbrace{\begin{bmatrix} \hat{v}_D^1 \\ \vdots \\ \hat{v}_D^L \end{bmatrix}}_{\mathbf{v}} = \underbrace{\begin{bmatrix} a_1^1 & \dots & a_L^1 \\ \vdots & \dots & \vdots \\ a_1^L & \dots & a_L^L \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} s_1 \\ \vdots \\ s_L \end{bmatrix}}_{\mathbf{s}}$$

Here,  $\hat{v}_D^l$  denotes the  $l$ -th linear combination element estimated at the destination. Also,  $a_k^l$  denotes the linear coefficient of the  $k$ th source signal in the  $l$ th linear combination. The matrix  $\mathbf{A}$  contains the integer linear coefficients of the linear combinations. If  $\mathbf{A}$  is full rank,  $\mathbf{s}$  is obtained by inverting  $\mathbf{A}$  such that

$$\hat{\mathbf{s}} = \mathcal{D}_{D2}(\mathbf{v}) = \mathbf{A}^{-1}\mathbf{v}$$

where  $\hat{\mathbf{s}}$  is the estimate of the original source signals. It is assumed that, since the channel conditions are highly variant (typically in wireless networks like vehicular networks), the channel fades independently in each transmission. Consequently, the matrix formed by the rounded channel coefficients is full rank with high probability. As proposed in our scheme, the transmitter chooses the integer coefficients closest to the channel coefficients, therefore,  $\mathbf{A}$  is full rank with high probability.

## 3.5 Performance Analysis

In this section, we analyze the proposed precoding based CF implementation under the light of probability of error. Note that we will make use of the geometrical properties of lattices in obtaining closed form expressions of probability of error [40][41].

The end-to-end probability of error of the precoded CF is culminated by two error probabilities: (i) the probability of error from source to relay node (ii) the probability of error from relay-to-destination node. In this subsection we obtain two probabilities separately and then formulate the end-to-end probability of error expression.

### 3.5.1 Probability of error from source to relay node

According to Theorem 5, the received signal at the relay is the linear combination of original signals with additive Gaussian noise. Therefore, an error occurs if the noise vector is outside the fundamental Voronoi Region [40] of the lattice  $\Lambda'$  in (3.23), denoted by  $\mathcal{V}(\Lambda')$ . The probability of error in decoding a linear equation in  $n$ -dimensional lattice at the relay is given by

$$P_1(\mathbf{h}, n) = \Pr(\|\mathbf{z}_R\| \notin \mathcal{V}(\Lambda')) \quad (3.25)$$

The vector  $\mathbf{z}_R$  has Gaussian distribution in each dimension with zero mean and variance  $\sigma_1^2$ . It is well known that the above probability expression depends on the shape of the Voronoi Region. The bounds of such probability of error can be found in [42].

For simplicity, we now consider a subset of an integer lattice namely,  $\mathcal{L} = \mathbb{Z}^n$  to draw the original signals  $s_l$ . This is because the Voronoi Region for the integer lattice is a hypercube and hence it can be used to obtain an exact probability of error expression. The following theorem gives this probability.

**Theorem 7.** *The probability of error in decoding a linear equation of signals drawn from  $\mathcal{L} = \mathbb{Z}^n$  using IFP is given by*

$$P_1(\mathbf{h}, n) = 1 - \left( \operatorname{erf} \left( \frac{c}{2\sqrt{2}\sigma_1} \right) \right)^n$$

where  $c = \gcd(h_1^z, h_2^z, \dots, h_L^z)$

*Proof.* We firstly use induction to prove that the Voronoi Region of  $\Lambda' = \{\lambda' \mid \lambda' = \sum_{i=1}^L h_i^z \lambda_i, h_i^z \in \mathbb{Z}, \lambda_i \in \Lambda\}$  is a hypercube of side  $c$ . Let  $L = 2$  and  $n = 1$ . The Bezout's lemma [43] states that for any two (non-zero) integers  $p$  and  $q$ , there exist two integers  $u$  and  $v$  such that

$$pu + qv = d$$

where  $d$  is the common divisor of  $p$  and  $q$ . In addition, if  $d > 0$  is the greatest common divisor of  $(p, q)$ , it is the smallest positive integer satisfying this equation for any  $(u, v)$  integer pair. Therefore, using Bezout's lemma, any point in  $\Lambda'$  can be written as

$$\lambda' = b_1 \lambda_1 + b_2 \lambda_2 = d\mathbb{Z}$$

where  $d = \gcd(b_1, b_2)$ . Similarly, this result can be extended to  $L = L + 1$  by induction. Since this result is independent of the integers  $\lambda_i$ , therefore, it can be extended to  $n$ -dimensional lattices. Using this result, we can write

$$\Lambda' = c\mathbb{Z}^n$$

where  $c = \gcd(h_1^z, h_2^z, \dots, h_L^z)$ . Hence the Voronoi Region of  $\Lambda'$  is a hypercube of side  $c$ .

To prove the second part of the theorem, we rewrite (3.25) as,

$$\begin{aligned} P_1(\mathbf{h}, n) &= 1 - \Pr(\|\mathbf{z}\| \in \mathcal{V}(\Lambda')) \\ &= 1 - \left( \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-c/2}^{c/2} e^{-\frac{u^2}{2\sigma^2}} du \right)^n \\ P_1(\mathbf{h}, n) &= 1 - \left( \operatorname{erf} \left( \frac{c}{2\sqrt{2}\sigma_1} \right) \right)^n \end{aligned} \quad (3.26)$$

where  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ . This concludes our proof.  $\square$

*Remark 8.* For the special case of  $n = 1$ , (3.26) reduces to

$$P_1(\mathbf{h}, 1) = \operatorname{erfc} \left( \frac{c}{2\sqrt{2}\sigma_1} \right) \quad (3.27)$$

where  $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ .

Note that in the proposed scheme, the effective behavior of the channel is reduced to a point-to-point AWGN channel with no channel fading. This makes its characterization possible. In the previous formulations of CF, the effective noise is Gaussian noise along with the self-noise, which results in non-Gaussian distribution of total effective noise [21]. Hence, the characterization in existent formulations has been done by providing limits of error probability and not the accurate value. But with IFP, CF can be implemented without the limitation of any self-noise adding to the effective noise. However, the penalty that is paid is

the extra power required at the transmitter. We analyze the effect of this penalty on performance in terms of outage probability in the next section.

### 3.5.2 Probability of error from relay-to-destination node

The signal received at the destination, given by (3.8), is normalized to the channel coefficient resulting in

$$\frac{\mathbf{y}_D}{h_{RD}} = \hat{\mathbf{v}}_R + \frac{\mathbf{z}_D}{h_{RD}} \quad (3.28)$$

The following theorem gives the probability of error at the destination node.

**Theorem 9.** *The probability of error in decoding the linear combination of signals sent from the relay to the destination for  $\mathcal{L} = \mathbb{Z}^n$  is given by*

$$P_2(d, n) = 1 - \left( \frac{2}{\pi} \left( \arctan \left( \frac{1}{2d} \right) \right) \right)^n$$

where  $d = \frac{\sigma_2}{\sigma_{h_{RD}}}$  such that  $h_{RD} \sim \mathcal{N}(0, \sigma_{h_{RD}}^2)$  and the variance of added noise is  $\sigma_2^2$ .

*Proof.* Using (3.28), the probability of error at the destination is given by

$$P_2(d, n) = Pr \left( \left\| \frac{\mathbf{z}_D}{h_{RD}} \right\| \notin \mathcal{V}(\Lambda') \right)$$

where  $\Lambda'$  is defined in (3.23). The distribution of  $\mathbf{z}_D/h_{RD}$  is the quotient of two normal distributions with zero mean. Such a distribution is given by the Cauchy's distribution [44] with zero mean. Let us define ratio  $d$  as  $d = \frac{\sigma_2}{\sigma_{h_{RD}}}$ . The probability density function of Cauchy's distribution takes the form of

$$\phi(x) = \frac{d}{\pi(x^2 + d^2)}$$

An error occurs in the decoding if the noise falls outside the Voronoi Region. Therefore, we integrate the distribution of effective noise from  $(-1/2, 1/2)$  to obtain the probability of error-free transmission and consequently, subtract it from 1 to find the probability of error. Hence, the resulting expression is given by

$$P_2(d, n) = 1 - \left( \int_{-1/2}^{1/2} \phi(x) dx \right)^n$$

The above integration can be written as

$$P_2(d, n) = 1 - \left( \int_{-1/2}^{1/2} \frac{d}{\pi(x^2 + d^2)} dx \right)^n$$

Using  $\int \frac{1}{x^2 + a^2} dx = \frac{1}{a} \arctan \frac{x}{a} + C$ , we get

$$P_2(d, n) = 1 - \left( \frac{d}{\pi} \left( \frac{1}{d} \arctan \left( \frac{1}{2d} \right) - \frac{1}{d} \arctan \left( -\frac{1}{2d} \right) \right) \right)^n$$

Since  $\arctan(-x) = -\arctan(x)$ , therefore,

$$P_2(d, n) = 1 - \left( \frac{2}{\pi} \left( \arctan \left( \frac{1}{2d} \right) \right) \right)^n \quad (3.29)$$

where  $d = \frac{\sigma_2}{\sigma_{hRD}}$ . □

For the special case of  $n = 1$ , the above reduces to

$$P_2(d, 1) = 1 - \frac{2}{\pi} \left( \arctan \left( \frac{1}{2d} \right) \right)$$

We now give the overall probability of error.

### 3.5.3 Overall end-to-end probability of error

We now aim to evaluate the overall probability of error for CF. An error occurs in CF transmission, if the linear combination of source signals are not correctly obtained at the relay and/or the linear combination transmitted from the relay is not correctly obtained at the destination. Hence both the phases of transmissions are required to be free from error. Further, the entire scheme requires  $L$  transmissions because  $L$  linear combinations are required at the destination to obtain the original signals. Hence, the CF transmission is successful if and only if  $L$  independent linear combinations are obtained at the destination without any error. Since we assumed that the channel is highly variant in time, hence it is implicit that the  $L$  linear combinations will be independent. Therefore, the overall probability of error for CF scheme is given by (for  $n = 1$  dimension)

$$P_e = 1 - \prod_{l=1}^L (1 - P_1^l)(1 - P_2^l)$$

where the superscript indicates the  $l$ -th transmission. Therefore, inserting the expressions for  $P_1$  and  $P_2$  from (3.26) and (3.29) respectively, we obtain,

$$P_e = 1 - \left( \frac{2}{\pi} \left( \arctan \left( \frac{1}{2d} \right) \right) \right)^L \prod_{l=1}^L \left( 1 - \operatorname{erfc} \left( \frac{c^l}{2\sqrt{2}\sigma_1} \right) \right) \quad (3.30)$$

As in previous cases, the superscript indicates the value of the variable in the  $l$ th transmission. The above expression takes into account the necessity of all the linear combinations to be simultaneously error-free for error-free transmission. A simplified upper bound to the above expression can be given by (assuming fixed gcd of 1 for all channel realizations)

$$P_e^u = 1 - \left( \frac{2}{\pi} \left( \arctan \left( \frac{1}{2d} \right) \right) \right)^L \left( 1 - \operatorname{erfc} \left( \frac{1}{2\sqrt{2}\sigma_1} \right) \right)^L \quad (3.31)$$

In the next section, we will present the numerical results to demonstrate the performance of the proposed scheme.



## 3.6 Numerical Results

In this section, we demonstrate the performance of the proposed scheme with the help of numerical simulations. We focus on mainly three performance metrics: (i) Probability of error (ii) Achievable Rates (iii) Outage probability.

### 3.6.1 Probability of error

In order to analyze the probability of error performance for the proposed IFP, we perform the simulations under two parts. In the first part, we study the stand-alone performance of the IFP precoder by observing the probability of error behavior at the relay node. Further, in second part, we analyze the effect of the implementation of CF with IFP on the overall end-to-end system from source to destination.

#### 3.6.1.1 Probability of error at the relay

Two sources are considered transmitting the signals ( $s_1$  and  $s_2$ ) drawn from one-dimensional integer lattice, i.e  $n = 1$  and  $\mathcal{L} = \mathbb{Z}$ . There is one relay which decodes the linear combination of these signals. The performance in terms of probability of error in decoding the linear combination of original signals at the relay is simulated. The additive noise at the relay node has a Gaussian distribution.

In order to compare with the existent schemes, the baseline lattice network coding (LNC) scheme, which is used to practically implement CF in [19] is considered. To make a fair comparison, we have adjusted the power of the transmitters in the schemes with and without precoding such that the total power available at transmitters in both schemes is equal. We consider precisely the same channel as considered in [21], for sake of comparison, given by  $\mathbf{h} = [ -1.274 \quad 0.602 ]$  however, the results are not sensitive to choice of channel gain vector.

We observe in Figure 3.2, that the probability of error by the proposed IFP based scheme has a gain of up to 2 dB over the existent baseline LNC.

For comparison, we also plot the probability of error which can be achieved when the channel gain vector is integral. In this case, the additional noise is the only source of error at the relay. Therefore, this indicates a lower bound of the probability of error for CF. Note that, with our proposed precoding, the only source of error at the relay is the additional noise. However, as shown in Figure 3.2, there is a gap between the proposed scheme and the lower bound. This gap is due to the penalty of extra power required for precoding. More precisely, the proposed scheme requires more power at the transmitter to achieve the same probability of error as the CF with integral channel.

#### 3.6.1.2 Probability of error at the destination

In order to analyze the overall performance of using IFP with CF system, we consider the extension of above model to the destination node. The aim of the destination is to obtain  $L$  original source signals. Hence, we consider  $L$  transmissions to study the performance of error at the destination. After  $L$  transmissions from source to relay, the  $L$  transmissions from relay-to-destination are orthogonalized in time. The channel from relay-to-destination is random real valued channel. Figure 3.3 shows the probability of error of CF implemented with IFP. To make a comparison with the state-of-the-art, we plot the error performance of the system when Time Division Multiplexing (TDM) is used to transmit the signals from source to destination via relay. In case of TDM, one source transmits at one time instant to the relay. The relay decodes the signals and sends it to the destination. In the next time instant, the relay sends the decoded signal to the destination and the destination decodes the original signal from the first source. After  $L + 1$  time slots, all the  $L$  signals are received at the destination. Figure 3.3 shows that CF with IFP shows a gain of approximately 5dB at an error of 1% over TDM. This gain is basically due to multiple copies of the same signals received at the

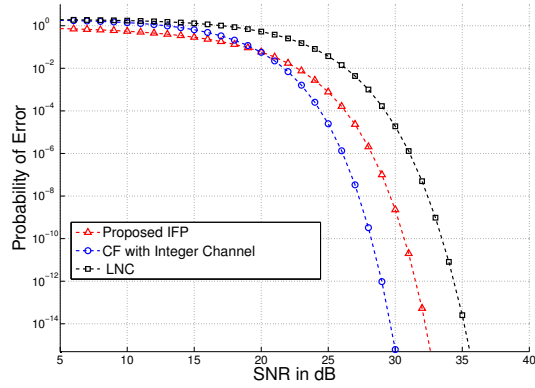


Figure 3.2: Comparison of probability of error at the relay for proposed scheme with varying SNR for  $L = 2$ . The total signal power is adjusted for fair comparison with other schemes.

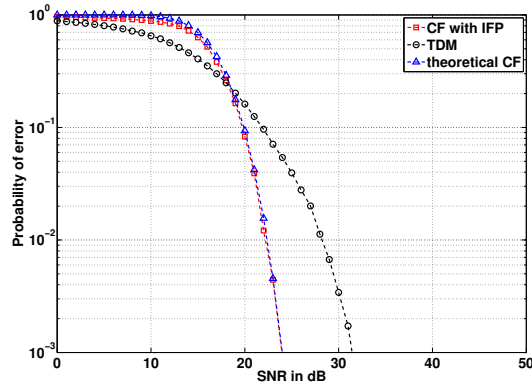


Figure 3.3: CF with IFP compared to TDM performance at the destination

relay in case of CF with IFP, which is not the case in TDM. Hence, in case the channel is in a deep fade, it results in an error in TDM whereas, such deep fading situations are overcome due to existence of multiple copies of same signals in CF. Note that we have again considered fixed channel gains here which reduces the theoretical integration of equalized noise in (3.28) to point to point error probability expression.

### 3.6.2 Outage Probability

In this subsection, we formulate the probability of an outage event to occur in the system. When the total power available at the transmitter is limited, the integer forcing precoding may fail depending on the total power requirement of the precoder. We define the outage probability  $P_{out}$  as the probability when the integer forcing precoding requires more power than the power available at the transmitter. Let the total power available at the transmitter be given by  $\gamma_a$ . This power  $\gamma_a$  is the maximum power which the transmitter can use to precode. The power required by the precoded signal is given by  $\gamma_w$ . Therefore, if the precoder is such that it

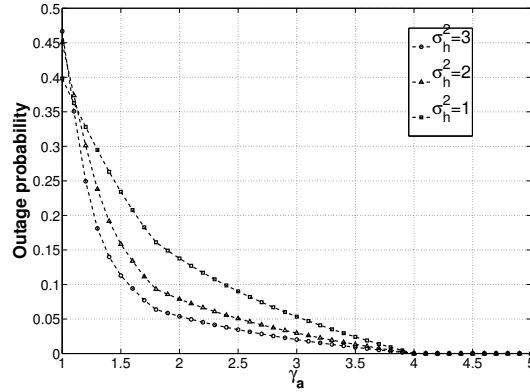


Figure 3.4: Outage probability curve. The horizontal axis shows the power available at the transmitter. The power available at the transmitter is assumed to be always greater than or equal to the signal power hence  $\gamma_a > 1$ . With the increasing channel variance  $\sigma_h$ , the outage decreases.

needs more power to precode than the available power  $\gamma_a$ , we have a case of outage as,

$$P_{out} = \Pr(\gamma_w > \gamma_a)$$

The distribution of above probability depends on the distribution of the channel fading coefficients because IFP is designed as  $w_l = h_l^2/h_l$ . The outage probability for a Gaussian channel with varying available power has been shown in Figure 3.4. The outage probability decreases with increasing available power. When  $\gamma_a > 4$  the outage probability goes to zero because  $\gamma_w < 4$  as shown in Theorem 1.

Figure 3.4 indicates that as the channel variance increases, the outage probability decreases. This is due to the fact that with increasing channel gain  $h_l$ , the power required by precoder decreases. Hence, the higher is the instantaneous channel gain, the better is the outage performance of IFP.

### 3.6.3 Achievable Rates

In this subsection, we compare the theoretical rates achieved by IFP in (3.22) with existent schemes. We consider the rates achieved by three other schemes: (i) the decode and forward scheme which requires to decode the original signals at the relay, (ii) compute and forward scheme without precoding and (iii) the compute and forward obtained by considering integral channels which serves as upper bound for achievable rates in CF. For fair comparison, we have adjusted the total power available at the transmitters in all the schemes to be equal to  $\gamma_x = \max(\gamma_w)$ . The channel gains  $\mathbf{h} = [h_1 \ h_2]$  are approximated by integers  $\mathbf{a} = [a_1 \ a_2]$ . The  $\log$  is taken to the base 2. The rates are measured in bits per channel use (bpcu). For decode and forward, the relay decodes one of the signals treating the other signal as noise. Hence,

$$R_{DF} = \log^+ \left( \frac{\gamma_x |h_1|^2}{\sigma^2 + \gamma_x |h_2|^2} \right)$$

For compute and forward without precoding, the relay computes the linear combination of original signals, and the non-integer part of the channel contributes to additional self-noise. The achievable rate is given by

$$R_{CF} = \log^+ \left( \frac{\gamma_x \|\mathbf{a}\|^2}{\sigma^2 + \gamma_x (\|\mathbf{h} - \mathbf{a}\|^2)} \right)$$

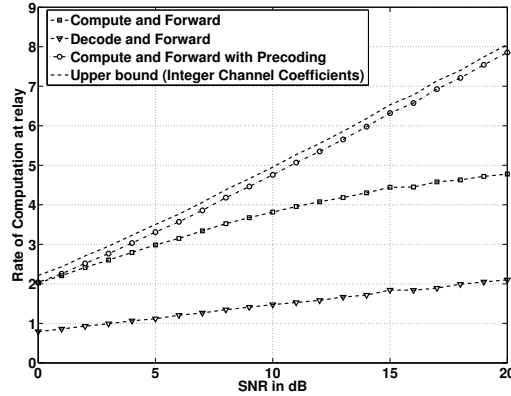


Figure 3.5: Comparison of achievable rates for various schemes for  $L = 2$ . The average value is taken over 10000 different random channel realizations. The rates are measured in bits per channel use (bpcu).

The upper bound of CF is computed considering the channels as integers ( $\mathbf{h} = \mathbf{a}$ ) in which no additional self-noise is present. The rate achieved is given by

$$R_{CF}^u = \log^+ \left( \frac{\gamma_x \|\mathbf{a}\|^2}{\sigma^2} \right)$$

We compare these rates with the rates for our proposed scheme. Figure 3.5 shows the achievable rates of different schemes for varying SNR. Clearly, the proposed implementation approaches the upper bound at high SNR's. The CF scheme without precoding suffers from increased self-noise as the signal power increases.

## 3.7 Concluding Remarks

### 3.7.1 Contributions towards state-of-the-art

In this chapter, we have proposed a novel scheme to implement the fundamental concept of CF. We have proposed an integer forcing precoder to precode the signals such that the signal obtained at the relay is naturally a linear combination of original signals. The IFP effectively reduces the channel into an additive white noise channel. The proposed scheme can remove the self-noise which arises in CF due to approximation of channel by an integer but it incurs the penalty of extra power usage at the transmitter. We have shown that this extra power used is in a finite range for most channels. Our scheme is a novel development in the direction to use precoding with CF and is a promising direction because we illustrate that our scheme outperforms the existent implementation for a variety of channels.

Our scheme is inspired by the motivation to remove the gap between the channel coefficients and integer coefficients in CF. This problem has been tackled using other ways in the literature like integer forcing receivers [45], interference alignment using antennas [46], MIMO techniques [47] etc.

Implementation of CF using precoders in the presence of CSI at transmitters has been limited in literature. For example, precoding has been applied using eigen direction alignment in two way relay channel framework with MIMO in [25]. Precoding from the relay to the receivers has been studied in [24]. An information theoretic insight of precoding using multiple antennas in CF was also provided in [47] which basically inspired our work. Our work departs from the state-of-the-art by providing a very basic analysis of applying

integer forcing precoding to CF framework proposed by Nazer and Gastper from a practical perspective with single antenna. The consequent analysis of our scheme is novel in nature due to use of simple mathematical tools like Bezouts theorem to obtain the performance expressions.

The main contributions of this chapter can be summarized as follows [48, 49]:

1. We propose a novel precoding technique to implement physical layer network coding using CF. Our precoder design is based on integer forcing and allows bypassing the self-noise limitation of existing formulations, thus providing higher achievable rates than existent relaying schemes.
2. We developed a decoder for relays of CF and characterized it using the generalization of Bezout's theorem. We also made an analytical derivation of the end-to-end probability of error for cubic lattices.
3. We analyze the two phases of transmission in the CF scheme, thereby characterizing the end-to-end behavior of the CF based on different performance metrics and not only one-phase behavior, as in available studies. Compared to conventional designs, we obtain a gain of approximately 5 dB in terms of probability of error as compared to conventional time division multiplexing (TDM) transmission.

We also add that our published work has been followed by some more recent research with precoding in CF using CSIT (Channel State Information at Transmitter) in [50, 51], which reaffirms the potential in our proposed framework.

### 3.7.2 Related Publications

This chapter has resulted in the following two publications:

#### *Journal*

**S. Gupta**, M. A. Vázquez-Castro, "Physical Layer Network Coding Based on Integer Forcing Precoded Compute and Forward." *Future Internet* 5, no. 3: 439-459, 2013.

#### *Conference Proceedings*

**S. Gupta**, M. A. Vazquez-Castro, "Physical-layer network coding based on integer-forcing precoded compute and forward," *Wireless and Mobile Computing, Networking and Communications (WiMob), 2012 IEEE 8th International Conference on*, vol., no., pp.600-605, 8-10 Oct. 2012

### 3.7.3 Possible Future lines of work

CF is a natural transmission framework for upcoming wireless relay networks. Achieving practically the theoretical limits of performance in such networks is a challenge and CF is a promising approach to improve this performance. Being in its initial phase, there are number of future lines of work possible in this area.

The precoder introduced in this work can be optimized according to specific scenarios and such designs are required to be explored further. Recently, [51] has attempted to identify the region of achievable rates using precoding from information theory perspective. A coding-theoretic assessment is still required to be developed.

Another important study is the estimation of the effect of imperfect CSI availability at the transmitter, which can reduce the efficiency of the scheme using precoder. An attempt in this direction has been made in [52] from the perspective of post-processing of the information.

In general, this scheme also reinforces the importance of the study of networking techniques at the physical layer and the immense potential in exploiting such techniques. Implementation of PNC using CF is still a highly unexploited work and sophisticated techniques of its practical implementation is crucial.

## Chapter 4

# Networking at PHY-MAC layer

### 4.1 Introduction

After the physical layer wireless signals are transferred to higher layer of the protocol stack, namely the link and network layer, different error correction mechanisms, quantization into bits and then into packets, header additions etc. are applied. This leads to formation of packets which are further passed to higher layers. In addition to improvement in networking techniques at the physical layer presented in previous chapter, the networking techniques at the link and network layers can also be improved in different ways. In this chapter, we focus on the contributions towards the improvement in networking techniques beyond pure physical layer, aiming at the data link layer and the Network layer.

#### 4.1.1 Objective and technique in overall framework

Our focus is to improve the throughput of the wireless networks using novel techniques at the link layer and network layer. To this end, different tools are used in order to design novel networking techniques. However, the implementation of these tools is required to be coherent with the existing optimization frameworks. Therefore, our overall objective is to propose coherent implementation of novel techniques in coherence with other frameworks and assess the issues and trade-offs which arise as a result.

We present two distinct contribution studies towards networking techniques in wireless relay scenarios at link and network layer. The first contribution aims to maximize the throughput in by using network coding along with the cross-layer optimization in the networks whereas, the second contribution aims to maximize the throughput in satellite networks by using cognitive transmission techniques.

**Objective1** “*Maximize throughput with Network coding and Cross-layer Optimization*”

The recent increase in demand of high quality video applications over wireless channels has led to widespread research in optimal usage of resources to support the user requirements. Particularly, the wireless links are time varying and hence it is beneficial to adapt the rate of transmission across these links by cross-layer optimization techniques (CL). For video applications, it is required to optimize the transmission rate based on providing maximal data rate, for a high definition video, and also maximize the flow continuity of the video, thereby providing minimal freezing of the video during playback. The Quality of Experience (QoE) driven Cross-layer optimization [53], for point-to-point data transmission across wireless networks is one such example of optimizing the video perceived by the user.

**Technique** *Network Coding*

The natural advantage of wireless networks covering a huge geographical region leads to a high ubiquity in their services, which is particularly advantageous for video transmission. As discussed previously, wireless relay networks form a fairly essential part of the modern communication systems and are a perfect platform to implement advanced relaying techniques like Network Coding. A number of recent advances in the field of network coding (NC) have established it as an efficient routing mechanism which can achieve multicast capacity [5]. The use of NC with CL techniques is imperative to the new generation technologies and has been studied in [54, 55].

The network coding opportunities are highly maximized with maximal overhearing of unintended packets at different nodes [14]. Otherwise termed as interference, an intelligent treatment of these overheard packets, can lead to an efficient resource utilization [56]. This overhearing is dependent upon the geographical location of the different user terminals [57]. Therefore, the geographical location of users should be considered in order to implement network coding in wireless network services [58].

### Problem Statement

Inspired by these works, we aim to study a joint implementation of QoE-driven cross layer optimization and network coding in two different topologies: bent pipe topology and X-topology. While we study the existent implications of the joint implementation, we also aim to take into account the geographical distribution of users in such a way, that the joint scheme improves the overall QoE of the end-user while optimizing the resource utilization. The aim is to design a solution which not only admits a flexibility of supporting heterogeneous user demands in terms of video quality, but also allows a uniform maximization of video quality in terms of both clarity and continuity.

**Contributions:** The main contributions of this work are as follows [59, 60]:

1. We propose a joint implementation of network coding and cross layer optimization over wireless relay channels. In particular, we present a video streaming solution over relay channels which implements network coding for maximization of throughput and QoE-driven cross layer optimization to provide seamless video quality experience to the end-user. We study such an implementation under the light of two different topologies: bent-pipe topology [59] and X-topology [60].
2. We identify that the reason for the successful combination of the component schemes is the existing trade-off between QoE and throughput. We analyze the behavior of this trade-off and the parameters affecting the performance in the light of different target requirements.
3. We further optimize the scheme by utilizing the geographical location of the users. To this end, we propose a location-adaptive algorithm to adapt the transmission rates in the joint network coding and QoE-maximization implementation such that based on the geographical distribution of the end-users, the joint implementation can be optimized.
4. Furthermore, in order to optimally utilize the geographical location information of the users, we propose the development of the *switching maps* as a novel tool for characterizing the performance.

**Objective2** “Maximize throughput with Cognition in satellite networks”

Terrestrial wireless networks have been well studied in literature and a number of networking techniques have been proposed for efficient and robust communications via terrestrial networks. However, the global infrastructure of communications systems faces an important challenge of increasing demands for ubiquitous and high quality services with limited resources at disposal. Satellite networks play a vital role in achievement of such ubiquity and efficiency in providing high quality services. Therefore, in this part of the chapter we will focus on satellite networks as a case study for wireless networks. The satellite networks have the advantages

such as higher coverage, limited maintenance after one-time installation etc. However, the satellite networks have disadvantages such as higher delays, installation costs, channel sensitivity to different frequencies etc. This leads to need of developing optimal techniques of maximizing the throughput of satellite networks.

In order to provide the satellite services, there is a need to explore not only more complex and efficient systems but also ensure their coherent existence with already operative networks without any disruptions in services. In recent years, the concept of cognition has been proposed for introducing additional levels of intelligence inspired by human cognition. Cognition aims to improve the overall system performance in terms of both quality and efficiency. In this part of the chapter, we set out to introduce the multi-satellite systems and the relevance of cognition in such systems, with emphasis on the practical dual satellite system case.

**Technique** *Cognitive transmission techniques in multi-satellite systems*

The growing traffic of satellites around the globe [61] has led to voluntary or involuntary creation of multi-satellite systems in which more than one-satellite operates over a geographical region. Therefore, it is mandatory to study the coexistence and implications of such systems.

- **Voluntary Systems:** The voluntarily formed multi-satellite systems are basically those multi-satellite networks *which are designed* such that more than one satellite operates over a geographical area to provide services to the users at the ground terminal in a coordinated manner [62]. Coordinated transmission is needed in such multi-satellite networks.
- **Involuntary Systems:** The involuntarily formed multi-satellite systems are those which *come into existence* due to increasing space traffic and limited spatial and temporal domains. For example, the overlapping coverage of multiple satellites leads to a ground user obtaining the signals from more than one satellite. Such systems are designed such that the incumbent satellite system services are not degraded and the different satellites can coexist. Cognitive transmission is needed in such multi-satellite networks.

It should be noted that in order to design the voluntary systems, coordinated transmission techniques are used. These techniques are primarily optimized in a centralized fashion. In case of involuntary systems, the techniques required should be more dynamic and adaptive thus predisposed for cognition-based optimization. We focus on involuntary systems and how cognition can help to significantly improve quality and efficiency in these systems. We will now address the basic principles of cognition and its applicability on involuntary multi-satellite scenarios.

The optimal allocation of resources is desired among the coexisting networks in involuntary systems, for both efficiency and optimal service. In such cases, the traditional allocation schemes are rendered sub-optimal. The task force report by Federal Communication Commission in [63] shows that a significant amount of radio spectrum remains underutilized almost 90% of the time based on the current static spectrum allocation policy which is based on grouping the services with similar technical characteristics. Another aspect which was pointed out in the report is that there is a very limited allocation to the Mobile Satellite Services (MSS) and the future wireless networks demand a high allocation to these services specially in L-band. Apart from the limited spectrum availability, it is important to ensure that the coexisting satellite networks do not interfere with each other and hence the quality of service provided is not degraded. Furthermore, in the land mobile scenarios (LMS), the ground unit should be provided ubiquity in service along with the quality and consequently the resources provided due to multiple satellites should be exploited. Therefore, an intelligent and dynamic resource allocation and management is required for next generation multi-satellite networks. Cognitive transmission techniques serve as an adequate solution to these problems. Such techniques have been well-studied in literature for different terrestrial and hybrid network and can be categorized under three headings to maximize the throughput as:



1. Improvement in Spectrum Utilization - See e.g. [64, 65, 66].
2. Improvement in Interference Management - See e.g. [67, 68, 69, 70].
3. Improvement in Channel Availability - See e.g. [71, 69, 72].

**Contributions:**

1. We present a critical review of the different cognitive communication techniques in satellite scenarios. In particular, we focus on the dual satellite scenarios for the study [73].
2. We develop a taxonomic analysis of the existing cognitive techniques on the basis of the different cognitive processes in the cognitive cycle. We also propose a concrete mapping between the different processes of the cognitive cycle and the corresponding processes of the engineering design. Our classification of various cognitive techniques is based on the aim with which the cognition is applied to the dual satellite system.
3. Based on our analysis, we conclude that a technological assessment of cognition is required in place for optimal combination of different aims of cognition.

### 4.1.2 Outline of the Chapter

This chapter is organized as follows: in section 4.2, we focus upon the study of network coding with cross layer optimization for video transmission over relay networks. We present different topologies and models for the proposed technique. In addition, a number of experimental results are presented to analyze the different paradigms of the scheme. In section 4.3, an in-depth taxonomical analysis of various methods of cognition in satellite systems, particularly dual satellite systems, is shown. In section 4.4, the conclusions are presented.

## 4.2 Maximizing throughput using QoE-driven Cross-layer optimization for video transmission

In this section, we present our first contribution study which is towards maximizing the throughput in wireless relay networks using the technique of network coding along with cross-layer optimization.

### 4.2.1 Cross-layer design preliminaries

**Cross layer design** We now describe the cross application/transport layer design to transmit the video content. It is assumed that the video content is transmitted using UDP (User Datagram Protocol) over RTP (Real-Time Transport Protocol). The source node is equipped with a codec responsible of compression of video content. With the availability of open source codecs such as VP8, the codec can be reconfigured to deliver a target bit-rate. Therefore, at the source node, the output rate can be adjusted as desired. This output rate from any node say  $x$  where  $x \in \{A, B\}$  at time  $t$  is denoted by  $r_x(t)$  and it is measured in packets per second. The packets are further passed down to the network layer maintaining coherence with standard protocol stack. The rate of transmission of video payload is adapted to the network conditions by optimization of the QoE at the end-user as shown in [9]. With the use of Real Time Control Protocol (RTCP) signaling, the source node collects the feedback information about the Round trip time (RTT) from the destination. The RTCP feedback signaling provides the source node with information to re-configure the codec rate to suit the target needs.

**Buffer and delay model** Let the node  $x$  transmit the packets at time  $t$  at rate  $r_x(t)$ .  $x$  also maintains a transmission buffer of size  $B_c$  packets. Let  $R_{xy}(t)$  be the channel capacity between the nodes  $x$  and  $y$  at time  $t$ . Therefore, if  $R_{xy}(t) < r_x(t)$  then, if buffer is empty, only  $R_{xy}(t)$  packets are transmitted to the destination. Out of the rest of  $r_x(t) - R_{xy}(t)$  packets,  $B_x(t) \leq B_c$  packets are stored in the free space of buffer and are transmitted with a delay while the rest of the packets, if any, are lost. The packets stored in the buffer are retrieved in FIFO (First In First Out) manner and transmitted in the next transmission slot. We assume that the size of the buffer  $B_c$  is less than the channel throughput capacity at all times. This assumption implies that if some packets are stored in the buffer at any time instant, these packets will be transmitted through the channel in the next transmission. Therefore, the maximum delay that any packet can face is one transmission cycle. Note that the buffer structure at both the nodes  $A$  and  $B$  is identical.

**QoE-driven rate adaptation Algorithm** The source updates the rate of transmission using the following rate control update as shown in [9]. The source node  $x \in \{A, B\}$  changes the rate as,

$$r_x(t+1) = r_x(t) + \Delta r_x \quad (4.1)$$

where

$$\Delta r_x = \delta [U_1'(r_x(t)) - \alpha(k)\tau(k)] \quad (4.2)$$

Here  $\delta$  is the step size,  $U_1(r(t)) = m \log(r(t)) + h$  is a utility function to characterize the video quality in terms of rate as shown in [9] with  $m, h$  are numbers suitably selected. The delay is given by  $\tau(k)$  as observed by the receiver at  $k$  where  $k = (t+1) - \tau_{fd}$  and  $\tau_{fd}$  is feedback delay. The factor of  $\alpha(k)$  is the penalty value which controls the adaptation to react faster to increasing delay constraints and packet loss. It is updated as

$$\alpha(k) = \begin{cases} \alpha(k-1) & \text{if } \alpha(k-1) = \alpha_{init} \text{ AND } \tau(k) \leq \bar{\tau} \\ \lambda \alpha(k-1) & \text{if } \Delta_p(k) = \bar{\Delta}_p \text{ AND } \tau(k) > \bar{\tau} \\ \frac{1}{\alpha(k-1)} & \text{if } \alpha(k-1) > \alpha_{init} \text{ AND } \tau(k) \leq \bar{\tau} \end{cases} \quad (4.3)$$

Here  $\bar{\tau}$  is the threshold delay. If the delay value is above this threshold value, it causes an observable freezing of the video. Also,  $\bar{\Delta}_p$  is the threshold packet loss. If none of the conditions are met, the value of  $\alpha(k)$  remains  $\alpha(k-1)$ . Here, the values of  $\alpha_{init}$  and  $\lambda > 1$  are design parameters that are fitted as per the system considerations.

## 4.2.2 System Topologies

### 4.2.2.1 Bent-pipe Topology

The first topology that we focus upon is based on the scenarios of wireless video transmission system where two users want to exchange their videos via a relay node. Such scenarios model situations like video conferencing between two parties across geographically separated area. The node  $A$  (or  $B$ ) transmits the video packets to the relay and the relay sends these packets to node  $B$  (or  $A$ ) as shown in figure 4.1a. The wireless channel between any two nodes has variable capacity to transmit packets across the nodes depending on various factors like environmental conditions (rain, cloud etc), congestion due to other users, etc. We model channel links between the the nodes to have a certain throughput capacity which varies with time. The throughput capacity is the maximum packets that can be transmitted between any node  $x$  and node  $y$  at time  $t$  and is denoted by  $R_{xy}(t)$ . It is measured in packets per second (pps). As shown in figure 4.1a,  $R_{AB} = 0$  for all  $t$ . Throughout this paper, at any time instant  $t$ , all links are assumed to have equal throughput capacities  $R_c(t)$ , that is,  $R_{xy}(t) = R_c(t) \forall xy \in \{AR, BR, RA, RB\}$ .

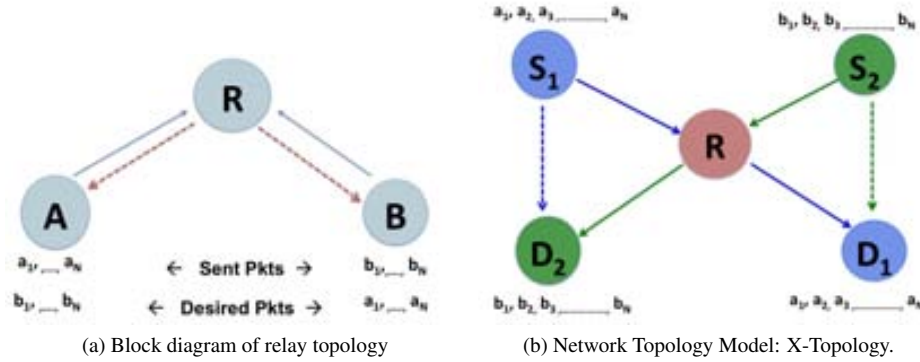


Figure 4.1: System topologies

#### 4.2.2.2 X-topology

The second topology that we focus upon is based on the scenarios where wireless relay channels are used for video transmission to different users, which are geographically separated from the source. Consider the system model topology as shown in 4.1b which is commonly known as the X-topology. The sources  $S_1$  and  $S_2$  send their video streams to the destinations  $D_1$  and  $D_2$  respectively via the relay  $R$ . The wireless channel between all the nodes varies with time due to various factors like rain, clouds etc, and distance, power, interference etc.. The maximum capacity of wireless link between any two nodes  $x$  and  $y$  at any time  $t$  is given by  $R_{xy}(t)$  and measured in packets per second (pps). It is assumed that the link capacity of all the links between nodes and the relay node is equal to  $R_c(t)$ , therefore  $R_{xy}(t) = R_c(t) \forall xy \in \{S_1R, S_2R, RD_1, RD_2\}$ . This is a reasonable assumption because channel conditions variation over a large part of atmosphere largely remains the same.

Further, we model the maximum capacity of wireless links between the (terrestrial) nodes using [57] in coherence with IEEE 802.11. The maximum rate supported by the link is determined by the transmit power, distance, thermal noise, bandwidth, interference, etc. The maximum rate between any two nodes  $x$  and  $y$  is defined as

$$R_{xy} \leq \frac{w}{2} \log \left( 1 + \frac{\rho \|x-y\|^{-\alpha}}{\eta + I} \right) \quad (4.4)$$

where  $\rho$  is the transmitting power of the node  $x$ ,  $\eta$  is noise power spectral density,  $w$  is the system bandwidth under consideration,  $\alpha$  is the attenuation factor,  $\|x-y\|$  is distance between two nodes  $x$  and  $y$ , and  $I$  is the amount of external interference power. We assume continuous rate region in order to illustrate the possible rates supported irrespective of the standard being used. In coherence with IEEE 802.11, we note that any rate supported by the standard in the rate region can be used. Assuming equal transmission power, bandwidth, noise power spectral density and attenuation factor, the maximum capacity between any two nodes is a function of the distance between the nodes. We assume that due to the geographical proximity, the packets from  $S_1$  can be overheard by  $D_2$  and the packets from  $S_2$  can be overheard by  $D_1$ . We also assume a symmetrical model such that  $R_{S_1D_2} = R_{S_2D_1} = R_O$ , and their value is determined by eq. (4.4). It should be noted that  $R_{S_1D_1} = R_{S_2D_2} = 0$  at all times.

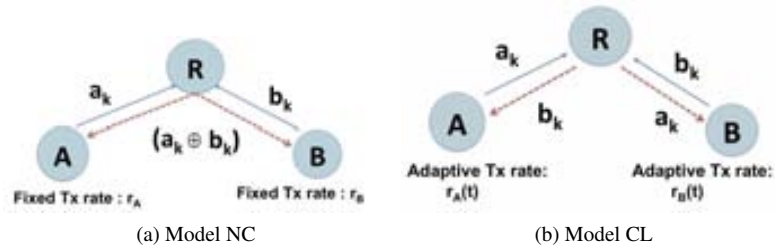


Figure 4.2: Model NC and CL in bent-pipe topology

## 4.2.3 Proposed Joint Network Coded Cross-layer Optimized Model

### 4.2.3.1 Bent-pipe Topology

**Model NC** We consider Model NC as the network coding model as shown in figure 4.2a. Here the key characteristic is the implementation of practical network coding at the relay node. The nodes A and B send their packets to node R in alternate time slots at a fixed rate, which remains constant at all times. Therefore,  $r_A = r_B = r_C$ . The node R decodes packets received from A and B in alternate time slots. Since the packets are received in alternate time slots, they are free from interference and only additional noise is removed while decoding at the relay. R maintains a queue of the incoming packets from each node A and B. The node R performs a bit by bit XOR of two packets, one from A and B each and sends the combination to both A and B using the multicast over UDP. When the node A and B receive a packet from R, they use the XOR operation with their own packet to obtain the packet of the other node. Note that in this model, the role of R is to mix the packets from A and B. The rate at which it mixes and transmits the packets is the same rate at which it receives the packets. Note that all successfully received packets at R can be sent to A/B as the links have identical throughput capacity.

**Model CL** We consider Model CL as shown in figure 4.2b. The key characteristic in this model is the cross layer optimization (CL) based on QoE of end-user video. The nodes A and B transmit the RTP packets over UDP to node R in alternate time slots. R collects the packets from A and B and sends the packet from A to B and from B to A in next two time slots. Each of these links A to B and B to A uses end to end QoE-driven adaptive video transmission as in [9]. The relay simply forwards the packets it receives from one node towards the other node. After the destination node obtains a packet, it sends back the RTCP to relay, which sends this RTCP to the source.

**Model NC+CL** The system model has been shown in figure 4.3. A and B transmit their packets towards R. The relay R, obtains the packets from nodes A and B and performs XOR. The new packet, which is XOR of A and B is multicast to A and B by R. The received packet is then XORed at A and B with their own respective packets to obtain the packet of the other node. The destination node observes the delay and packet loss statistics, based on which, it updates the codec rate at which it transmits for the next transmission. This is because, we assume identical channel links between R-A and R-B and the rate is updated identically at both nodes using (4.1)-(4.3).

In this model, clearly, less number of time slots are used for transmission, thereby increasing the throughput. Additionally, the rate is being adapted to optimize QoE to improve the end-user video quality. Note that the rate adaptation is transparent to network layer mixing of packets which allows both NC and CL to be implemented coherently.

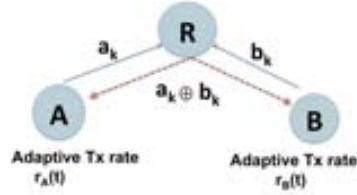


Figure 4.3: Model NC+CL

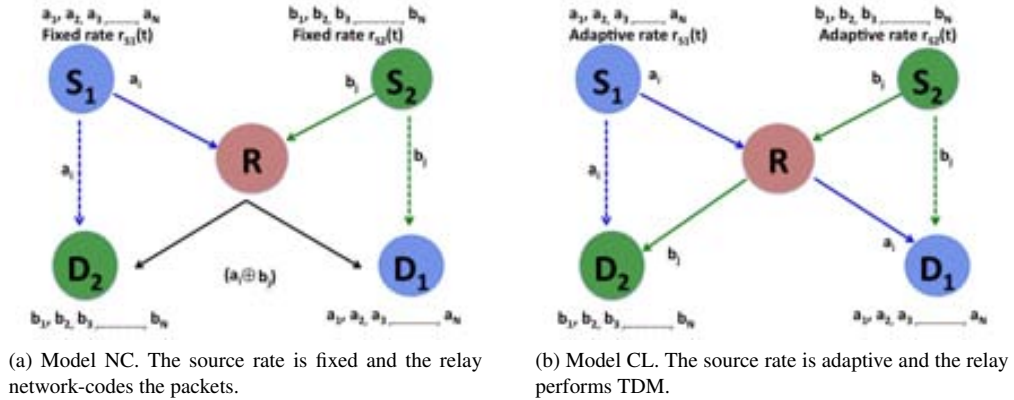


Figure 4.4: Model NC and CL with X-topology. The dotted line shows the overhearing links.

### 4.2.3.2 X-topology

**Model NC** The network-coding model, or Model NC, implements only network coding at the packet level at the relay node with no cross-layer optimization. As shown in figure 4.4a, the sources  $S_1$  and  $S_2$  send their RTP packets to the relay  $R$  in alternate time slots at a fixed rate  $r = r_{S_1}(t) = r_{S_2}(t)$  at all times. The relay node obtains the packets from both the sources and performs a bit-by-bit XOR of the packets from the two sources. The relay then multicasts the new XOR-ed packet to both the destinations  $D_1$  and  $D_2$ . If the rate  $r$  is less than the link capacity  $R_{S_1D_2} = R_{S_2D_1} = R_O$ , then the nodes  $D_1$  and  $D_2$  overhear the packets from  $S_2$  and  $S_1$  respectively. When the destination nodes receive the XOR-ed packets, they decode the intended packet by doing an XOR between the received packet and overheard packet. *Note that the destination nodes can decode the packets only if  $r < R_O$  because in the absence of overhearing the destinations cannot eliminate the packet of the other source from the received packets.* Also note that all the packets that are successfully received at the relay  $R$  are successfully transmitted to the destination also due to the same channel capacity  $R_c$ . Clearly, it takes three time slots to transmit  $r$  packets from source to destination in this case, thereby increasing the overall throughput of the network as compared to traditional Time division multiplexing (TDM).

**Model CL** The cross-layer optimization model, or Model CL, is a model implementing the QoE driven cross transport/application layer optimization along with time division multiplexing (TDM). The cross-layer optimization is designed such that the source rate is adapted to the channel conditions to improve the Quality of Experience of the video as perceived by the end-user. With an improved QoE, the end-user views a continuous video without any freezing of video frames during the playback. As shown in fig.4.4b, the source nodes  $S_1$  and  $S_2$  transmit their RTP packets over UDP to node  $R$  at the rate  $r_{S_1}(t) = r_{S_2}(t) = r(t)$ . The node

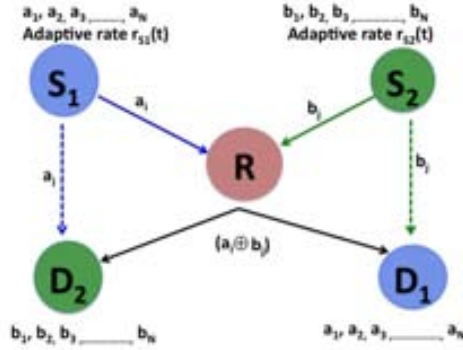


Figure 4.5: Model NC+CL. The source rate is adaptive and the relay network-codes the packets.

$R$  collects the packets from nodes  $S_1$  and  $S_2$  and transmits the packets of node  $S_1$  to  $D_1$  and those of node  $S_2$  to  $D_2$  in alternate time slots by TDM. When the destination receives the packet, it sends back RTCP to the respective source using the same channel through the relay. When the source observes the RTCP, and identifies the network statistics like, packet delay, packet losses etc, it updates its rate in order to optimize the QoE as perceived by the user. In order to update this rate, we use the QoE-driven optimization [53] as explained in section 4.2.1. This algorithm is based on adapting the source rate to the channel conditions in order to improve the Quality of Experience such that, the rate adaptation is transparent to what happens below the transport layer. The relay in this case simply forwards the packets it obtains.

Note that in model CL, it takes 4 slots for  $r(t)$  packets to reach the destination node; hence the overall throughput is lower as compared to Model NC for the same rate. There is also no dependency of packet decoding at the destination and overhearing capacity,  $R_O$  as opposed to Model NC and the user-end QoE is always optimized.

**Model NC+CL** In this model, we jointly implement the key characteristics of Model NC and Model CL to obtain the Model NC+CL such that it not only optimizes the QoE of the end-user, but also simultaneously optimizes the resource utilization. As shown in figure 4.5, in this model, the sources  $S_1$  and  $S_2$  transmit their packets to  $R$  at the rate  $r(t)$ . These packets are overheard by the unintended destinations as long as  $r(t) < R_O$ . The relay obtains the packets and performs a bit-by-bit XOR of the packets from both the sources and multicasts the XORed packets to both the destinations. The destination nodes obtain the packets and then decode them using the overheard packets as in Model NC. Upon decoding the packets, the destination nodes send back the RTCP to the source nodes. The source nodes update their rates using Eqs.(4.1)-(4.3) and further send the packets. As in previous cases, we assume identical channel links to and from the relay to all nodes; therefore, all packets successfully received at the relay are transmitted successfully to the destination.

Note that schemes in Model NC and Model CL can be implemented coherently because the rate adaptation at the transport layer is transparent to the lower layers, and network coding at network layer is transparent to the upper layers of the protocol stack. In Model NC+CL, only three time slots are required to transmit  $r(t)$  packets per time slot, as in Model NC. Also note that in Model NC+CL, the rate is adapted as per the channel conditions to optimize the QoE of the end-users as in Model CL. However, the maximum rate at which the packets can be transmitted cannot exceed  $R_O$  in order to allow overhearing and continue the use of reduced time slots.

## 4.2.4 Throughput-QoE Tradeoff

The channel conditions in our model are subjected to different environmental conditions, malicious user attacks etc. Such conditions can also cause a sudden drop in channel throughput capacity. In this section, we characterize a trade-off existent between QoE and throughput in case of such sudden channel drop conditions. We present this analysis primarily for bent-pipe topology but as it will be clear that it can be extended to any other topology.

### 4.2.4.1 Trade-off in sudden drop conditions

While NC provides a higher overall throughput in the scheme, QoE-driven adaptation ensures both, a higher throughput per link, and a better flow continuity by minimizing the delay and freezing of video at the user-end. The QoE-driven adaptation is done using the feedback from the channel, without the absolute knowledge of total channel throughput capacity. In realistic conditions, the channel throughput capacity may suddenly drop due to random unavailability, unexpected congestion etc which is common for wireless links. In good channel conditions, if the rate provided by the source is increased steeply to achieve higher throughput, the performance of the system is good. However, if suddenly the channel throughput capacity drops, there is a considerable packet loss and a considerable delay faced by the packets leading to loss in flow continuity. On the other hand, when the channel condition is good, if the source rate is increased more cautiously, analyzing the probability of channel to degrade, then in case of sudden channel drop conditions, the packet loss and delay will be lesser and QoE can be better controlled. The former scheme leads to a high throughput but risky QoE performance, whereas the latter compromises in throughput, but provides a better QoE in terms of sudden channel capacity drop. Therefore, we see a trade-off existent in the scheme between throughput and QoE. In the next subsection, we modify the Model NC+CL adaptation scheme to make the proposed scheme more robust against such channel variations.

### 4.2.4.2 Modified Trade-off aware optimization

We propose a modified 'trade-off aware' adaptation in Model NC+CL to maintain robustness in the proposed NC+CL model against the channel capacity variations of wireless links. We assume that the source is unaware of the time and magnitude of drop in channel throughput capacity. We assume that the source can only estimate a probability of drop to occur. This probability is given by  $p_d$ . We propose a basic 'trade-off aware' rate adaptation in which the equation (4.1) is updated as follows:

$$r_x(t+1) = r_x(t) + \Delta r'_x \quad (4.5)$$

where

$$\Delta r'_x = (1 - p_d)(\Delta r_x) + p_d \Delta r''_x$$

such that

$$\Delta r''_x = \delta \left[ \frac{U'_1(r_x(t))}{n} - \alpha(k) \tau(k) w \right]$$

$n, w > 1$  and  $\Delta r_x$  is as in eq. (4.2). The motivation behind this rate increment is as follows: in case of sudden channel misfunction, if it is expected that channel quality will drop, the source sets the probability of this drop to  $p_d$ . The rate increment is then partially influenced by the traditional algorithm in [9] only with a weight of  $(1-p_d)$ . However, with a weight of  $p_d$ , the rate increment is influenced by a more conservative increase given by  $\Delta r''_x$  which can be set to slower increase and fast decrease using appropriate values of  $n$  and  $w$ . This is a preliminary method with which we illustrate the trade-off in the section 4.2.7.

### 4.2.5 Location Adaptive Joint Network Coded Optimized Model

We note in section 4.2.3.2 that the Model NC+CL has the advantages of both Model NC and Model CL. However, it also bears the disadvantage of Model NC: the destination (say  $D_1$ ) cannot decode the packets of the desired source ( $S_1$ ) from the received network coded packets if the unintended source rate ( $r_{S_2}(t)$ ) exceeds the overhearing capacity ( $R_O$ ). This is because, in such cases, the destination is unable to eliminate the unintended source packets ( $S_2$ ) from the XOR of packets of  $S_1$  and  $S_2$ . In other words, when the source rate exceeds the overhearing capacity in Model NC+CL, the destination node is not able to decode the packets. Consequently, the source observes packet losses from the network statistics (RTCP signaling), and decreases its transmission rate. This decrease is not due to degradation in channel capacity, but due to inability to decode the packets at the destination as a result of network coding. Hence, the network capacity is not well utilized.

In this section, we propose to eliminate the restriction of overhearing capacity rate on maximum source rate in model NC+CL pointed above, by using the information about the location of the destination node. We assume that both the sources and the relay node are aware of the location of the destination nodes. This information can be conveyed using a dedicated control channel flag before the transmission starts [58]. Using the information of the location of the destination nodes, the overhearing capacity  $R_O$  can be computed with eq. (4.4). Now the key idea is to implement the model as long as the source rate is less than  $R_O$ . When the source rate exceeds the overhearing capacity as a result of QoE-driven adaptation, the Model NC+CL switches to either Model NC or Model CL. Therefore, we define two types of switching mechanisms:

1. NC-based Switching - The mechanism for NC-based switching is as follows: the transmission is according to Model NC+CL whenever the source rate is within the overhearing capacity. However, when the overhearing is not available due to high source rate, the Model NC+CL is switched to Model NC. Therefore, the rate adaptation of the source node is given by

$$r(t+1) = \begin{cases} r(t) + \Delta r & \text{if } r(t) + \Delta r < R_O \\ r(t) & \text{otherwise} \end{cases} \quad (4.6)$$

where  $\Delta r$  is given by eq. (4.2). When NC-based switching is used in Model NC+CL, the relay performs NC at all times, however, the source nodes adapt the rate using eq. (4.6) with eqs. (4.1)-(4.3). This switching mechanism requires decision-making eq. (4.6) at the end of the source nodes.

2. CL-based Switching - The mechanism for CL-based switching is as follows: the transmission is according to the Model NC+CL as long as source rate is within the overhearing capacity. However, when overhearing is not available due to high source rate, the Model NC+CL is switched to Model CL. Therefore, the relay transmits the packet using the following decision making scheme:

$$\text{Relay routing scheme} = \begin{cases} \text{NC} & \text{if } r(t) < R_O \\ \text{TDM} & \text{otherwise} \end{cases} \quad (4.7)$$

When CL-based switching is used in Model NC+CL, the source nodes adapt the rate according to eqs. (4.1)-(4.3) at all times whereas the relay either implements NC or TDM as per eq.(4.7). This switching mechanism requires decision-making eq. (4.7) at the end of the relay node.

In order to switch from NC+CL model to another model, we propose the flowchart of algorithm shown in figure 4.6. It will be shown from the simulations for bent-pipe topology in next section, that the Model NC gives an advantage of overall throughput whereas Model CL gives an advantage of overall QoE. Further, the location of the user determines its overhearing capacity, and the decision making point for the switching in eq.(4.6) or eq.(4.7). Therefore, based on the user location and its preference to throughput (QoS) or



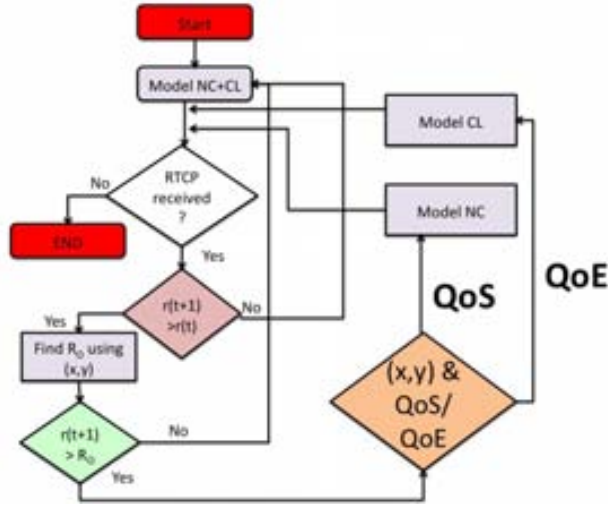


Figure 4.6: Flowchart to implement switching mechanism. The coordinates  $(x,y)$  represent the location of destination nodes.

QoE, the NC-based or CL-based switching is used. Firstly, the user-location is used to determine the  $R_o$ . Consequently, the user preference is utilized to determine the switching mechanism. If the users prefer to have a high throughput at all times, then the Model NC+CL is implemented with NC-based switching. Therefore, as soon as the rate  $r(t)$  exceeds  $R_o$ , the rate adaptation at source is done using eq.(4.6) while the relay continues to perform NC until the the CL-optimization algorithm causes the sources to lower the rate. However, if the users prefer to have a high QoE at all times, then the Model NC+CL is implemented with CL-based switching. Therefore, as soon as the rate  $r(t)$  exceeds  $R_o$ , the relay stops network coding, and begins to transmit the packets using TDM. The sources continue to adapt to the channel conditions. Consequently, when the sources reduce the rate below  $R_o$  as a result of adaptation, the relay begins network coding again. Note that following our assumption of symmetry, both the users will have the same demand thereby allowing the system to follow one of the switching mechanism, and there is no conflict of interest.

## 4.2.6 Performance Metrics

In this section, we present the performance metrics used to assess the three models. We firstly present the analytical expression of buffer and delay in order to describe the performance metrics.

### 4.2.6.1 Preliminaries

#### Buffer Occupancy

The buffer occupancy is the number of packets in the buffer of the source node, at time  $t$  and is given by

$$B(t) = \min(B_c, \max(0, (r(t) + B(t-1) - R_c(t)))) \quad (4.8)$$

with  $B(0) = 0$ . The expression is derived based on the buffer modeling as described in section 4.2.1. The number of packets stored in buffer are the total previous packets and new packets without those transmitted through the channel. If all packets are transmitted, the max constraint ensures that buffer is empty. However, the buffer occupancy cannot exceed the buffer capacity, which is ensured by the min constraint.

### Delay

The delay as observed at the destination is given by the time difference between the actual arrival of the packet at the destination and the expected time of arrival of the packet at the destination. However, by our assumption, the packets face a maximum delay of one transmission cycle (the time slots between the next transmission arriving at the destination) when the buffer is non-empty. Assuming a transmission cycle comprises of  $d_i$  seconds, the delay can be defined as

$$\tau(t) = \begin{cases} d_i & B(t - (d_i + 1)) > 0 \\ 0 & B(t - (d_i + 1)) = 0 \end{cases} \quad (4.9)$$

The values of  $d_i$  depend upon the model being used. For example, in the bent-pipe topology with Model NC and NC+CL,  $d_i = 3$  because each session comprises of three time slots, whereas, for Model CL,  $d_i = 4$ .

#### 4.2.6.2 QoS Metrics

We evaluate the Quality of Service in terms of two metrics namely throughput and packet loss.

**Throughput** The network throughput, or throughput, is defined as

$$\rho = \frac{\text{Packets obtained at A} + \text{Packets obtained at B}}{\text{Total time}}$$

It is measured in packets per second (pps). The average throughput per node is given by  $\rho/2$ .

**Proposition 10.** *The throughput can be analytically expressed as*

$$QoS = \rho = \frac{1}{T} \left( 2 \times \sum_{t=1}^T \min(R_c(t), r(t) + B(t-1)) \right) \quad (4.10)$$

where  $B(0) = 0$ .

*Proof.* The packets obtained at the destination are either the previous buffer content and/or the new packets inserted. However the total packets transmitted cannot exceed the channel throughput capacity. Therefore, the instantaneous throughput at both nodes is given by  $2 \times (\min(R_c(t), r(t) + B(t-1)))$ . Averaged over  $T$  time slots, we get the expression in eq.(4.10).  $\square$

We also define network utility ( $\eta$ ) as the ratio of the total throughput achieved to the total offered throughput by the network. It is measured in percentage.

**Packet Loss** The packet loss is defined as

$$\Delta_p = \frac{\text{Total packets lost}}{\text{Total packets desired to be transmitted}}$$

It is measured in percentage of packets which are desired to be transmitted.

**Proposition 11.** *The packet loss can be analytically expressed as*

$$\Delta_p = \frac{\sum_{t=1}^T \max(0, [r(t) + B(t-1) - R_c(t) - B_c])}{\sum_{t=1}^T r(t)} \quad (4.11)$$

*Proof.* The expression can be derived by computing the difference between the total packets to be transferred (packets inserted and the previous buffer occupancy) and the packets possible to be transmitted and stored (channel throughput capacity and the buffer capacity). However, if the difference is negative, it implies there is no packet loss and therefore, the instantaneous packet loss expression is given by  $\max(0, [r(t) + B(t-1) - R_c(t) - B_c])$ . Averaging over the total time and dividing by the total transmitted packets, we get the expression eq.(4.11).  $\square$

#### 4.2.6.3 QoE Metrics

**Flow Continuity** The flow continuity is defined as the probability that the delay faced by a packet is less than a tolerable threshold. Therefore, let the packet delay is  $\tau$  and the tolerable threshold is  $\bar{\tau}$  then the flow continuity is defined as

$$t_f = Pr(\tau < \bar{\tau})$$

A high value of  $t_f$  yields better performance.

**Proposition 12.** *An analytical expression of flow continuity is given by*

$$t_f = 1 - \frac{d_i}{T} \sum_{t=1}^T (\text{sgn}(\min(B_c, \max(0, v(t)))))) \quad (4.12)$$

where  $v(t) = (r(t-1) + B(t-2) - R_c(t-1))$  and  $\text{sgn}$  is the signum function, defined as

$$\text{sgn}(x) = \begin{cases} 0 & x = 0 \\ \frac{x}{|x|} & \text{otherwise} \end{cases}$$

*Proof.* The flow continuity is given by  $t_f = Pr(\tau < \bar{\tau})$ . The instantaneous delay is given by eq.(4.9). Therefore, the dependency of delay is directly on the buffer occupancy assuming  $\bar{\tau} < d_i$ . Hence, we can rewrite the flow continuity as

$$QoE = t_f = 1 - \frac{d_i}{T} \sum_{t=1}^T (\text{sgn}(B(t-1)))$$

where  $B(t)$  is as defined in eq.(4.8). Therefore, inserting it above we get eq.(5.4).  $\square$

#### 4.2.6.4 Joint Utility

We define a joint utility metric to assess the performance of the models in terms of trade-offs between QoE and QoS. Inspired from [7], the joint utility metric is defined as the weighted average of the network utility percentage and flow continuity percentage of the network,

$$u(\rho, t_f, \mu) = \frac{\lambda_1 \rho + \lambda_2 t_f}{\lambda_1 + \lambda_2} = \frac{\mu \rho + t_f}{\mu + 1} \quad (4.13)$$

where  $\mu = \lambda_1/\lambda_2$ . If the weight factor is set to  $\mu < 1$ , it indicates the joint utility metric gives more weightage to flow continuity, whereas if  $\mu > 1$ , it indicates that joint utility metric gives more weightage to throughput. A weight factor of  $\mu = 1$  shows that both throughput and flow continuity are equally important.

### 4.2.7 Performance Analysis

In this section, we present numerical results to prove the gains in performance of the proposed scheme in different scenarios.

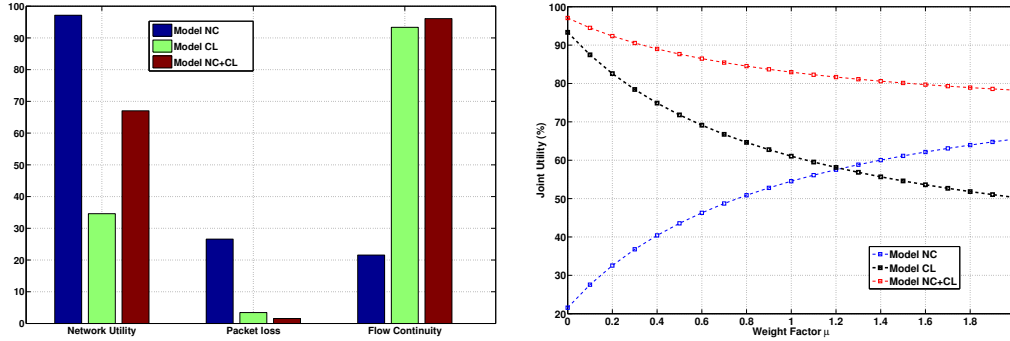


Figure 4.7: Model NC+CL provides a fine balance in terms of throughput and flow continuity giving high joint utility

#### 4.2.7.1 Bent Pipe Topology

**Scenario 1** We consider a system scenario where all the links are identical capacity links. The performance is analyzed for  $T = 200$  s. The channel provides a throughput capacity of 350 Kbps in first 40 seconds and then falls to 200 Kbps for rest of the time. The nodes A and B exchange the video packets via relay R. The three models are implemented under the given scenario. The constant rate in Model NC is fixed at 300 kbps, whereas in Model CL and Model NC+CL, the rate adapts according to channel conditions. Figure 4.7 shows the performance of the three models for the different metrics.

From the figure 4.7(top), it can be seen that the Model NC provides maximum network utility but also leads to a high amount of packet loss. On the contrary, Model CL, provides a lower network utility but the cautious rate adaptation leads to low packet loss rate. Further, in terms of flow continuity, Model CL outperforms Model NC, due to adaptation in rate with varying channel condition. Our proposed Model NC+CL, provides a fine balance between Model NC and Model CL. This balance can be seen from the joint utility metric defined in eq. (4.13) in figure 4.7 (bottom). As the weight factor  $\mu$  increases, the utility of Model NC increases due to higher weightage to throughput but as  $\mu$  decreases, the utility of Model CL increases due to higher weightage to flow continuity. However, the Model NC+CL gives a high utility for all values of  $\mu$  indicating a better performance in terms of both throughput and flow continuity.

**Scenario 2** In this set of simulations, we illustrate the QoE-throughput trade-off as described in Section 4.2.4. We consider a scenario in which the channel throughput capacity is fixed to  $R_{c1} = 350$  kbps for the first 40 seconds. For the next 160 seconds, the channel throughput capacity drops to  $R_{c2}$ . In figure 4.8, we show the throughput and QoE variations, for different magnitudes of  $R_{c1} - R_{c2}$ . It can be seen that our proposed trade-off aware adaptation gives a 10 kbps lower throughput as compared to the trade-off unaware adaptation, but gives more than twice the gain in flow continuity for a sudden drop of 300 kbps when  $p_d = 0.5$ . It is also clear that as probability of channel to drop approaches 1, the performance of trade-off aware NC+CL significantly outperforms trade-off unaware NC+CL.

To illustrate this gain more closely, we study the throughput and flow continuity while varying both  $p_d$  and  $R_{c2}$  as shown in figure 4.9a and figure 4.9b. It can be seen that as  $p_d$  tends to 1, the flow continuity gain varies from as small as 0.5% for a drop of 100 Kbps to 126% for a drop of 300 Kbps. The price paid in terms of throughput is : 18% loss in throughput for a drop of 100 Kbps whereas a 10% loss in throughput with a drop of 300 Kbps. Note that the performance of trade-off aware and unaware adaptation coincide at  $p_d \rightarrow 0$ .

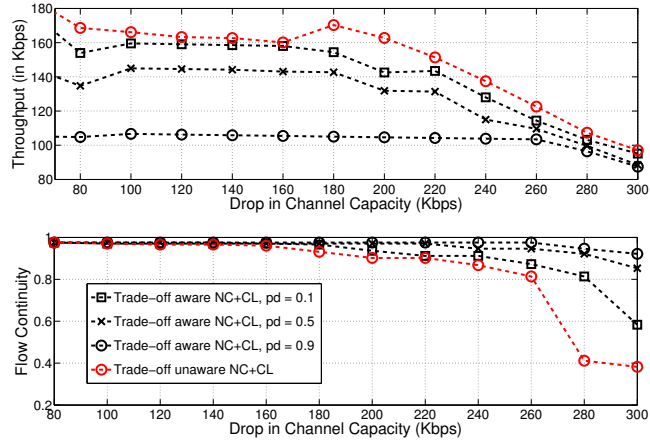
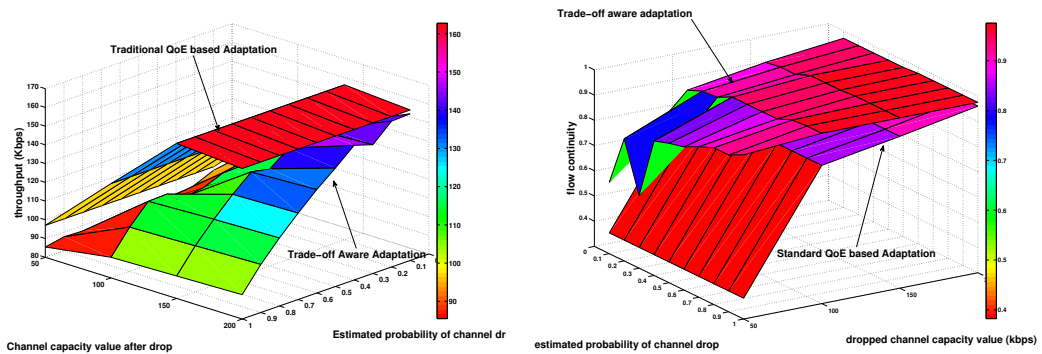


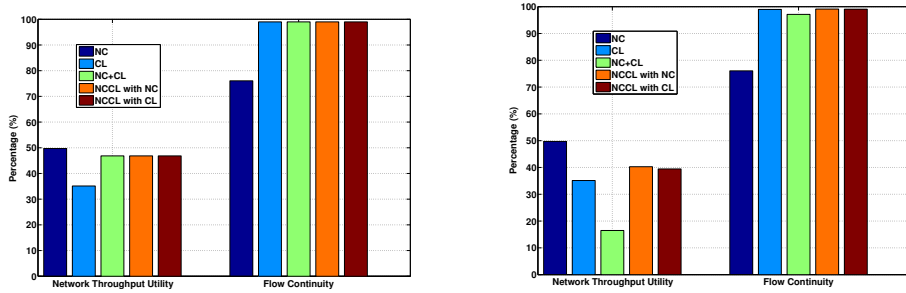
Figure 4.8: The higher the drop, the better QoE for trade-off aware adaptation with smaller throughput compromise.



(a) The smaller the drop from  $R_{c1} = 350$  kbps, the better is the throughput

(b) Trade-off aware adaptation gives better flow continuity

Figure 4.9: Comparison of throughput and flow continuity for varying channel capacity drops and varying probability of drop



(a) Destinations located such that there is no loss of over-hearing (b) Destination nodes located such that there is an intermediate loss of overhearing.

Figure 4.10: Performance Comparison in terms of QoS (Network Throughput) and QoE (Flow continuity). Note that in case there is no loss in overhearing, the switching mechanisms are not utilized .

#### 4.2.7.2 X-topology

In this section, we will present the numerical simulations to illustrate the performance of different models in X-topology.

**Performance Comparison of Models** On the basis of the performance metrics described in section 4.2.6, we now compare the models NC, CL and NC+CL along with the different switching in X-topology. We consider synthetic channel traces for a span of 1000 time slots. The channel capacity is initially 350 Kbps for first 250 seconds, then drops to 200 kbps for 250 seconds, again rises to 350 Kbps for next 250 seconds and drops for 200 Kbps for last 250 seconds. The initial rate is set to 150 Kbps.

In figure4.10a, we show the network throughput utility and flow continuity performance of different models in case the destination nodes are at a fixed geographical location, such that the overhearing capacity of the links  $R_O$  is high and the source rate never exceeds it. It can be seen that the network utility of Model NC exceeds that of Model CL, and flow continuity of Model CL exceeds that of Model NC. However, Model NC+CL provides a fair trade-off between the two. Also it can be seen, that the Model NC+CL has the same performance with or without switching because there is no loss of overhearing and there is no switching required.

In figure4.10b, we compare the performances of the models when the destination nodes are placed such that the overhearing capacity between the source and the destination  $R_O$  , is fixed at 220 kbps. Hence, the Model NC+CL switches to Model NC or Model CL. It can be seen in figure4.10b, that NC+CL with NC-based switching, outperforms NC+CL with CL-based switching in case of network throughput utility, whereas CL-based switching in NC+CL marginally outperforms the NC-based switching in terms of flow continuity. However, switching between models outperform Model NC+CL without switching in all cases.

**Location based variation of QoS and QoE in NC+CL Models with switching** In this set of simulations, we focus upon the performance of NC+CL model and illustrate the variation in performance for different locations of the destination nodes. As we consider symmetric model, we plot the geographical distribution of one pair ( $S_1-D_2$  or  $S_2-D_1$ ). We place the source node at the origin and place the unintended destination node at different integral coordinates on a space where both x-coordinate and y-coordinate vary from -500 km to 500 km.

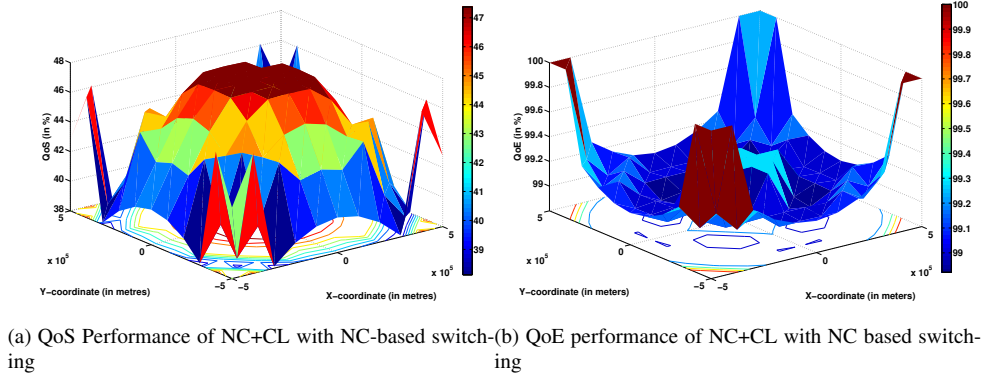


Figure 4.11: QoS and QoE variation in Model NC+CL with NC-based switching for different geographical placements of destination nodes

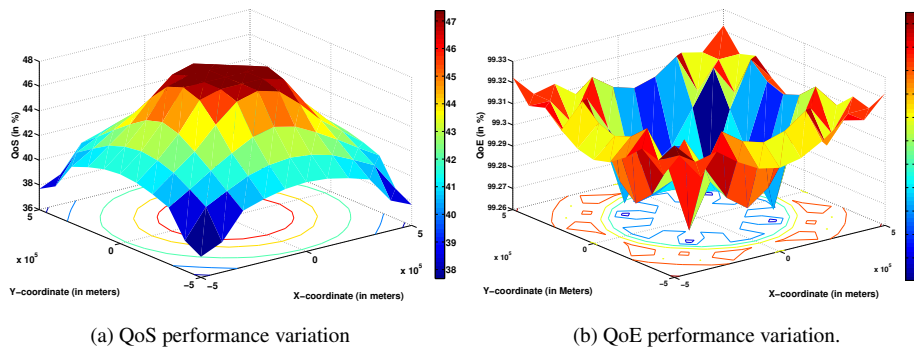


Figure 4.12: QoS and QoE variation in NC+CL with CL-based switching for different geographical placement of destination nodes

In figure4.11, we show the variation of QoS and QoE at different geographical locations with NC-based switching with NC+CL model. It can be seen that the QoE decreases towards the origin where the overhearing capacity is higher (close to the source). This is due to the fact that with higher overhearing, there are more chances of network coding, which optimizes QoS. As expected, the QoS close to the origin is highest and it reduces as the node moves away from the source.

In figure4.12, we show the variation of QoS and QoE at different geographical locations with CL-based switching in NC+CL Model. As opposed to NC-based switching, the overall QoE is higher in this case, although the gradient decreases away from center. This is because, whenever overhearing is lost, CL optimization ensures a high QoE. The QoS profile remains the same, with a lower magnitude due to no use of NC after loss of overhearing.

In order to compare the performance of the two switching mechanisms in NC+CL model, we plot the optimal switching with NC+CL model in terms of QoS and QoE, for different locations of the node. As shown in figure4.13, the dominant switching in most locations for optimal QoS is NC-based switching, whereas the dominant switching mechanism for optimal QoE is CL-based switching.

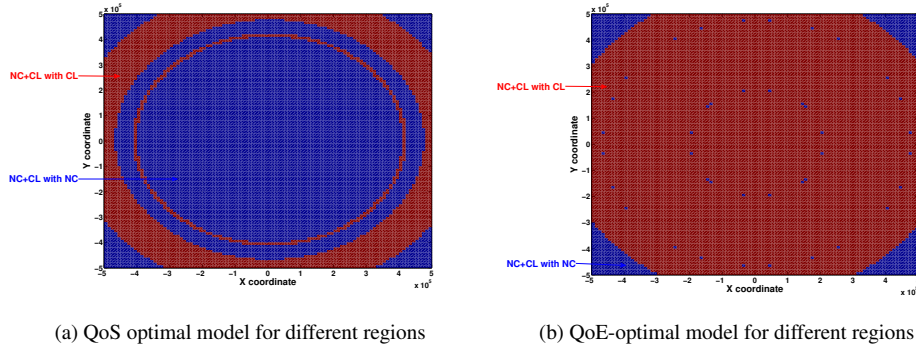


Figure 4.13: Location-optimal Model Map: the optimal switching to be applied at different geographical placement of destination node.

However, there are some exceptional places towards the boundary of the space, where NC-based switching is dominant strategy for both high QoE and QoS. This is due to the following reason: as the distance of the destination node from the origin increases, the overhearing capacity decreases. Towards the boundary of space, the overhearing capacity reduces to be less than the channel capacity after the drop has occurred. Therefore, in case the source rate reaches the overhearing capacity (which is very low), in NC-based switching, the transmission occurs at low rate. However, no delay is seen by the receiver. In case of Model CL, the rate is adapted gradually, which results in delay and packet losses when the source rate exceeds the network capacity. Therefore, in such geographical regions, with low overhearing capacity, NC-based switching outperforms CL-based switching in terms of QoE. Another exception is, for a circular ring of width 100 km, there is a region where CL based switching is optimal for both QoS and QoE. The reason is as follows: the average QoS is directly proportional to the number of packets obtained at the receiver and inversely proportional to the time taken to deliver those packets. With NC model, a reduced number of time slots is required for transmission, which leads to gain in QoS as opposed to CL model at comparable rates. However, when the rates in model CL are much higher as compared to model NC, the gain in QoS due to reduced number of time-slots does not outperform the gain in CL due to higher rate. In this region, the source rate exceeds the overhearing capacity at a value such that the amount of gain using CL is more than the reduced time slot utilized using NC.

Hence, this demonstrates that in order to supply an optimal QoS and/or QoE to the user, the service provider should consider the location of the users, and choose the optimal model for the same.

### 4.3 Maximizing throughput using cognition

We will now present another contribution towards the networking techniques for satellite networks with the aim of maximization of throughput. The aim of this study is to analyze the networking techniques using the concept of cognitive radio in satellite scenarios. The use of cognition is well-known to optimize resource utilization and enhance the performance and it is a promising technique for future wireless networking.



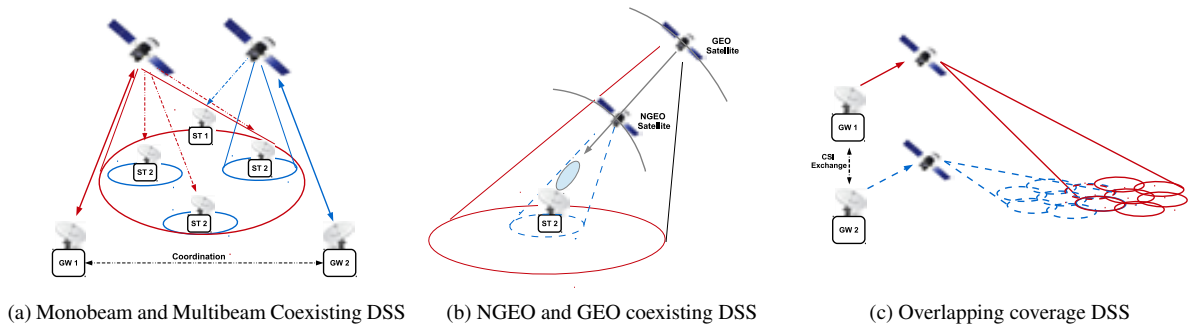


Figure 4.14: Some typical DSS scenarios

### 4.3.1 Concept of Cognition

The concept of cognition was proposed in the seminal paper of Mitola in 1999 [74]. It was further applied to wireless communications by Haykin in 2005 in [75] where he applied cognition to radio spectrum utilization and coined the term cognitive radio. Haykin defines cognitive radio as follows [75]:

*Cognitive radio is an intelligent wireless communication system that is aware of its surrounding environment (i.e., outside world), and uses the methodology of understanding-by-building to learn from the environment and adapt its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g., transmit-power, carrier-frequency, and modulation strategy) in real-time, with two primary objectives in mind:*

- *highly reliable communications whenever and wherever needed,*
- *efficient utilization of the radio spectrum.*

Cognitive radio (CR) therefore provides intelligent transmission protocols in which the communication is associated with observation of the available resources dynamically and efficiently, processing the observation and acting based on this processing in order to provide seamless services with limited resources. Cognition can therefore be naturally applied in satellite networks to adapt to changing conditions and demands.

### 4.3.2 Dual Satellite Systems

#### 4.3.2.1 Scenarios

Dual Satellite System (DSS) is the satellite network which has two satellites operating simultaneously over a coverage area in a same spectrum band. In such systems, the two satellites share spatial and spectral degrees of freedom. The DSS models a number of satellite scenarios as follows:

- **Mono-beam and Multi-beam co-existent DSS** - The traditional monobeam satellites can coexist with the multibeam satellites thereby forming a DSS. An example of such systems has been shown in figure 4.14a. The users in coverage area of these two satellites may or may not be served by both the satellites.
- **NGEO and GEO co-existent DSS** - The GEO satellites can coexist with LEO/MEO satellites (also called the Non-GEO Satellites (NGEO)) forming a DSS. An example of such systems has been shown in figure 4.14b. Since GEO satellites cover a fixed geographical area all the time, whereas N GEO satellites may cover the same area over some period of time, the two satellites are required to coexist. Such coexisting systems need to combat not only the interference due to the signals from co-located satellites, but also the inline interference due to loss of line of sight of GEO owing to N GEO.

- **Overlapping coverage DSS** - Two similar satellites having overlapping coverage areas form a DSS. Due to increasing space traffic and high demands, a number of satellites are being deployed for different specific purposes and/or are deployed in close proximities. This leads to the overlapping area of coverage. An example of such scenarios has been shown in figure 4.14c.
- **Replacement phase DSS** - During the replacement phase of an old satellite with a new satellite, there are long phases of coexistence before the old satellite is completely eliminated from the orbit. Such replacement phases form a system when the two satellites coexist thereby forming a DSS.

It is clear that the DSS can model a number of satellite scenarios and therefore, the utility of study of the DSS is imperative. There are a number of existing satellite systems which can be analyzed as DSS. For example, the Other Three Billion network proposed O3b satellites in the MEO due to the advantages of MEO satellites like reduced free space attenuation, small propagation delay and reduced in-orbit injection cost [76]. However, these satellites use parts of Ka band which is also being used by GEO satellites thereby forming a DSS as studied in [77]. Similar example of LEO and GEO coexistent network is IRIDIUM constellation which consists of a network of 66 LEO satellites coexisting with GEO satellites in different coverage areas [78]. An example of replacement phase DSS is the replacement of existent Iridium network with the Iridium NEXT constellation (to be launched in 2015) which aims to replace the older network completely by 2020 [79]. Therefore, both Iridium and Iridium Next with form a DSS for approximately 5 years. Other examples of such DSS are NGE0 satellite networks like Orbcomm and Globestar coexisting with the GEO satellites.

#### 4.3.2.2 Challenges

The DSS described above possess some inherent challenges/limitations which are required to be addressed for an optimal utilization.

1. **Limited Spectrum:** The DSS consists of two satellites sharing the spectral degree of freedom. With the high amount of space traffic, the problem of spectrum scarcity is well known [65]. Therefore, the DSS needs an efficient and optimal technique for utilizing the available spectrum. For example, by ITU RR No. 5.523A, certain parts of Ka band have been assigned to GEO satellites. However, some of the upcoming satellite networks of NGE0 satellites utilize those parts of spectrum for their transmission. This leads to spectrum sharing which is an important challenge which is needed to be addressed in DSS.
2. **High interference:** In the coexistence of two satellite networks sharing both spatial and spectral domain, there is an important problem of interference between two networks which is required to be addressed. The interference in DSS can be classified in two categories. Firstly, the interference arising from the wireless signals of the two coexisting networks. This interference is constantly present and is required to be tackled by interference mitigation techniques like interference alignment, beamforming, precoding etc. Secondly, the interference which arises in DSS due to the orbital position of the satellites leading to inline-interference. This type of interference is observed in DSS formed by GEO and NGE0 satellites when NGE0 satellite falls in line of sight of the GEO satellite and its ground terminal.
3. **Low Channel Availability:** The DSS faces an important challenge of the availability of wireless channels because two systems sharing a limited spectrum coexist. This problem is specifically important in Land mobile satellite (LMS) scenarios, although it also exists in Fixed Satellite Scenarios (FSS). In LMS scenarios, the channel conditions between the satellites and the ground terminal varies continuously, hence the joint effects of shadowing of channel and uncoordinated sharing of the available channel with the other satellite network leads to low channel availability. The low channel availability in such scenarios has been well studied in [80, 81].

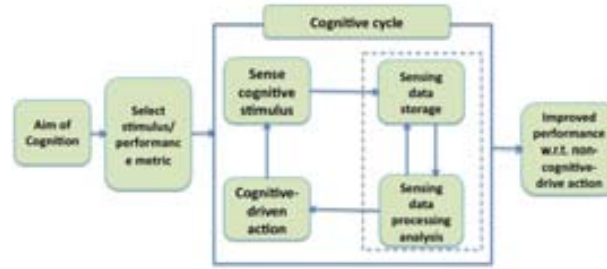


Figure 4.15: Proposed Block Diagram for Cognitive Cycle

### 4.3.3 Cognition in DSS

#### 4.3.3.1 Proposed Cognitive Cycle

The concept of cognition is applied to the communication systems using a precise cognitive cycle [82]. The cognitive cycle proposed here is shown in figure 4.15. The cognitive radio node in the network observes or senses the external environment stimulus. The sensed data is processed and based on the processed data, the most efficient action is taken in order to achieve the aim. This process leads to enhancement of performance in order to achieve the predefined aim. This cognitive cycle shows the basic steps of cognitive radio, namely, observe-process-act.

Furthermore, in cognitive scenario, users are classified as primary users (PU) and secondary users (SU). The PU are the incumbent users which are given a priority in the cognitive transmission and it is ensured that their services are not disrupted due to the presence of other systems. The SU are the users which use the cognitive techniques and adapt themselves to coexist with PU without lowering the quality of services of PU. For example, in hybrid satellite terrestrial networks, the terrestrial network is (typically) considered as PU and the satellite network utilizes the resources to provide the services to its users without compromising the services of terrestrial networks.

In the seminal paper of [74] on cognitive radio based communications, Mitola describes the implementation of communication based on cognition to be driven by the repeated stimulus from external environment. Such stimuli leads to series of 'intelligent' techniques employed for efficient and adaptive communications. The continuous stimuli observation and cognitive actions lead to a cognitive cycle. Inspired by Mitola's Cognitive cycle based communications, we propose a cognitive cycle to assess the concept of communication applied to the DSS. The cognitive cycle has been shown in figure 4.15.

The CR in satellite communication systems could be employed at either ground terminal or the satellite terminal depending on both, the requirements and the feasibility [83]. The cognitive radio (CR) is comprised of the basic cycle of observe-process-act. Each of the steps of the figure 4.15 are explained below.

- Independent of the terminal used for cognition, the cognition is employed with a predetermined aim/target, just as any intelligent or brain-empowered action is performed with an aim or a set motive. The aim for cognition further drives the CR.
- The primary requirement of cognition is the observation of external stimuli. The external stimuli observation implies monitoring the available information of the radio environment.

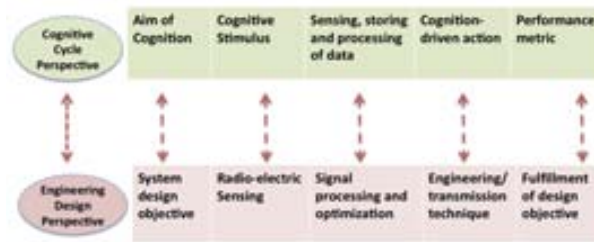


Figure 4.16: Translation of different steps of Cognitive Cycle into Engineering Design Steps

- Using the stimulus (raw data) which is sensed from the environment, the CR stores the sensed data in databases. Such data is processed and quantified using appropriate metrics. Such information is again stored in the database, analyzed and classified into different categories.
- Based on this classification, it is determined what cognitive action should be taken in order to realize the initial aim of the cognition and when such an action should be taken. Cognition also implies that the processed information is constantly monitored to determine the expected trend of the stimulus and implement the suitable action.
- Furthermore, this cognitive cycle of observe-process-act is performed repeatedly during the transmission to realize the aim of cognition. The overall improvement from this cognitive cycle results is measured on the basis of different performance metrics like spectral efficiency, power efficiency, diversity etc.

#### 4.3.3.2 Cognitive Cycle from Engineering Design Perspective

The challenges in DSS are required to be addressed in order to tap the potential of DSS in increasing both the quality of service and efficiency of resource utilization. Cognitive radio promises to be an efficient solution. This is because precisely each step of the cognitive cycle can be mapped into an engineering design requirement and can be tackled using accurate methods. The different phases of cognitive cycle proposed in figure 4.15 can be mapped to different phases of engineering design problem as shown in figure 4.16. The basic aim of cognition can be translated as the system design objective from perspective of engineering design. Furthermore, the radio-electric sensor outputs serve as cognitive stimulus which can be sensed and processed using signal processing and optimization techniques. The cognition-driven action can be further seen as the engineering/transmission techniques used for the communication leading to fulfillment of design objective. Such a precise mapping indicates that the realization of the abstract principles of cognitive cycle into practical implementation can be translated with concrete and well defined steps. It also reaffirms the applicability of cognition in satellite networks like DSS because in such systems, precise and robust protocols are required due to cost benefits and complexity involved.

There are a number of cognitive DSS studied in literature. Based on the proposed cognitive cycle, each of these cognitive DSS can be categorized to be associated with different aim of cognition. The different steps followed to achieve this aim can be associated with different phases of the proposed cognitive cycle (Table 4.1). We now develop a taxonomy of the different existing cognitive DSS and identify the phases of cognitive cycle proposed with different steps taken in each of these cognitive DSS categories.

The existent cognitive DSS can be divided into three categories:

1. Spectrum Management based Cognitive DSS
2. Interference Management based cognitive DSS

Aim of Cognition	Cognitive stimulus	Sense data, storage and processing	Cognitive-driven action	Performance metric to be improved
Increase spectrum Reuse	1. Spectrum utilization 2. Polarization utilization	1. System/Channel Modeling 2. Compute Power Spectral Density 3. Implement algorithms of analysis	1. Multiplexing 2. Scheduling 3. Transmission in Spectrum holes	1. Spectral efficiency 2. Power efficiency
Optimize interference management	Interference	1. System/Channel Modeling 2. Compute Interference Temperature 3. Classify interference levels	1. Interference alignment 2. Precoding 3. Beamforming 4. Rate splitting	1. Spectral efficiency 2. Power efficiency
Increase system availability	Channel Availability	1. System/Channel Modeling 2. Compute Shadowing threshold 3. Implement algorithms to classify shadowing	1. MIMO Techniques 2. Network Coding	Diversity (Spatial, Temporal, Angular)
<b>System Design objective</b>	<b>Radio-electric sensor output</b>	<b>Signal processing and optimization</b>	<b>Engineering/transmission technique</b>	<b>Fulfillment of design objective</b>

Table 4.1: Taxonomy of Different cognitive techniques based on the aim of cognition. The columns within each row indicate the phases of cognitive cycle related to each aim of cognition. The top row shows the steps of cognitive cycle and the bottom row indicates the engineering design equivalent of each step.

### 3. System Availability based cognitive DSS

In the next subsections, we explain each of these categories in detail.

## 4.3.4 Taxonomic Analysis of Cognitive DSS

We have seen that there are a number of scenarios where DSS arise and they can be modeled in different ways. In addition, in order to ensure an efficient coexistence of two satellite systems in a DSS, cognitive techniques have been shown to be very useful as described in section 4.3.3. In this section, we explore a taxonomy of the cognitive techniques applied to DSS.

### 4.3.4.1 Spectrum Management based Cognitive DSS

The cognitive DSS which employ cognitive techniques to deal with the problem of scarce spectrum fall in this category. The under utilization of spectrum in satellite systems has been well pointed out in [63]. The two satellite systems in DSS share a spectrum which leads to necessity of intelligent techniques to manage the limited available spectrum. The use of such cognitive techniques [83] entails all the steps of cognitive cycle and the engineering design equivalent as explained further. All the DSS models explained in section 4.3.2.1 and section ?? can be used to implement these techniques. In general, the multibeam/monobeam coexistence DSS considers monobeam satellite as PU. In N GEO/GEO coexistence DSS, GEO satellite is considered as PU.

**Cognitive Stimulus** In order to implement the cognitive techniques with an aim of spectrum management, the external stimulus sensed by the satellite system is the availability of free spectrum. This is the primary step of sensing the spectrum using an appropriate spectrum sensing technique. There are a number of spectrum

sensing techniques which are provided in the literature, not specific to DSS, but are applicable in all multi-satellite scenarios [64, 65, 66]. The three mainly used signal processing techniques for sensing are Matched Filter Detection, Energy Detection and Cyclostationary feature detection [84]. However, these techniques are limited by received signal strength which cannot be guaranteed in variable environment. In such cases, cooperative sensing is more efficient.

**Storage and Processing** After sensing the spectrum utilization (both spatially and temporally), the sensed spectrum is quantized using the metrics like power spectral density (PSD). Based on the PSD, different algorithms are applied to identify if the PSD observed indicates presence of PU with certain probability of detection. In addition, the spectral usage in space-time-frequency domain identified via the spectrum sensing techniques can be classified into three different categories [83]: (a) Black Space - spectral portion occupied by high power satellites or local interferers some of the time, (b) Grey Space - spectral portion which are partially occupied by low power interferers and (c) White Space - spectral portion free of RF interferers except for ambient noise consisting of natural and artificial forms of noise. In the different DSS scenarios, the SU transmits in preferably white space using the opportunistic transmission. To transmit in grey or black spaces, an appropriate interference management technique is required, which will be discussed in next subsection. It should be noted that the opportunistic transmission in spectral holes also includes the transmission in 3D space spectral holes. This implies that the spectral holes can be identified depending on transmission angle and polarization, and such parts of spectrum can also be opportunistically utilized for cognitive transmission.

**Cognitive technique** The cognitive-driven action or the transmission techniques for spectrum management can be studied under two categories:

1. Opportunistic Transmission: A very basic and very effective cognitive technique applied to DSS is the opportunistic transmission by secondary user in the spectrum holes of the primary user. More specifically, the CR terminal at the SU utilizes the radio spectrum on a non-interfering basis in a spatio-temporal manner by giving a higher priority to incumbent spectral users.

An implementation of opportunistic transmission technique has been studied in [85] using the concept of cognitive beamhopping in overlapping coverage coexistent DSS. The primary satellite network is allowed to transmit at a predetermined spectrum bandwidth. However, there is dynamic spectrum sensing applied at the SU and the idle bands are assigned to the SU. In a traditional beamhopping system, each active beam uses full frequency instead of fractional frequency reuse as in the conventional multibeam systems. This property is exploited in the cognitive beamhopping DSS. Since only a certain fraction of total available beams are active in a particular time slot, the unused frequencies can be used by secondary satellite network of DSS. Basically, the primary satellite system is a beamhopping system with larger beams. The secondary satellite system can also be considered to be a beamhopping system with smaller beams and lower peak power.

2. Cooperative Transmission: Conventional techniques of spectrum management are cognitive in nature, namely multiplexing and scheduling. In case of cognitive DSS (e.g.. in case of N GEO/GEO coexistence), the spectrum can be shared between PU and SU to an extent that is within tolerable limits of PU. The resources can be shared in frequency (Frequency Division Multiplexing) or time (Time Division Multiplexing). Furthermore, in case of monobeam/multibeam coexistence DSS, in the uplink, spectrum assignment based on the demands of the users which adds another level of cognitiveness to spectrum management. Such a case has been very well presented in [86] and can be applied straight-forward to the DSS.

The output of these spectrum management cognitive techniques is a reduction in spectrum scarcity which can be measured using the metrics such as spectral efficiency or power efficiency.

#### 4.3.4.2 Interference Management based Cognitive DSS

The cognitive DSS which employs the cognitive techniques with the aim of interference management are categorized under this subsection. The spectral coexistence of two satellite systems in DSS can be modeled as Cognitive Radio (CR) network with interference channels between primary and secondary systems. A number of interference management techniques are employed in cognitive DSS in order to reduce/eliminate the interference of SU towards the PU while maintaining the QoS of the SU. It should be pointed out here that a cognitive system should be also able to embrace interference. Such techniques can be also studied in-line with our proposed cognitive cycle as follows:

**Cognitive Stimulus** The external stimulus used to assess the implementation of cognitive techniques to combat interference is the radio-electric sensing of interference at the primary user node. This is precisely done by assessing the radio-electric signals obtained at the PU. These signals are observed over stipulated domains of time/space/frequency before proceeding to the next step.

**Storage and Processing data** The observed signals are further stored and processed to assess the amount of interference in at the PU. This processing involves quantization of observed signals to identify interference. The FCC task force report [63] defines the metric of interference temperature to model the interference faced by the PU from the SU. Hence in cognitive DSS, the interference management techniques aim to ensure that the interference temperature at the PU does not exceed a predetermined threshold.

**Cognitive technique** The interference management techniques considered in cognitive DSS can be studied under two sub-categories: Interference management via flexible/beamforming system: The interference management depending on the domain used has been extensively explored in [87] using beamforming techniques. These techniques are valid for all the scenarios discussed above in downlink transmission. In [88] and [77], these techniques are studied for DSS. Note that in these techniques, we assume that the interference management is used at both primary SAT1 and Secondary SAT2.

1. Interference Alignment: The basic principle of IA is based on aligning the interference on a signal subspace with respect to the non-intended receiver, so that it can be easily filtered out by sacrificing some signal dimensions. The fundamental assumptions which render IA feasible are that there are multiple available domains (space, frequency, time or code) and that the transmitter is aware of the CSI towards the non-intended receiver. Interference Alignment (IA) in satellite system was initially studied in [89] for downlink scenarios. It was also proved in this work that subspace interference alignment is feasible in downlink satellite scenarios using only certain geometrical arrangement constraints. IA was later studied in uplink transmission because in downlink, in general, the coverage of the satellites is geographically large and it is complex to align the signal from satellite in certain dimensions for the terrestrial receivers. DSS had later been discussed in [88] for uplink transmission in which IA is applied to multibeam satellite terminals to mitigate interference towards monobeam satellite. Further, depending on the degree of cognition, different IA techniques are possible:
  - (a) Static IA: The simplest case of IA employed at the ST2's is based on the initial concept of IA introduced in [90] according to which a non-zero reference vector  $\mathbf{v}$  is chosen along which the interference should be aligned. The selection of alignment direction can be predetermined with signaling from intended gateway. The interference aligned across  $\mathbf{v}$  can be removed using zero-forcing filter. This basic IA approach leads to a unit multiplexing gain and is not optimal.
  - (b) Coordinated IA: Adding more amount of cognition to IA leads to coordinated IA, where the primary and secondary systems coordinate to exchange CSI information and the alignment vector dynamically depending on best conditions.

- (c) Uncoordinated IA: In this approach cognition is on side of PU and the primary and secondary systems do not coordinate. The ST2s choose the  $\mathbf{v}$  to maximize their throughput and subsequently, SAT1 senses the  $\mathbf{v}$  and applies the appropriate filter.

Further, in [91] IA techniques with frequency packing have been suggested which involves reducing spacing between adjacent signals in the frequency domain, while employing advanced techniques for suppressing or exploiting the additionally reduced interference.

2. Interference mitigation by power control: The DSS also faces another type of interference namely inline interference, which is particularly significant in N GEO-GEO coexistence scenarios. Cognitive DSS employ inline interference mitigation by power control techniques. The GEO satellite system is considered the primary system and the N GEO satellite system is considered as the secondary user. Such scenarios have been shown in figure 4.14b. The interference mitigation is tackled by formulating optimization problems for uplink and downlink separately [77]. In uplink transmission, the optimization problem for power allocation requires the N GEO earth station (ST2) to control its transmit power such that the interference at the GEO satellite terminal  $(I/N)_{sat1}$  is below the threshold level while simultaneously ensuring its own QoS. In downlink transmission, the optimization problem for power allocation requires the N GEO satellite (SAT2) to transmit at power levels such that the N GEO terminals (ST2) on ground obtain the desired  $(C/N)_{st2}$  ratio whereas the GEO terminals on the ground (ST1) have the  $(I/N)_{st1}$  below the threshold level. Furthermore, the on-board power constraint of the N GEO satellite should be satisfied.

#### 4.3.4.3 System Availability based Cognitive DSS

Assessing how to improve the availability is a problem that has to be often tackled whenever two systems sharing a limited spectrum coexist, as is the DSS case. Developing efficient techniques becomes even of more importance when the scenario is mobile, i.e. LMS-DSS, which means that besides of sharing the spectrum users have limited and time varying visibility of the satellites, i.e. limited and varying spectrum holes from which to send or gather the data. Although there exist no specific works assessing the LMS-DSS scenario, it is well known that Network Coding (NC) and Multiple Input Multiple Output (MIMO) techniques can help to improve the availability in mobile scenarios. These techniques are based on the idea of embracing interference rather than mitigating it as pointed out in section 4.3.4.2. We have explored in detail the implication of LMS-DSS with these techniques in [73] which interested readers may refer to.

## 4.4 Concluding remarks

### 4.4.1 Contributions over state-of-the-art

In this chapter, we have taken a very elaborate view of the different networking techniques applied beyond purely physical layer at the PHY/MAC layer. At this layer, we considered the target wireless scenarios to be either general relay scenario or satellite scenarios to focus on accurate wireless model for our study. These networking techniques have been studied to be applied in coherence with two aspects of communications.

- The first part of our studies has shown application of network coding at link layer along with QoE-driven cross application/transport layer optimization. This work develops a novel scheme to implement QoE-driven cross-transport/application layer optimization using network coding at the network layer for video transmission in a wireless relay network. This work also presents novel trade-offs which are a crucial consideration for design of such schemes. Another important contribution of this work is the use of novel geographical distribution based tools to optimize the joint implementation of such schemes whose potential has not been exploited in literature so far.



There are research work aiming at joint implementation of different types of cross layer optimization schemes with network coding [55, 54]. Our work departs from the state-of-the-art due to the use of the cross transport/Application layer optimization which is transparent to the network coding at lower layers. Furthermore, we have shown the implication of the use of QoE-driven cross-layer adaptation and the effective end-user perception with network coding, which is a novel contribution. We illustrate that the joint implementation of CL and NC in X-topology and in bent-pipe topology, using Model NC+CL leads to the existence of a QoE-throughput trade-off. This trade-off is crucial for choosing the optimal parameters to achieve desired performance. Furthermore, we also used the location of the terrestrial users to optimize the throughput by adapting the transmission rate to the user-location constraint. The extra information conveyed due to user location further consolidates the overall performance in terms of QoE and throughput. Via numerical simulations, we illustrate that for each location of the user, an optimal switching with NC+CL model for QoS or QoE can be identified, which should be adopted by the service provider for maximizing its user benefits. This work provides a primary insight into the joint practical implementation of QoE driven CL and NC and demonstrates that the interplay between location based information and the novel concepts of NC+CL should be exploited to maximize the transmission quality in future scenarios.

- The second part of our studies entails a very extensive and critical review of the various cognitive techniques involved in satellite channels primarily over mac/physical layer. We focus on particular satellite networks in this case, namely dual satellite systems, which allows for a large arena to apply cognition in the system. This part of the chapter has provided an introductory overview of dual satellite systems, their modeling and justified the relevance to apply cognitive communication techniques to such systems. In particular, the existing trends of the research in this area have been critically reviewed and classified based on different system design objectives. To the best of our knowledge, this is among the first surveys on the topic of Dual Satellite systems. An important contribution of this chapter is to provide a concrete mapping between the phases of cognitive communication cycle and the corresponding phases of engineering design applied in dual satellite system thereby bridging the gap between the abstract and the practical design of cognitive techniques.

#### 4.4.2 Publications

The contributions presented in this chapter have resulted in following three publications:

##### *Book Chapter*

- **S. Gupta**, M. A. Vázquez-Castro, R. Alegre-Godoy, "Dual Satellite Systems", Chapter in the book *Cooperative and Cognitive Satellite Systems*, Elsevier 2014.

##### *Conference Proceedings*

- **S. Gupta**, M. A. Pimentel-Niño, M. A. Vázquez-Castro, "Joint network coded-cross layer optimized video streaming over relay satellite channel", In *3rd International Conference on Wireless Communications and Mobile Computing (MIC-WCMC)*, Valencia, June 2013.
- **S. Gupta**, M. A. Vázquez-Castro, "Location-adaptive network-coded video transmission for improved Quality-of-Experience" In *31st AIAA International Communications Satellite Systems Conference (ICSSC)*, Florence, (2013).

#### 4.4.3 Future lines of work

This chapter provides a comprehensive insight into the two aspects focussed upon. A number of open challenges still remain to be solved which form the possible lines of future research.

- The trade-off study emerging from the joint implementation of NC and QoE driven cross transport and application layer rate adaptation, can be further generalized for various topologies thereby building a concrete structure to characterize such system performances.
- The location-aware transmission in the joint implementation is a theoretical study however, a number of practical aspects of the same remain to be seen like antenna structures for overhearing, realistic overhearing conditions, etc.. Such a study coupled with the model presented forms an interesting future line of research.
- The cognitive DSS techniques have a number of technological aspects of actual implementation which have not yet been dealt with in the literature. These aspects will ultimately provide actual gains of cognition based designs and hence after this conceptual/academic research in cognitive DSS, it is time for looking into technological/implementation issues.
- In addition, using the technological assessment of cognition based design implementation, a future direction of work is the optimal combination of the three aims of cognition presented in this chapter.
- Another important open direction to be explored is the security issues involved in inter-system transmission and selection of a cognitive strategy which satisfies the security demands.

Overall this chapter is a comprehensive study of the future generation wireless networks, like satellite networks, imbibing the aspects from different independent techniques to function in coherence and provide effective performance improvements over state-of-the-art.

## Chapter 5

# Networking at Transport -Application layer

### 5.1 Introduction

Having explored the networking techniques at the lower layers, we now focus on the transport and application layer. These layers are the ones which are perceived by the end-user more closely and it is important to maximize the effective service provided by these layers to the users while ensuring an optimal resource utilization as well. Therefore, in this chapter we study a particular application, aim to maximize not only the throughput but also the end-user experience in using this application.

#### 5.1.1 Objective and technique in overall framework

**Objective:** *“Maximizing the joint utility using game theoretical tools”*

Recently there has been a remarkable development in video transmission over hypertext transfer protocol (HTTP) [92] which uses transport control protocol (TCP) and its variants [93]. However, video transmission over TCP is the best-effort transmission which does not account for specific characteristic requirements of multimedia applications, particularly video applications, like disruptions in user-perceived video quality, overall user satisfaction etc. In order to address different specific characteristics, several cross-layer optimization techniques have been proposed to optimize video transmission over wireless media [22]. These techniques are based on the optimization of different aspects of the video transmission like optimization of the Quality of Service (QoS) [94], optimization of Quality of Experience (QoE) [95, 96, 59], optimization of a general application oriented user satisfaction [97] etc.

An important demand of the users of the video applications is to be provided a desired quality of experience of the perceived video. The QoE-driven cross-layer optimization accounts for video transmission such that the perceived video quality at the end-user is optimized in terms of minimizing both distortions in video quality and disruptions in video playback. This is because the QoE at the end user is composed of temporal structure and spatial structure of the video. The temporal structure involves the continuity in the video observed at the end-user without freezing or disruptions, implying timely playback of the video. The spatial structure involves the avoidance of formation of artifacts due to packet losses at the end user [98]. An inherent benefit of focussing on QoE-driven cross layer adaptation is that it generalizes the TCP based transmission and hence can be adopted by the existing systems. However, the wireless nodes involved in the video transmission are largely distributed and are autonomous devices. Therefore, the practical dynamics of

the interaction between them is governed by their independent behaviors which maximize their individual video quality. This independent behavior in the interactions affects the quality of the video observed at the end user.

Our objective is to assess the affect of such interactions. In particular, in a video exchange scenario between two selfish autonomous users, the ability of a user to obtain a desired quality video depends on the optimization performed by the other user. However, in order to perform such an optimization, the other user incurs some cost. This leads to a natural competition between the users which optimize the quality of video they transmit selfishly such that they incur minimum cost but are still able to obtain a desired quality video from other user. This chapter addresses such competitive video exchange between two users.

**Technique:** *Game Theory*

A relevant framework to model this type of interactions is provided by game theory. A game theoretical framework also helps to analyze the prospective benefits of possible cooperation among the nodes. The basics of game theory have been explained in Chapter 2. We will develop an extension of the basic framework to develop the game theoretical tool in this chapter.

**Contributions**

The main contributions of this work can be stated as follows[99, 100]:

- We propose a novel video exchange game between two nodes where each node seeks to optimize a QoE-dependent metric to adapt the video transmission rate to varying channel conditions while minimizing the cost. Particularly, we identify a natural game in which the players use the QoE-driven rate adaptation algorithm as their strategy. There is no external/artificial price mechanism used to control the outcome of the game.
- We analyze the possible outcomes of the game when it is played once and when it is played repeatedly and thereby establish the clear conditions required for cooperation to be sustained for long duration using the tools in [101]. It is shown that when the game is played once or a finite number of times, the players have no incentive to cooperate and rate adaptive video transmission is not sustained. However, a non-trivial solution is obtained only when the game is played repeatedly and the players are unsure of the final stage of the game. We illustrate that in such an interaction, a heterogeneous quality video exchange between two users can be sustained, depending on the length of the interaction.
- *Game Theoretical perspective:* An important contribution of this work is that we not only identify the region of sustainable action profiles achievable by the users in an infinite horizon repeated game, but also identify which of these profiles are likely to be achieved by the users over a long duration of time. We use a two-fold characterization to identify such profiles: firstly, we model the tolerance of the users using a tolerance index which measures the ability of the users to agree to asymmetric action profiles. Secondly, we show the efficiency of these achievable profiles under the light of pareto optimality. Using this two fold characterization of tolerance and pareto efficiency of all the sustainable action profiles, we identify the action profiles (and consequently the video QoS and QoE) likely to be obtained by the heterogeneous users.
- *Video Transmission Perspective:* Our approach allows us the flexibility to tailor the solution to varying needs of the users by providing different weightage to Quality of service and Quality of experience of the video in the payoff formulation. In addition, we establish a framework which allows the achievability of different video QoS and QoE controlled by flexible system parameters in order to control outcome of the game played by selfish users.

- *Satellite Case Study*: We consider a special case of exchange of video applications over satellite channels and with the help of repeated games' framework, we identify the best performing QoE-driven rate adaptation strategies for adaptive video transmission in different satellite networks. Using this case study, it has been shown that the speed at which the users adapt the transmission rate to the channel conditions is influenced by the two fold effect of expected length of interaction between the users and the expected channel conditions. Consequently, the best strategies are chosen by the users based on available parameters which affects the QoS and QoE of the video achieved [100].

### 5.1.2 Outline of the chapter

The rest of the chapter is organized as follows: in section 5.2, we describe the system model and the QoE-driven rate adaptation performance metrics which are used in the chapter. We present the game theoretical framework for video transmission in section 5.3. Based on game theoretical insights, possible solution selection is discussed in 5.4. The numerical results are presented in section 5.5 followed by conclusions in section 5.6.

## 5.2 System Model and QoE-driven Rate Adaptation

### 5.2.1 System Model

In this section, we describe the system model and the QoE-driven rate adaptation scheme used. We also define the performance metrics which will be considered in the analysis further.

#### 5.2.1.1 Network Topology and Channel Model

Our focus is on wireless video transmission systems in which two terrestrial users want to exchange their videos via a relay as explained in Chapter 4 with bent-pipe topology. Such scenarios model situations like video streaming between two parties across geographically separated areas. The node A (or B) transmits the video packets to the relay and the relay forwards these packets to node B (or A) as shown in figure 1. The wireless channel between any two nodes has variable capacity to transmit packets across the nodes depending on various factors like environmental conditions (rain, cloud, etc.), congestion due to other users, etc. We model channel links between the nodes to have a certain throughput capacity which varies with time. The throughput capacity is the maximum number of packets that can be transmitted between any node  $k$  and node  $l$  at time  $t$  and is denoted by  $R_{kl}(t)$ . It is measured in packets per second (pps). As shown in figure 1,  $R_{AB} = 0$  for all  $t$ . Throughout this chapter, for simplicity, at any time instant  $t$ , all links are assumed to have equal throughput capacities  $R_c(t)$ , that is,  $R_{kl}(t) = R_c(t)$ ,  $\forall kl \in \{AR, BR, RA, RB\}$ . Each time instant is described by  $t = n\Delta$ , where  $n \in \mathbb{Z}^+$  and  $\Delta$  is the time interval size, i.e., a precision parameter. We will therefore, refer to instantaneous quantities using the integer value  $n$ . We assume in this work, that the channel conditions remain constant over a large period of time, therefore,

$$R_c(n) = \begin{cases} R_1, & n \leq N_1 \\ R_2, & N_1 < n \leq N_2 \end{cases} \quad N_1, N_2 \in \mathbb{Z}^+ \quad (5.1)$$

for large values of  $N_1, N_2$ .

#### 5.2.1.2 Cross-layer Design Model

We now describe the cross application/transport layer design to transmit the video content. The video content is transmitted using UDP (User Datagram Protocol) over RTP (Real-Time Transport Protocol). The source

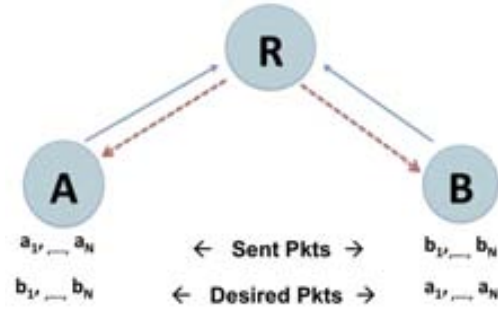


Figure 5.1: Network Topology. The nodes A and B exchange information via R. The arrows indicate the links with positive channel capacity.

node is equipped with a codec responsible for compression of the video content. The output rate of the codec can be reconfigured to deliver a target bit-rate. Therefore, at the source node, the output rate can be adjusted as desired. This output rate from any node  $k \in \{A, B\}$  at time  $t = n\Delta$  is denoted by  $r_k(n)$  and it is measured in packets per second. The packets are further passed down to the network layer maintaining coherence with standard protocol stack. The rate of transmission of video payload is adapted to the network conditions by optimization of the QoE perceived at the end-user. With the use of Real Time Control Protocol (RTCP) signaling, the source node collects the feedback information about the Round Trip Time (RTT) from the destination. The RTCP feedback signaling provides the source node with information to re-configure the codec rate to suit the target needs.

### 5.2.1.3 Buffer and Delay Model

We will now illustrate how the transmission buffer, delay and packet losses are modeled in this chapter. Each of the source nodes (A and B) maintains a transmission buffer of size  $B_c$  packets. The number of packets stored in transmission buffer at time  $t = n\Delta$  is given by  $B(n)$ . The buffer is maintained using a First In First Out (FIFO) queue. Let the source node  $k$  transmit the packets at time  $t = n\Delta$  at the rate  $r_k(n)$ . The source node is unaware of the channel link capacity  $R_c(n)$ . Assuming that the buffer is initially empty, there are two cases. If  $r_k(n) < R_c(n)$ , all the packets are transmitted through the channel with rate  $r_k(n)$ . If  $r_k(n) > R_c(n)$ , then packets are transmitted with a rate of  $R_c(n)$  through the channel link. From the remaining  $r_k(n) - R_c(n)$  packets per time slot,  $B(n) \leq B_c$  packets are stored in the empty space in the transmission buffer, while the remaining packets are lost. In the next time slot, the packets stored in the buffer are transmitted through the channel. These packets face a delay in reaching the destination. The same process as above is applied to rest of the new packets to be transmitted, but now the effective channel capacity is  $R_c(n) - B(n - 1)$ .

Using the above system model, we will now describe the Quality of Experience-driven rate adaptation model which is used in the rest of the chapter. Furthermore, we describe the pertinent performance metrics which take into account the applicability of the system model.

## 5.2.2 QoE-driven Rate Adaptation and Performance Metrics

The cross layer rate adaptation model shows the framework to optimize the output codec rate of a source node with feedback from the network in the form of RTCP signals. The rate adaptation can be based on optimizing various aspects like packet losses, throughput, etc. We will focus on the adaptation based on optimizing Quality of Experience of the video obtained by the end-user. QoE of the end user can be quantified using different metrics in spatial domain like SSIM (structural symmetry) and temporal domain like flow continuity

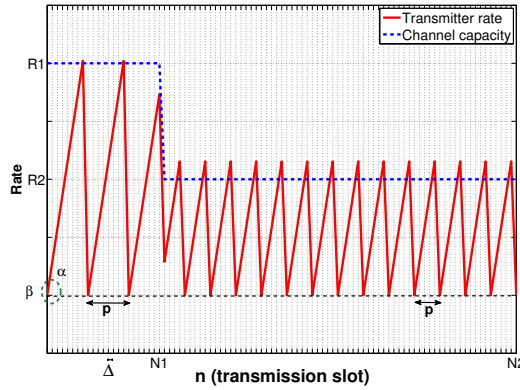


Figure 5.2: Rate adaptation curve. The red line indicates the rate  $r_k(n)$ . The blue line indicates the channel capacity. Angle  $\alpha$  remains constant for all time  $n$ , whereas period  $p$  depends on channel capacity.

. In this work, we focus on using flow continuity of the end-user video as a measure of the QoE in the time domain. This is done in order to ensure that the end-user is able to view the video without any freezing. The rate is adapted to the channel conditions as follows. The source begins the transmission at a certain initial rate. The rate is increased linearly with time, as long as the source observes no delay. However, as soon as the source observes a delay, it reduces the rate to its initial value. The process is again repeated. This adaptation is shown in figure 5.2. Mathematically, any source  $k$  begins the transmission at rate  $r_k(1) = \beta < R_c(1)$ , where  $\beta \in \mathbb{R}^+$ . With each RTCP received, the rate is updated by

$$r_k(n, \alpha_k) = \beta + \delta r_k(\alpha_k) \quad (5.2)$$

where the increment in rate is

$$\delta r_k(\alpha_k) = \alpha_k(n \bmod p)\Delta$$

such that  $\Delta$  is the time interval size,  $\alpha_k \in (0, \tan(\pi/2))$  is the slope of the rate adaptation curve and  $p$  is the time period of the waveform in figure 5.2 given by

$$p = \begin{cases} \left\lceil \frac{R_1 - \beta}{\alpha_k \Delta} \right\rceil + 1 & n \leq N_1 \\ \left\lceil \frac{R_2 - \beta}{\alpha_k \Delta} \right\rceil + 1 & N_1 < n \leq N_2 \end{cases}$$

Note that this rate adaptation is done at the source node  $k$ , and it is adapted according to the delay statistics observed at node  $l \neq k$ . Henceforth, we denote any node other than  $k$  as  $-k$ . We now describe the different performance metrics using this model which will be used consequently to analyze the performance of the scheme.

### 5.2.2.1 Quality of Service metric

The network utility is defined as the ratio of the network capacity utilized to transmit packets to a node and the total network capacity provided by the network. The number of packets obtained at node  $k$  depends on the rate of transmission at the other node  $r_{-k}(n, \alpha_{-k})$ . However, the packets obtained at the node  $k$  cannot exceed

$$f_k^{QoS}(\alpha_{-k}) = \frac{1}{2} \left\{ \frac{1}{N_1} \sum_{n=0}^{N_1} \mu(R_1, r_{-k}(n, \alpha_{-k})) + \frac{1}{N_2 - N_1} \sum_{n=N_1+1}^{N_2} \mu(R_2, r_{-k}(n, \alpha_{-k})) \right\} \quad (5.3)$$

$$f_k^{QoE}(\alpha_{-k}) = 1 - \left\{ \frac{\sum_{n=0}^{N_1} \varphi((r_{-k}(n, \alpha_{-k}), R_1) + \sum_{n=N_1+1}^{N_2} \varphi(r_{-k}(n, \alpha_{-k}), R_2))}{N_2} \right\} \quad (5.4)$$

the instantaneous channel capacity. Therefore, the instantaneous network utility at time  $t = n\Delta$  is given by

$$\mu(R_c, r_{-k}, n, \alpha_{-k}) = \frac{\min(R_c(n), r_{-k}(n, \alpha_{-k}))}{R_c(n)}$$

where  $R_c(n)$  is given by eq. (5.1) and  $r_{-k}(n, \alpha_{-k})$  is given by eq. (5.2). We define the averaged network utility from  $n = 0$  to  $n = N_2$  in providing the video to node  $k$  as the *Quality of Service* at node  $k$  and it is given by eq. (5.3). It can be seen that the QoS of node  $k$  depends on the rate adapted by the other node and consequently on  $\alpha_{-k}$ .

### 5.2.2.2 Quality of Experience metric

The flow continuity is defined as the probability of the delay in the network to exceed the threshold delay. The delay in the network leads to video packets arriving at the end user in more than expected time, leading to a visible freezing of the displayed video and degrading the quality of experience. In this chapter, flow continuity is chosen as the QoE metric which is a simplified choice for the analysis. Assuming the threshold delay to be 0 (allowing no freezing of video), we notice in the buffer and delay model described in section 5.2.1.3 that the delay is observed at the node  $k$  when the rate  $r_{-k}$  exceeds the channel capacity. Let us define a delay counter function  $\varphi(\cdot)$  which determines if there is an instantaneous delay in the network or not. It takes the instantaneous rate and the channel capacity as the input and generates 0 if there is no delay or 1 if there is a delay in the network. It is defined as

$$\varphi(r(n, \alpha), R) = \max(0, \text{sgn}(r(n, \alpha) - R))$$

where

$$\text{sgn}(r(n, \alpha) - R) = \begin{cases} -1, & \text{if } r(n, \alpha) < R \\ 0, & \text{if } r(n, \alpha) = R \\ 1, & \text{if } r(n, \alpha) > R \end{cases}$$

Therefore, we can write the total number of events when the delay occurs as

$$\sum_{n=0}^{N_1} \varphi((r_{-k}(n, \alpha_{-k}), R_1) + \sum_{n=N_1+1}^{N_2} \varphi(r_{-k}(n, \alpha_{-k}), R_2)$$

We define the *Quality of Experience* metric as the average flow continuity of the network and it is given by eq. (5.4). We note that the flow continuity at  $k$  is a function of the rate adaptation gradient at the other node  $\alpha_{-k}$ .



### 5.2.2.3 Cost

In order to transmit packets, the sender node  $k$  incurs a cost of transmission due to the power usage, hardware requirements etc. This cost is dependent on not only the magnitude of resources used for transmission but also on the gradient of increment in the rate. This is because the higher the gradient in rate increment, the more is the difference between two consecutive instantaneous rates. This asserts a higher energy demand on the resources to make a sharper change in the rate leading to higher cost. In order to model this cost of transmission, for simplicity, we use a logarithmic function of the rate adaptation slope or  $\alpha$  given by

$$f_k^{COST}(\alpha_k) = \log_2(1 + \arctan(\alpha_k)) \quad (5.5)$$

Note here that the cost of transmission incurred at the node  $k$  is determined by the rate adaptation gradient used by the node  $k$ .

With these metrics, in the next section, we will define the utility function given by

$$w_1 f_k^{QOS}(\alpha_{-k}) + w_2 f_k^{QOE}(\alpha_{-k}) - w_3 f_k^{COST}(\alpha_k)$$

where  $w_i \in \mathbb{Z}$  and it will be optimized further in the game-theoretical analysis.

In the next section, we will illustrate the video exchange between the nodes  $A$  and  $B$  using the QoE-driven rate adaptation described above when the nodes  $A$  and  $B$  are selfish entities. Note that the rate adaptation is applied at the transport/application layer at the source nodes and the adaptation is done on the basis of the video quality at the destination nodes. Also note that since the channel capacity  $R_{AR} = R_{RB}$ , therefore, all packets transmitted from  $A$  to  $R$  are transmitted from  $R$  to  $B$  and vice versa. Therefore, we can assume the video exchange rates between  $A$  and  $B$  to be transparent to the relay because the relay does not alter the rates. This is a realistic assumption in several scenarios such as networks with relay satellites having no on-board processing as bent pipe network. With slight abuse of notation, in rest of the study, we refer to the transmission between  $A$  and  $B$  implying the transmission between  $A$  and  $B$  via relay  $R$ .

## 5.3 Game Theoretical Framework for QoE-driven Adaptive Video Transmission

### 5.3.1 One-shot Non-cooperative Game

It can be seen from section 5.2.2 that the quality of the video obtained by a user, say  $A$ , depends on the rate of transmission of user  $B$  and vice versa. Additionally, in order to transmit the video packets to user  $A$ , user  $B$  incurs a cost and vice versa. Therefore, for two selfish users  $A$  and  $B$ , there is a conflict of interest as both users want to incur minimum cost (affecting the video quality of other user) and also obtain good quality of video themselves (affected by the rate of transmission by other user). Therefore, there is an interaction arising naturally among the two nodes which is modeled using game theory.

#### 5.3.1.1 One-shot Game Formulation

We define a one-shot non-cooperative game by the following tuple:

$$\mathcal{G}_{\mathcal{O}} = \{\mathcal{P}, \{\mathcal{A}_k\}_{k \in \mathcal{P}}, \{u_k\}_{k \in \mathcal{P}}\} \quad (5.6)$$

where  $\mathcal{P}$  refers to the set of players which selfishly maximize their own payoffs given by  $u_k$  by choosing their actions from the action set  $\mathcal{A}_k$  for every  $k \in \mathcal{P}$ . With the QoE-driven rate adaptation model to exchange video, we define the following component sets of the game tuple  $\mathcal{G}_{\mathcal{O}}$ :

Players Set  $\mathcal{P}$ : The set of players is given by  $\mathcal{P} = \{A, B\}$  which are the two nodes  $A$  and  $B$  that exchange the video packets.

Action Set  $\mathcal{A}_k$ : The set of actions that can be taken by the player  $k$ , where  $k \in \{A, B\}$ , is given by  $\mathcal{A}_k = [0, \lambda \frac{\pi}{2}]$  and  $\lambda \in (0, 1)$ . The action chosen by  $k$ -th player from  $\mathcal{A}_k$  is denoted by  $\alpha_k$ . Hence the player can choose the gradient of rate adaptation as its action in order to maximize its own benefits. The factor  $\lambda$  ensures that the gradient is always  $\alpha < \frac{\pi}{2}$  to prevent infinite increase in rate.

Utility  $u_k$ : The utility, or payoff function of the  $k$ -th user is defined as follows

$$u_k(\alpha_k, \alpha_{-k}) = w_1 f_k^{QOS}(\alpha_{-k}) + w_2 f_k^{QOE}(\alpha_{-k}) - w_3 f_k^{COST}(\alpha_k) \quad (5.7)$$

The utility of the  $k$ -th player is a function of the action  $\alpha_k$  taken by the node  $k$  and the action  $\alpha_{-k}$  taken by the other node. The components of utility function are as follows: the utility is a joint measure of the quality of video experienced by the node  $k$  and the cost incurred in the transmission of video to the other node. The video quality at node  $k$  is affected by the throughput and the flow-continuity given by  $f_k^{QOS}(\alpha_{-k})$  and  $f_k^{QOE}(\alpha_{-k})$ , using (3) and (4) respectively, which are both determined by the action of the other node  $\alpha_{-k}$ . Since the node  $k$  wants to maximize its quality of video perceived, the utility is directly proportional to both throughput and flow continuity. However, the node  $k$  incurs a cost  $f_k^{COST}(\alpha_k)$ , given by (5), to transmit video to the other node, which is inversely proportional to its utility. Finally the weight  $w_i \in (0, 1)$  assigns the appropriate dimensions to different factors such that  $\sum_{i=1}^3 w_i = 1$ .

A natural solution concept of this game is the Nash Equilibrium as described in the next subsection.

### 5.3.1.2 Nash Equilibrium for one-shot game

The Nash Equilibrium (NE) of a non-cooperative game  $\mathcal{G}$  is defined as a set of actions of the players  $(\alpha_k^*, \alpha_{-k}^*)$ , from which no player has an incentive to deviate unilaterally. Hence, the selfish rational players are expected to play the action at NE irrespective of the other players action. In the next proposition, we describe the NE for the QoE-driven adaptive video exchange game.

**Proposition 13.** *The unique Nash Equilibrium of one-shot QoE-driven adaptive video exchange game  $\mathcal{G}_{\mathcal{O}} = \{\mathcal{P}, \{\mathcal{A}_k\}_{k \in \mathcal{P}}, \{u_k\}_{k \in \mathcal{P}}\}$  is given by  $(\alpha_1^*, \alpha_2^*) = (0, 0)$ .*

*Proof.* Consider the utility function defined in eq. (5.7). Node  $k$  can maximize its utility by minimizing the cost given by eq. (5.5). The cost is a logarithmic function which is an increasing function of  $\alpha_k$ , and therefore, is minimized at  $\alpha_k = 0$ . The node  $k$  has no control over the action taken by the other user, therefore, the part of its utility affected by the other user's action, i.e., its QoS and QoE, cannot be controlled. In the one-shot game, when the players cannot build trust with each other, they selfishly choose the action to maximize their utility. Hence, the NE for video exchange is given by  $(0, 0)$ .  $\square$

The NE in the one-shot game shows that the two selfish users will not exchange any video if there is only a single stage exchange (lasting for  $N_2 \Delta$  slots) of videos. No selfish end user would be willing to incur a cost to send the video when there is no guarantee of receiving any video in return.

## 5.3.2 Finite Horizon Repeated Game

In this section, we consider that the players interact repeatedly under the same conditions, i.e., the same game is played repeatedly. The repeated game can lead to a change in the outcome because the players can observe the previous interactions and decide their present action based on the past actions taken by the other player. We will firstly formulate the game tuple for the repeated game, in coherence with game tuple defined in section 5.3.1. We will consider two cases: (i) finite-horizon repeated game: the players know in advance how many times they will interact (ii) infinite horizon repeated game: the players are unaware of how many times

they will interact. These two cases model the situations where the channel is repeatedly changing from one state to other and the users want to make a video exchange over such a channel repeatedly while adapting to the channel conditions. However, in the first case the users are aware of duration of video exchange whereas in the second case, it is not known apriori how long the video exchange would take place. We will study the subgame-perfect equilibrium of these two cases.

### 5.3.2.1 Repeated Game formulation

A repeated game, in which the game  $\mathcal{G}_\mathcal{O}$  defined in eq. (5.6) is played repeatedly, is defined using the following tuple:

$$\mathcal{G}_\mathcal{R} = \{ \mathcal{P}, \{ \mathcal{S}_k \}_{k \in \mathcal{P}}, \{ v_k \}_{k \in \mathcal{P}}, T \}$$

We now describe the components of this game.

The set  $\mathcal{P}$  refers to the set of players, given by the set of nodes  $\{A, B\}$ .

The set  $\mathcal{S}_k$  is the strategy set of user  $k$ . The strategy set is different from the action set in eq. (5.6) because it describes the actions taken by the player  $k$  based on the history of the play and on the actions at each instant. More precisely, let the actions taken by the players in the  $\tau$ -th stage be  $a^{(\tau)} = (a_1^{(\tau)}, a_2^{(\tau)})$ . Then the history of the game at the end of stage  $t \geq 1$  is given by  $h^{(t+1)} = (a^{(1)} a^{(2)} \dots a^{(t)})$ . The set of all possible histories up to  $t$  is given by  $\mathcal{H}^{(t)} = \{ \mathcal{A}_k \times \mathcal{A}_{-k} \}^t$  where  $\mathcal{A}_k$  is the action set in eq. (5.6) given by  $[0, \lambda \frac{\pi}{2}]$ . The strategy of a player  $k$  is  $s_k = (s_k^{(1)}, s_k^{(2)}, \dots, s_k^{(T)}) \in \mathcal{S}_k$  and maps each possible history to an action, as  $s_k^{(t)} : \mathcal{H}^{(t)} \rightarrow \mathcal{A}_k$  such that  $s_k^{(t)}(h^{(t)}) = a_k^{(t)}$ .

The long term utility for the repeated game is a discounted utility function  $v_k$  such that for a strategy profile  $s = (s_1, s_2)$ , is the weighted sum of the stage utilities obtained by user  $k$  following the strategy  $s$ , such that the user discounts the future payoffs by a discount factor  $\delta$ . It is given by

$$v_k(s) = \frac{(1-\delta)}{(1-\delta^T)} \sum_{t=1}^T \delta^{t-1} u_k(a^{(t)}) \quad (5.8)$$

where  $a^{(t)}$  is the action profile at stage  $t$  induced by strategy  $s$  (i.e.  $s_k^{(t)}(h^{(t)}) = a_k^{(t)} \forall k$ ), discount factor  $\delta \in (0, 1)$  and  $u_k$  is the one-stage payoff function defined in eq. (5.7).

The parameter  $T > 1$ , refers to the number of times the interaction takes place.

Before we discuss the outcome of the repeated game  $\mathcal{G}_\mathcal{R}$ , we briefly describe the concept of sub-game perfect equilibrium for repeated games [101]. In coherence with the NE of one-shot game, the NE of repeated game is a strategy profile from which no player gains by unilaterally deviating. However, there are some strategy profiles, which are not expected to occur due to rational behavior of the players. Hence the NE region can be narrowed down to define a subgame perfect Equilibrium region, as a subset of NE. A subgame is a game from stage  $t$  onwards having a history  $h^{(t)}$ , and is denoted by  $\mathcal{G}_\mathcal{R}(h^{(t)})$ . The final history for this subgame is given by  $h^{(T+1)} = (h^{(t)}, a^{(t)}, \dots, a^{(T)})$ . The strategies and payoffs used for the sub-game are the functions of possible histories that are consistent with  $h^{(t)}$ . Any strategy profile  $s$  of the whole game induces a strategy  $s | h^{(t)}$  on any subgame  $\mathcal{G}_\mathcal{R}(h^{(t)})$  such that for all  $k$ ,  $s_k | h^{(t)}$  denotes the restriction on  $s_k$  to the histories consistent with  $h^{(t)}$ . A subgame perfect equilibrium (SPE) is defined as a strategy profile  $s^* = (s_1, s_2)$  such that for all  $h^{(t)} \in \mathcal{H}^{(t)}$ , the strategy  $s^* | h^{(t)}$  is a NE for any subgame  $\mathcal{G}_\mathcal{R}(h^{(t)})$ .

In the next subsection, we will investigate the SPE for the repeated video exchange game.

### 5.3.2.2 Subgame perfect equilibrium for finite horizon game

In this section, we consider the repeated game  $\mathcal{G}_R$ , where  $T$  is finite. In other words, we study the case when the nodes  $A$  and  $B$  exchange the video (lasting for  $N_2\Delta$  time slots) with each other  $T$  times and both the nodes are aware of the value of  $T$ . The SPE of this game is given by the following proposition.

**Proposition 14.** *The unique sub-game perfect equilibrium of the video exchange game between two nodes, defined by  $\mathcal{G}_{\mathcal{R}} = \{\mathcal{P}, \{\mathcal{S}_k\}_{k \in \mathcal{D}}, \{v_k\}_{k \in \mathcal{D}}, T\}$ , where  $T < \infty$  and is known to both nodes is given by*

$$s_k^{(t),*} = 0, \quad \forall k \in \mathcal{D}, \forall t \in \{1, 2, \dots, T\}$$

*Proof.* This proof follows by the backward induction principle [26]. We consider the last stage of the game. When the game is played at the  $T$ -th stage, the players are aware that they interact for the last time. They have no incentive to transmit the video for the other player while incurring a cost themselves, when there is no guarantee that the other player will transmit the video or not. Hence, their optimal strategy is  $s_k^{(T),*} = 0$ . Now, when the players play the game at time  $T - 1$ , given that in stage  $T$  they will not transmit, there is no incentive to transmit for any history  $h^{(T-1)}$ . Therefore,  $s_k^{(T-1),*} = 0$ . Following the backward induction principle, the players have no incentive to transmit at any stage in the repeated game, therefore, the SPNE is given by  $s_k^{(t),*} = 0 \forall t \in \{1, \dots, T\}$ . The discounted pay-off is then  $v_k(s^*) = 0$ . Note that this result is based on the principle that a rational player will never choose an action that is strictly dominated [26]. At each stage of the game, the strategy  $(0, 0)$  is strictly dominating.  $\square$

In case of finite horizon game, the nodes are aware of the number of times the video exchange will occur and hence, the nodes act selfishly at each stage of the game in order to maximize their payoffs. There is no incentive in building long-term trust in this game for any player, because it is known that the game will be played only  $T$  times.

### 5.3.3 Infinite Horizon Repeated Game

We will now study the SPE for infinite horizon repeated game. The reason why cooperation (i.e., any strategy apart from  $(0, 0)$ ) is not sustainable in finite horizon games is that the players know precisely when the interaction ends. However, in infinite horizon repeated games, the players in the game  $\mathcal{G}_{\mathcal{R}}$  do not know precisely when the game will end, or equivalently,  $T \rightarrow +\infty$ . We will now describe some of the achievable SPE for such games. Note that the overall achievable SPE region for the infinite horizon repeated games is an open problem and not known in general [26]. We will describe some of the possible SPE for such games.

**Proposition 15.** *A sub-game perfect equilibrium of the video exchange game between two nodes in an infinite horizon repeated game described by*

$\mathcal{G}_{\mathcal{R}} = \{\mathcal{P}, \{\mathcal{S}_k\}_{k \in \mathcal{D}}, \{v_k\}_{k \in \mathcal{D}}, +\infty\}$ , *is given by*

$$s_k^{(t),*} = 0, \quad \forall k \in \mathcal{D}, \forall t \in \{1, 2, \dots, \infty\} \quad (5.9)$$

*Proof.* In order to prove the above, we cannot use backward induction as for Proposition 2, because the players are not aware of the last stage of the game. We use the one-step deviation principle to prove this SPE [26]. The one-step deviation principle states that a strategy profile  $s^* = (s_1^*, s_2^*)$  is an SPE if for every user  $k$ , there exists no strategy  $\hat{s}_k \neq s_k^*$  such that at any stage  $\tau$  and history  $h^{(\tau)}$ , the strategy  $\hat{s}_k | h^{(\tau)}$  is a better response than  $s_k^* | h^{(\tau)}$  in a subgame  $\mathcal{G}_{\mathcal{R}}(h^{(\tau)})$ . This principle is based on the fact that if there is a strategy which offers the incentive to a player to deviate at a single stage in the game then the initial strategy is not a SPE. However, if the player has no incentive for any deviation from its current strategy at any stage, such a strategy is an SPE.

To use the one-step deviation principle, firstly, we prove that the stage payoffs in eq. (5.7) are uniformly bounded. Consider the expression in eq. (5.7):  $u_k(\alpha_k, \alpha_{-k}) = w_1 f_k^{QOS}(\alpha_{-k}) + w_2 f_k^{QOE}(\alpha_{-k}) - w_3 f_k^{COST}(\alpha_k)$ . The first term can be easily written as (using (3))

$$\frac{\beta}{2} \left( \frac{1}{R_1} + \frac{1}{R_2} \right) \leq f_k^{QOS}(\alpha_{-k}) \leq 1$$

This is due to the fact that  $\beta \leq \min(R, r_k(n))$  because  $\beta < R_1, R_2$  by assumption. Similarly, the second term can be written as (using (4))

$$0 \leq f_k^{QOE}(\alpha_{-k}) \leq 1$$

This follows the fact that  $0 \leq \max(0, \text{sgn}(x)) \leq 1$ . The third term can be written as (using eq. (5.5))

$$0 < f_k^{COST}(\alpha_k) < \log \left( 1 + \frac{\pi}{2} \right)$$

This is due to the limit on  $\alpha_k \in [0, \tan(\frac{\pi}{2}))$ . Using the above limits, we can write loose bounds of stage payoffs as

$$w_1 \frac{\beta}{2} \left( \frac{1}{R_1} + \frac{1}{R_2} \right) - w_3 \log \left( 1 + \frac{\pi}{2} \right) < u_k(\alpha_k, \alpha_{-k}) \leq w_1 + w_2$$

Since  $w_i \in [0, 1]$ , hence, the stage payoffs are uniformly bounded.

Secondly, we will evaluate if any deviation from the profile in eq. (5.9) offers a profit to one of the players. Lets assume that the player  $k$  deviates from the strategy  $s_k^*$  at stage  $\tau$  and history  $h^{(\tau)}$  with the strategy  $\hat{s}_k^{(\tau)}(h^{(\tau)}) = \alpha_k \in (0, \lambda \frac{\pi}{2}]$  and from then on conforms again to the strategy  $s_k^*$  such that  $\hat{s}_k^{(t)} = s_k^{(t)*}$  for all  $t > \tau$ . This implies that the node  $k$  transmits video with the rate adaptation gradient  $\alpha_k > 0$ , at an intermediate game played at the  $\tau - th$  stage and then conforms to  $\alpha_k = 0$  for the rest of the game. The node  $k$  therefore incurs some extra cost of transmission at the  $\tau$ -th stage given by eq. (5.5) which leads to  $u_k(\alpha_k, 0) < u_k(0, 0)$ . Therefore, the pay-off  $v_k(\hat{s}_k | h^{(\tau)}, s_{-k}^* | h^{(\tau)}) < v_k(s_k^* | h^{(\tau)}, s_{-k}^* | h^{(\tau)})$ . Hence the node  $k$  has no incentive in sending any video packets at any stage of the game, thereby making eq. (5.9) an SPE.  $\square$

An SPE given by eq. (5.9) is independent of the choice of the discount factor  $\delta \in (0, 1)$ . However, there are other possible SPE for the infinite time horizon repeated game  $\mathcal{G}_R$ . We will further show that depending upon the discount factor and other system parameters, an SPE that allows non-trivial video exchange is sustainable in long term. The intuition for existence of other SPE is similar to the infinite time horizon prisoner's dilemma [28]. At any stage of the game, the pay-off in eq. (5.7) has two components: (i) the component controlled by the opposite player's action which increases the utility, but the player  $k$  has no control over it (ii) the component influenced by the player's own action which decreases the utility. The player has no control over the first component, but in long term, it can build trust with the other player to influence the other player's decision and thereby both players can obtain a long term sustainable utility which is higher than utility achieved by the (0, 0). We will now identify such SPE.

Consider the action profile  $(\alpha_1^*, \alpha_2^*)$  such that

$$\begin{aligned} u_1(\alpha_1^*, \alpha_2^*) &> u_1(0, 0) \\ u_2(\alpha_1^*, \alpha_2^*) &> u_2(0, 0) \end{aligned} \tag{5.10}$$

We focus on such an action profile to evaluate the long term sustainability of the action profile whose utility is better than the one-shot NE for both players. In a finite horizon game, the NE action profile is the SPE due to no incentive of building up trust among the players. However, in case of infinite horizon games, an action profile with higher utility may be sustained, even if it has a higher single stage cost for the player,  $\alpha_1^*, \alpha_2^* > 0$ .

The players may be willing to take the risk of paying higher cost due to the trust that the other player will take the action which will lead to a higher payoff.

In the next proposition, we describe such SPE of the game  $\mathcal{G}_{\mathcal{R}}$  which is conditional to the value of the discount factor  $\delta$ .

**Proposition 16.** *In an infinite-horizon repeated game  $\mathcal{G}_{\mathcal{R}} = \{\mathcal{P}, \{\mathcal{S}_k\}_{k \in \mathcal{P}}, \{v_k\}_{k \in \mathcal{P}}, +\infty\}$ , and with an agreement profile  $(\alpha_1^*, \alpha_2^*)$  satisfying eq. (5.10), if the discount factor is bounded by*

$$1 > \delta > \delta_{asym}^{min}(\alpha_1^*, \alpha_2^*) \quad (5.11)$$

where  $\delta_{asym}^{min}(\alpha_1^*, \alpha_2^*)$  is given by

$$\delta_{asym}^{min}(\alpha_1^*, \alpha_2^*) = \max_{k \in \mathcal{P}} \frac{w_3[f_k^{COST}(\alpha_k^*) - f_k^{COST}(0)]}{w_1[f_k^{QOS}(0) - f_k^{QOS}(\alpha_{-k}^*)] + w_2[f_k^{QOE}(0) - f_k^{QOE}(\alpha_{-k}^*)]} \quad (5.12)$$

then the following strategy is an SPE: “A node  $k$  transmits with gradient  $\alpha_k^*$  at the first stage and continues to adapt the rate with this gradient as long as the other player adapts its rate at least by  $\alpha_{-k}^*$ . If any player has ever defected, then the players transmit at  $\alpha = 0$  for the rest of the interaction.”

*Proof.* See Appendix. □

*Remark 17.* It should be noted that the  $0 < \delta_{asym}^{min}(\alpha_1^*, \alpha_2^*) < 1$  for all values of  $(\alpha_1^*, \alpha_2^*)$  satisfying eq. (5.10). This can be shown as follows:  $\delta_{asym}^{min}(\alpha_1^*, \alpha_2^*)$  can be written as (as shown in Appendix)

$$\max_{k \in \mathcal{P}} \left\{ \frac{[u_k(0, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)]}{[u_k(0, \alpha_{-k}^*) - u_k(0, 0)]} \right\}$$

By condition eq. (5.10),  $u_k(\alpha_k^*, \alpha_{-k}^*) > u_k(0, 0)$ . Hence,  $-u_k(\alpha_k^*, \alpha_{-k}^*) < -u_k(0, 0)$ . Adding  $u_k(0, \alpha_{-k}^*)$  on both sides we get,  $u_k(0, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*) < u_k(0, \alpha_{-k}^*) - u_k(0, 0)$ . Therefore,  $\frac{[u_k(0, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)]}{[u_k(0, \alpha_{-k}^*) - u_k(0, 0)]} < 1$ . It can also be seen trivially that the numerator and denominator are both positive, because  $u_k(0, \alpha_{-k}^*) > u_k(\alpha_k^*, \alpha_{-k}^*)$  and  $u_k(0, \alpha_{-k}^*) > u_k(0, 0)$ . Therefore,  $0 < \frac{[u_k(0, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)]}{[u_k(0, \alpha_{-k}^*) - u_k(0, 0)]}$ . Consequently,  $0 < \delta_{asym}^{min}(\alpha_1^*, \alpha_2^*) < 1$ . Hence, eq. (5.11) does not imply any additional condition on system parameters. If an agreement point satisfies eq. (5.10), it is a sufficient condition to say that there is an admissible discount factor range within  $(\delta_{asym}^{min}(\alpha_1^*, \alpha_2^*), 1)$  for which the agreement point is sustainable. Intuitively, it can be noted that if any agreement point provides a higher utility than the one shot NE, such agreement point can be sustained. The probability of its sustenance is depending on the probability for the game to stop which is identified by  $\delta$  having a lower bound of  $\delta_{asym}^{min}(\alpha_1^*, \alpha_2^*)$ . This probability is always less than 1, by no additional condition, thereby conforming that the agreement point is always sustained with certain probability.

From the above proposition 4, we have identified the discount factors range which can lead to long term sustainable SPE other than  $(0, 0)$ . The discount factor can be seen as the perception of the players of the probability for the game not to end at every stage of the game. If the players perceive the probability of the game not to stop above certain limits, there can be a non-trivial SPE. In other words, if the probability of the game not to stop is large enough, the players would rather prefer to develop trust and obtain better payoffs than minimize their one-shot costs. We also note here, that only those asymmetric points which satisfy eq.(5.10) are sustainable thereby ensuring that the players agree upon the points which offer them incentive over one-shot NE. It is also worth noting, that all points which are achievable definitely satisfy eq.(5.10), and further the value of discount factor range in proposition 4 determines the achievability. Therefore, the discount factor range in proposition 4 does not imply any supplementary condition on the system parameters,

and for every achievable point through condition eq.(5.10), there is some discount factor range for which this point can be achieved.

## 5.4 Selection of sustainable agreement profiles

We have shown above that when the video exchange is done repeatedly, the players can transmit the video at an agreement profile defined in eq.(5.10). These agreement profiles are chosen based on the condition, that if both the players are able to obtain better payoffs with  $(\alpha_1^*, \alpha_2^*)$  than the payoff obtained at the NE  $(0, 0)$ , they may agree to such a profile  $(\alpha_1^*, \alpha_2^*)$ . A natural question that arises is that out of all these agreement profiles, which profile will be selected by the players. In general, this is an open problem and there is no unified theory to obtain the answer [26]. In this section, we present a qualitative solution to this problem: we use the system parameters specific to the video exchange scenario to assess the players' decision in order to select an agreement profile. Therefore, we illustrate which of the achievable agreement profiles are more likely to occur over a period of time. We will consider two factors influencing the choice of agreement profiles: the tolerance of the players and the pareto optimality of the agreement points.

### 5.4.1 Tolerance Index

It has been shown in section 5.3.3 that depending on the discount factor, the region of the sustainable agreement profiles is determined. However, these profiles are not always symmetric. In other words, some of these agreement profiles provide a higher payoff to the first user, some profiles provide a higher payoff to the second user and some profiles provide equal payoffs to both users. However, it is assumed here that every player  $k$  agrees to the action profile  $(\alpha_k^*, \alpha_{-k}^*)$  without considering the payoff obtained by the other player,  $u_{-k}(\alpha_k^*, \alpha_{-k}^*)$ , as long as its own payoff  $u_k(\alpha_k^*, \alpha_{-k}^*)$  is higher than the payoff it obtains at the NE. This means that the players are selfish but not malicious. However, in realistic scenarios, if the two selfish players have similar negotiation stand point, they will not agree to a profile  $(\alpha_k^*, \alpha_{-k}^*)$ , such that one player gets a higher payoff than the other. This is due to the following reasoning: no selfish player  $k$  will agree to an agreement action profile when it is aware that the other player could offer it a better payoff by reducing its own payoff  $u_{-k}$  to be equal to or closer to  $u_k$ . Therefore, it is imperative that the agreement point should be chosen on the basis of the demands of the players. There is also an implicit advantage in such a behavior for the service provider who wants a fair treatment for all the users but also does not want to provide more rate than necessary "for user to be happy". In order to model such demands, we propose the concept of tolerance index as follows:

**Corollary 18.** *If the players in the infinite horizon repeated game are aware of the utility obtained by the other players at the agreement point, an action profile  $(\alpha_1^*, \alpha_2^*)$  can be an agreement point for the strategy in Proposition 6, if it satisfies, in addition to eq.(5.10), the following condition*

$$|u_k^* - u_{-k}^*| < \xi \min(u_k^*, u_{-k}^*) \quad (5.13)$$

where  $u_k^* = u_k(\alpha_k^*, \alpha_{-k}^*)$ ,  $u_{-k}^* = u_{-k}(\alpha_k^*, \alpha_{-k}^*)$  and  $\xi > 0$  is the tolerance index .

The agreement points, in eq.(5.10), do not imply a condition that the utilities of both players need to be equal. In other words, a player agrees to any action profile which provides them a higher payoff than the point  $(0,0)$ , no matter what the other player obtains with such an agreement. The players have infinite tolerance to whatever payoff the other player obtains, as long as they obtain a higher payoff than  $(0,0)$ . However, the introduction of tolerance index  $\xi$  adds to this condition as follows: In addition to eq.(5.10), the condition eq.(5.13) implies that the player obtaining a lower payoff will agree to any profile  $(\alpha_1^*, \alpha_2^*)$  only if it gets a payoff higher than NE (implied from eq.(5.10)) and the other player obtains a payoff which is not more

than  $\xi$  times higher than its own payoff. Intuitively, the tolerance index  $\xi$  models how much the players can tolerate the difference in the payoff of its own and the other user at an agreement point. In other words, when the players exchange the video, a player can tolerate, or agree to, a poorer quality video than the other player, only to certain point while knowing the other player is obtaining a higher quality. If the quality of video of a player is worse than allowed by the tolerance index condition, the player would not agree to such an agreement point, even though it offers a better quality than a much worse video quality obtained at the NE. This shows that the nodes will rather exchange video at symmetrical (0, 0) than an unfair asymmetrical point, subjected to their tolerance level.

## 5.4.2 Pareto Optimality

It has been shown that depending on the discount factor and the tolerance index, different agreement profiles in the sustainable agreement region can be achieved by the players. However, these agreement profiles may provide different payoffs to both the players, that means, some agreement profiles may give equal payoffs to both players, while some agreement profiles do not, depending on player tolerance. In order to determine which profiles the players are likely to agree upon, we use the concept of pareto optimality [26]. Pareto optimal profiles are those agreement profiles from which no user can improve its own payoff without making the other user's payoff worse. Any change from the pareto optimal profiles leads to worsening of payoffs of at least one of the players. It is intuitive that the players will tend to agree upon the profiles which are pareto optimal. This can be proved as follows: assume that the players choose an agreement profile which is not pareto optimal. Since this profile is not pareto optimal, this means by definition, either or both players can improve their payoffs without worsening the payoff of the other player. Hence, the players are likely to choose such a point. On the other hand, if the players agree upon a pareto optimal point, no player can improve its own payoff without worsening the payoff of the other user. Therefore, the players will tend to agree upon the agreement profiles which are pareto optimal such that both users can obtain the maximum possible payoff at a given discount factor and tolerance index, and there is no scope for unilateral improvement in payoff. Intuitively, the pareto optimal points are the points on the outer boundary of the whole achievable region.

Therefore, using the joint effect of tolerance index and pareto efficiency, the most likely agreement profiles to be chosen by the players can be determined. We will show in the next section, that such profiles are the pareto optimal points located at the boundary of the region of achievable action profiles region by both users based on their tolerance index.

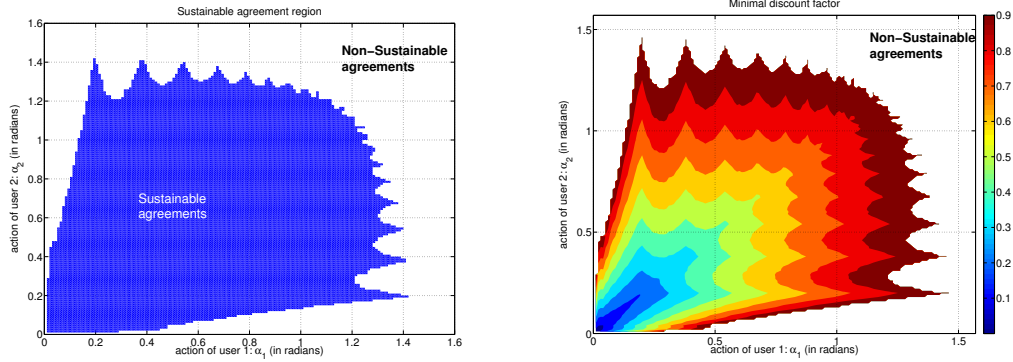
## 5.5 Numerical Results

We will now present the simulation results to identify the region of all the agreement points  $(\alpha_1^*, \alpha_2^*)$  which are sustainable in the long term under sufficient conditions on the discount factor (which ensures that the infinite horizon repeated game lasts long enough for these points to be achieved). The channel parameters we have considered in these simulations are as follows: the channel has a capacity of  $R_{c1} = 350$  kbps for  $N_1 = 1000$  time slots and  $R_{c2} = 200$  kbps for  $N_2 = 3000$  time slots. The parameters of the rate adaptation are :  $\beta = 100$  kbps,  $\Delta = 1$ . The weights of the utility function are fixed at  $w_1 = 0.45$ ,  $w_2 = 0.45$  and  $w_3 = 0.1$ .

### 5.5.1 Discount factor variation

In figure (5.3a), we plot the region of all the sustainable agreement points for infinite tolerance  $\xi$ . The point (0,0) is both NE and SPE. All the other achievable points are in blue region, and are achievable on the basis of the value of discount factor  $\delta$ . Note that the achievability is shown on the basis of eq.(5.10) : if  $u_k(\alpha_1^*, \alpha_2^*) > u_k(0,0)$ , it is achievable depending on the value of the discount factor. Also note that the points on the axes, namely,  $(\alpha, 0)$  and  $(0, \alpha)$  for  $\alpha > 0$  are not achievable for any value of discount factor. This is





(a) Sustainable agreement region for the infinite horizon repeated game with infinite tolerance index and weights as  $w_1 = w_2 = 0.45$  and  $w_3 = 0.1$ . Video transmission with gradients in non-tation to varying channel conditions is achievable. sustainable agreements region is not possible due to worse payoff than NE.

Figure 5.3: Sustainable agreement region and corresponding discount factor region.

because no user has an incentive to sustain such points because  $u_k(\alpha, 0) < u_k(0, 0)$  and  $u_{-k}(0, \alpha) < u_{-k}(0, 0)$  hence a unilateral deviation is always preferred. Also note, the curve has spikes periodically due to the non uniform variation of utility function which can be explained as follows: the utility function  $u_k(\alpha_k, \alpha_{-k})$  varies uniformly with  $\alpha_k$  because  $\frac{du_k(\alpha_k, \alpha_{-k})}{d\alpha_k}$  is monotonously decreasing for increasing  $\alpha_k$ . This implies that increasing the user's own gradient simply incurs more cost to the user and hence decreases its own payoff. However, the utility function does not vary monotonously with  $\alpha_{-k}$ . This is because, the utility of a user  $u_k$  is affected by the action of the other user  $\alpha_{-k}$  in two ways. Firstly, as  $\alpha_{-k}$  increases, the flow continuity decreases and thereby quality of experience decreases, due to more instances of rate exceeding the channel capacity. Secondly, as  $\alpha_{-k}$  increases, the quality of service increases for smaller values of  $\alpha_{-k}$ , when there are not many instances of rate exceeding the channel capacity. However, when such instances increase at high values of  $\alpha_{-k}$ , the quality of service decreases with increasing  $\alpha_{-k}$ . Therefore, due to the culmination of these two effects, the agreement point region does not show monotonous behavior. In fact, the derivative of utility function with action profile of other user is a complex trigonometric function which is periodic in nature leading to the shape of the curve being similar to periodic function.

We will now illustrate the range of discount factor for sustainable agreement points. In figure 5.3b, we show the discount factor variation at different achievable points. The points in achievable region shown in figure 5.3a, are achievable subjected to the discount factor value shown in figure 5.3b. The white region is the same non-sustainable region shown in figure 5.3a. The variation of minimal discount factor is shown by various colors. The higher is the minimal discount factor value, the higher the probability of the game not to stop. It can be seen that the higher is the probability of the game to stop, the closer the agreement points are to the one-shot NE of  $(0, 0)$ . When there is a low probability of the game to stop, the players would prefer to build trust and transmit at an agreement point further away from the NE, to obtain a higher payoff. It can be seen in figure 5.3b, that as we go radially away from origin  $(0, 0)$ , the minimal discount factor increases, i.e. the probability of game to end decreases. Therefore, if there is a lower probability of the game to stop, the players adapt to the channel capacity faster, or with a higher gradient in order to achieve a high utility and consequently high quality videos.

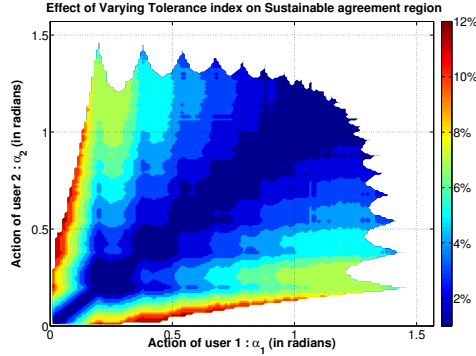


Figure 5.4: Variation of sustainable agreement region for varying tolerance index. As the tolerance index increases, the non-symmetric points become achievable. With the least tolerance, the players agree to the same adaptation rates where both of them get same quality video.

## 5.5.2 Tolerance Analysis and Pareto optimality

In the above figures, we assumed the tolerance index  $\xi$  to be infinite thereby implying that the players are tolerant to the higher payoffs obtained by the other player at the agreement points. Furthermore, in figure 5.4 we show the variation of the sustainable agreement points region with varying  $\xi$ . It can be seen that when the tolerance index is as low as 1%, the sustainable action profiles are symmetric. However, as the tolerance index increases, the asymmetric action profiles become sustainable and the sustainable agreement region increases. This shows that when the nodes are less tolerant, they do not agree to transmit the video as long as they do not achieve the video quality at least as good as the other player and definitely better than NE. However, as the nodes become more tolerant, they agree to transmit at the action profiles which gives them a video quality not as good as the other player's video, thereby making the asymmetric action profiles sustainable. This behavior might be also preferred by the service provider of the network as it provides more efficient use of resources throughout the network while ensuring the satisfaction of the users.

We investigate the efficiency of any agreement point chosen by the users for a given tolerance index. In figure 5.5, we plot the payoffs achieved at the sustainable agreement points. Clearly all the payoffs are higher than the payoff at the one shot NE. For a tolerance index of 0%, all sustainable agreement payoffs are symmetric, whereas, as the tolerance index increases, the asymmetric payoffs are sustainable. This illustrates that at higher tolerance, the players can agree upon different action profiles, leading to different video QoS and QoE, and at lower tolerance, the users agree upon only the symmetric action profiles, leading to symmetric video QoS and QoE (which is as good as the other player). It can also be seen that when  $\xi = 0\%$ , there is only one sustainable payoff that is pareto optimal which is shown as red-squared point. As the tolerance index increases, there are more sustainable payoffs that are pareto optimal. In other words, at lower tolerance, there is a unique payoff, for which the quality of video of the users is such that they cannot improve their quality further, without worsening the quality of the other user. However, if the players are more tolerant, there are many agreement profiles, which are pareto efficient and at which no user can improve its video quality further without worsening the other user's quality. The pareto efficient payoffs provide the optimal video quality to both the users subjected to their tolerance index.

Another interesting observation from figure 5.5 is that when the players have a higher tolerance index, the asymmetric payoffs which are pareto optimal are on both sides of the first bisector. If the players have different leverages over each other, the player that has higher leverage (or is more powerful) can influence the payoff in its favor, which is identified by the points either above or below the first bisector. If the user 1 has a higher leverage/power, for a tolerance index of  $\xi > 0\%$ , it would prefer to have a higher utility, and an

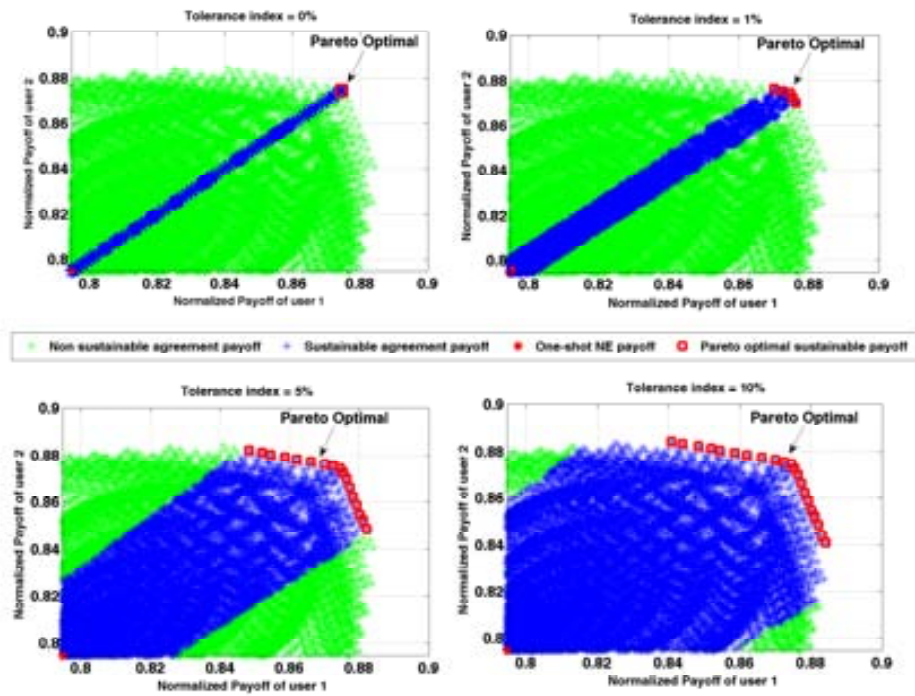


Figure 5.5: Comparison of the achievable payoffs with varying tolerance index. As the tolerance index increases, more payoffs are sustainable. The number of Pareto optimal payoffs reduce as the tolerance index decreases.

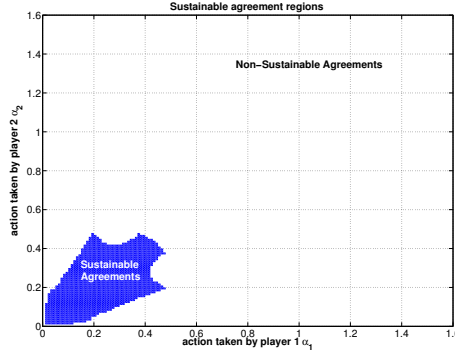


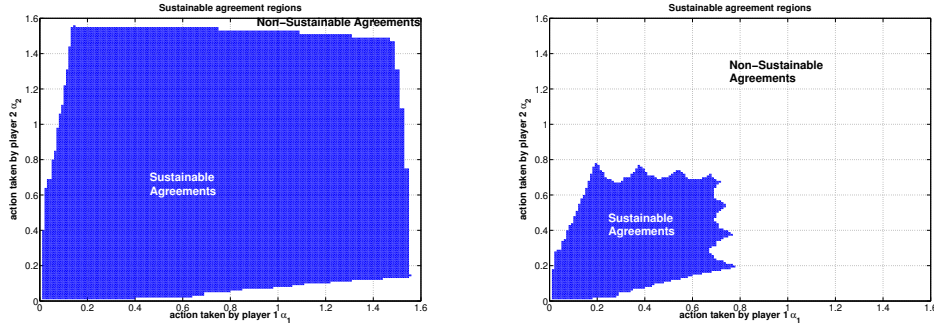
Figure 5.6: The region of all agreement points  $(\alpha_1^*, \alpha_2^*)$  with weights assigned as  $w_1 = w_2 = 0.4$  and  $w_3 = 0.2$ . The sustainable agreement region is reduced with a higher weightage to the cost. Due to the high cost for faster adaptation to transmit the video, the users agree to slow adaptation points.

agreement point below the first bisector is chosen, whereas, if user 2 has the higher leverage, an agreement point having payoffs above the first bisector is chosen.

### 5.5.3 Influence of weights of QoS, QoE and Cost

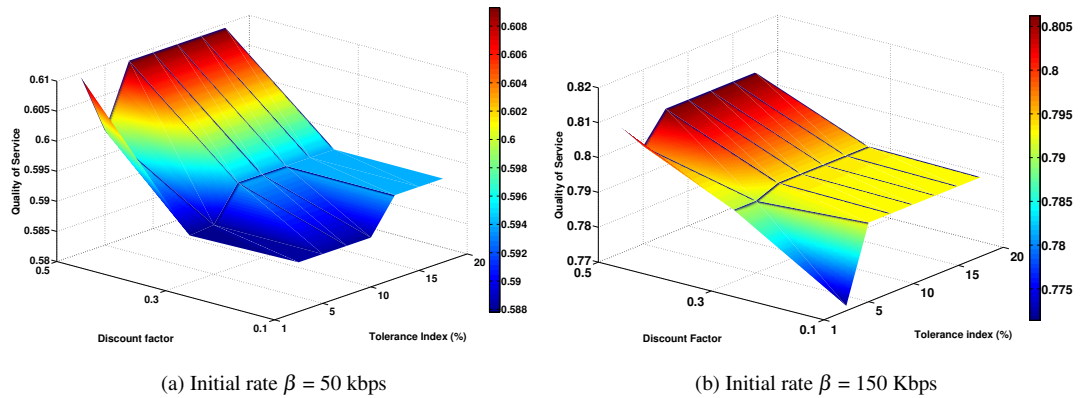
Now, we illustrate the dependence of the region of the sustainable agreement points on the different system parameters. We will now show that depending on the different weights  $w_i$  assigned to different factors affecting the utility in eq.(5.7), namely, QoS, QoE and Cost, the number of agreement points tremendously varies. In figure 5.6, the weight assigned to cost is increased to  $w_3 = 0.2$  as compared to figure 5.3a, where  $w_3 = 0.1$ . The other weights are evenly distributed as  $w_1 = w_2 = 0.4$ . The region of sustainable agreements has reduced in figure 5.6 as compared to figure 5.3a. This is because as the impact of cost increases, the utility  $u_k$  decreases as a function of  $\alpha_k$  with a higher gradient. This reduces the number of points satisfying eq.(5.10) and there are fewer agreement points giving utility higher than  $u_k(0, 0)$ . In general, the sustainable agreement region reduces with increasing weightage to the cost. This also shows that the players will adapt to channel capacity much slowly at the agreement point, when the cost is high.

Furthermore, in figure 5.7a, we fix the weightage of cost  $w_3 = 0.1$ , and consider a higher weightage of QoS over QoE. Therefore, we fix  $w_1 = 0.6$  and  $w_2 = 0.3$ . The region of sustainable agreements increases as compared to the one in figure 5.3a. In figure 5.7b, we assign a higher weightage to QoE over QoS such that  $w_1 = 0.3$  and  $w_2 = 0.6$ . We observe that the region for sustainable agreement points has reduced. This behavior can be explained as follows: as the weightage of QoS increases, the utility  $u_k$  increases with increasing  $\alpha_{-k}$  rapidly, due to higher gradient  $w_1$ . Hence, for any given point  $(\alpha_1, \alpha_2)$ , there is a higher probability for the point to satisfy eq.(5.10). On the other hand, for QoE, a higher  $\alpha_{-k}$  leads to a lower  $f^{QoE}$  as there are more instances when rate exceeds the capacity (figure 2). Therefore, if a higher weightage is given to QoE, there are more points having lower utilities than  $u_k(0, 0)$  thereby violating condition eq.(5.10). When a higher weightage is given to QoS, the players can agree to a faster rate adaptation which in turn provides a bigger sustainable agreements region in figure 5.7a. When a higher weightage is given to QoE, the faster rate adaptation points are not preferred by the players because, it leads to a higher number of instances when the rate exceeds channel capacity (figure 2), which in turn reduces QoE. Hence, the slower rate adaptation points are sustainable in this case. Therefore, we have an increased sustainability region in figure 5.7a as compared to figure 5.3a, and a reduced sustainability region in figure 5.7b as compared to figure 5.3a.



(a) Sustainable agreement points for weights  $w_1 = 0.6$ ,  $w_2 = 0.3$  and  $w_3 = 0.1$ . The maximum weightage is given to the QoS. (b) Sustainable agreement points for weights  $w_1 = 0.3$ ,  $w_2 = 0.6$  and  $w_3 = 0.1$ . The maximum weightage is given to the QoE.

Figure 5.7: Comparison of the region of sustainable agreement points for different weights to QoS and QoE. The weight to the cost is fixed. The higher the weightage to QoS, the more is the achievable region and faster rate adaptation are sustainable.



(a) Initial rate  $\beta = 50$  kbps (b) Initial rate  $\beta = 150$  Kbps

Figure 5.8: Variation of QoS of the video at the pareto optimal outcomes of the game for different initial rates. Higher the discount factor, higher is the QoS.

### 5.5.4 QoS and QoE analysis

We will now identify how our proposed game theoretical model can be used for the design of video streaming system based on user experience. In particular, we will identify a trade-off existent between achieving a desired QoE for the user and the speed of adaptation of transmission rate to the channel conditions.

In figure 5.8, the QoS (as defined in (3)) provided to the users for different initial rates  $\beta$ , at the pareto optimal outcomes of the game, is shown. Particularly, we plot the minimum QoS that is obtained among all the pareto optimal outcomes of the game, for different discount factors and tolerance index values. It can be seen from figure 5.8 that as the discount factor increases, the QoS increases. This is because, as the discount factor increases, the players are likely to agree to transmit at a higher gradient as there is lower probability of game to stop and higher trust among the players. This leads to a higher QoS due to faster adaptation to the channel conditions. Furthermore, as the tolerance index increases, the QoS also increases for a higher tolerance allows more non-symmetric gradients to be achieved thereby giving more flexibility to choose from higher gradients at pareto optimal outcomes. It can also be seen here that for a higher initial rate in figure 5.8b, higher magnitude of QoS is achieved as compared to lower initial rate in 5.8a. This is because for any adaptation gradient  $\alpha$ , a higher percentage of the network capacity is utilized when the initial rate  $\beta$  is higher.

Furthermore, in figure 5.9, we show the variation of minimum QoE (as defined in (4)) obtained at the end user, among all the pareto optimal outcomes of the game, for different discount factors and tolerance index values. It can be seen that the QoE decreases for higher values of discount factor as opposed to QoS variation. This is because at higher values of discount factor, higher rate adaptation gradients are achievable. Consequently, there is a higher probability of the rate to exceed the channel capacity, thereby leading to packet losses and delay in the network, which eventually affects the QoE. Another interesting point is that a higher initial rate gives lower QoE. The reason for this behavior is that a higher initial rate is closer to the channel capacity, which leads to more instances of the rate exceeding the channel capacity as opposed to lower initial rate. Furthermore, a higher discount factor leads to higher gradients achieved by the players at the pareto optimal points, thereby leading to more instances of rate exceeding the channel capacity. These instances are more at higher initial rate as compared to lower initial rate, hence the decay in QoE with increasing discount factor is more at  $\beta = 150$  than at  $\beta = 50$ .

It can be concluded from the above analysis that a trade off exists between speed of adaptation to the channel conditions and QoE. The faster adaptation to the channel conditions is desired for a higher QoS, but it leads to lower QoE. Furthermore, it has also been found that the choice of the initial rate should be relatively lower than the available channel capacity for an improved performance.

## 5.6 Concluding remarks

### 5.6.1 Contributions over state-of-the-art

In this chapter, we have considered a QoE-driven adaptive video exchange between two terrestrial nodes via a relay having symmetric channel conditions. We have considered an adaptive video transmission such that the rate adaptation provides not only a high quality of service but also maximizes the flow continuity of the perceived video, thereby minimizing freezing of the video during playback. A significant contribution made via this work is that the potential of the repeated game theoretical tools in a framework to optimize QoE based rate adaptation is concretely proved.

To the best of our knowledge, this is the first research to bring together the repeated game framework proposed in [101] with the QoE-driven rate adaptation. There are number of analysis using similar frameworks at the lower layers, however, we depart from existing work by providing a realistic and effective analysis of end-user perception driven game. In addition, our analysis models and characterizes a novel and practical metric of tolerance index to enhance the accuracy of the expected transmission outcome.

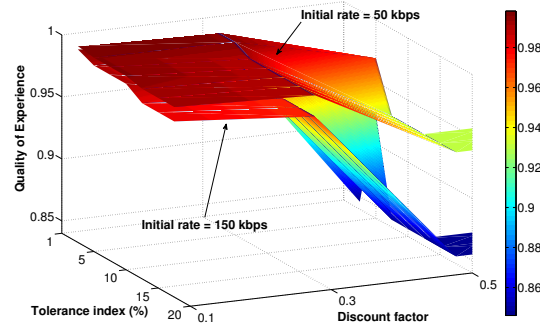


Figure 5.9: Variation of Quality of Experience at different initial rates. Higher the discount factor and higher the initial rate, lower is the QoE.

The game theoretical analysis of adaptive video protocols, like TCP, has been considered in the literature under different scenarios [102, 103].

However, there are some important limitations in the game theoretic analysis of the video transmission in existent works. *First*, in order to assess the dynamics of practical deployment of such schemes, it is required to examine the repeated interactions between the nodes, when independent rate adaptation mechanism is employed at each node, because such interactions tend to last a long period of time. The existent works do not account for such repeated interactions in QoE driven adaptation scenarios. *Second*, the existent works do not address the game theoretical assessment in which QoE of the video observed at the end user is also considered in the payoff function in QoE-driven rate adaptation scenarios. *Third*, the existent work does not account for the video QoS and QoE which are expected to be obtained at the end user in light of qualitative factors namely the rationality and tolerance of the users, and preference of users to the different aspects of the video quality.

In this chapter, we tackle these limitations in the following ways. *First*, we depart from the TCP study and we consider a general framework of QoE-driven rate adaptation to optimize the video quality at the end user. We further analyze not only the one-shot interaction of the users, but a repeated interaction of the users, thereby providing a more realistic video exchange with varying channel conditions. *Second*, we devise a joint utility function to model the preferences of the users which leads to optimization of not only the cost incurred to the user to transmit the video but also the quality of service and quality of experience of the received video. *Third*, we characterize the sustainable video QoS and QoE region which is achievable in repeated games through the analysis of sub-game perfect equilibrium and moreover we also identify which of these video QoS and QoE are likely to be obtained by the rational players via pareto efficiency analysis. In addition, we also model the heterogeneous users which have different tolerance level to the benefits of the other user and illustrate such affects on resulting video application quality.

## 5.6.2 Publications

The research work presented in this chapter has been presented in the following publication:

### *Journal*

- **S. Gupta**, E.V. Belmega, M. A. Vázquez-Castro, "A game-theoretical analysis of QoE-driven adaptive video transmission". *Submitted to Springer Journal on Multimedia Systems*, Nov. 2013.

### *Conference*

- **S. Gupta**, E.V. Belmega, M. A. Vázquez-Castro, "Game-theoretical Analysis of the tradeoff between QoE and QoS over satellite channels". *Submitted to ASMS/SPSC 2014*.

### 5.6.3 Future lines of work

In this work, we have considered the analysis on the basis of a particular QoE-driven rate adaptation scheme. There is a wide palette of practical rate adaptation protocols for media transmission such as TCP and its variants [93, 104] which are widely used nowadays for video transmission over HTTP [92]. By modifications in our rate adaptation model, these protocols can also be analyzed under the same game-theoretical framework and this can be explored as a future work. Another consideration for the future work is the asymmetric channel conditions as it brings us closer to practical scenario analysis. In addition, an open line of research work is the possibility of negotiation between the nodes and buffer period before adopting a change in strategy.

This contribution provides insights into a practical modeling of the user perception and predicting the attainable performance closely to realistic scenario. The focus on the higher layers emphasizes the networking concepts which lead to design of systems to meet the end-user demands. Extension of this work further in the directions pointed out is promising.



## Chapter 6

# Overall Conclusions and Future work

In this thesis, we had set out with a goal to contribute towards novel and advanced networking techniques for wireless relay channels and it is safe to say that we have successfully achieved this goal. We have made an extensive and in-depth study of new and promising techniques which will form potential candidates techniques towards the design of wireless networking in the future. Our methodology and contributions serve to develop the advanced wireless networking further, study and assess their benefits and improve their potential.

To make effective and concrete contributions, we have selected challenging problems in various aspects of advanced wireless networking techniques and presented neat and complete solutions to these problems. Our vision of the thesis has been to develop a work which not only gives hollistic view of networking techniques at different layers of the TCP/IP protocol stack but also models and characterizes them in light of more realistic factors like coherence with state of the art, random selfish behaviors, geographical factors etc. This vision has helped us to present a very deep and thorough analysis of different trade-offs and competitions which spring up in implementation of wireless networking and are otherwise do not seem crucial during the design phase.

Moreover, this thesis is backed by a good set of publications with 2 journal publications (1 submitted), 4 conference proceedings (1 submitted) and 1 book chapter besides contributions towards 3 different collaborative projects.

To conclude this thesis we will now summarize the main aspects of the contributions and future work of this thesis.

### 6.1 Conclusions

Lets summarize our contributions and consequent conclusions organized using the same classification that we have used in this thesis, i.e., based on the layer of protocol stack involved:

- **Physical Layer:** The first contribution we have made and presented in Chapter 3, is focussed upon physical layer. The novel tool of physical layer network coding applied using compute and forward is used with an aim to maximize the overall throughput. However, since this scheme is limited by non-integral channel coefficients, we introduce a novel integer forcing precoder to counter this limitation. Our results show that the IFP can reasonably improve the performance of CF at the rate of some extra complexity and power requirements. However, the overall benefits overshadow these additional inputs. This study reaffirms that physical layer based performance improvements are although most complex but are highly effective in overall sense. The proposed scheme outperforms the existent postcoding based implementation of CF. The proposed precoding techniques to maximize the CF performance can serve as a basis for realistic implementation of CF.

- **Link and Network layer:** After exploring the physical layer, we shifted our focus to MAC and network layer in Chapter 4. It was primarily motivated by the natural benefits of the tool of network coding explored at physical layer. We again formed our basis by the application of network coding to two way relay channel model based topologies (bent-pipe and X) with an aim to maximize the throughput. Moreover, we considered this application in coherence with a QoE driven cross layer optimization framework applied at transport and application layer. We proved by this study that there exists a crucial trade-off between throughput and QoE which in turn need to be considered for application of cross layer rate adaptation techniques, in general, along with network coding. The trade-off in fact models a more general competition between the network service provider's benefits and user demands. Another issue which is found to be very crucial in this implementation is the location of the end users. We designed a novel analytical tool called model maps, which can highly optimize the overall performance taking into account the geographical distribution of the end users. Further, generalizing the use of extra information like geography and designing intelligent tools and algorithms, we performed an extensive study towards the cognitive techniques in wireless networks. The wireless network we considered focussed on Dual Satellite system. This taxonomic analysis is one of the first surveys in the field and we are able to conclude that cognition or intelligent brain-empowered communication would be irrefutably required to manage the growing wireless network demands in future. This work provides a novel design scheme to implement network coding and cross-layer optimization, using novel basis of optimization, namely, implicit trade-offs and geographical distribution.
- **Transport and Application Layer:** A number of aspects which we studied in Chapter 4 inspired to extend the study at higher layers focussing on end user behavior and corresponding networking techniques. To this end, in chapter 5, we focussed upon a two way exchange of information between two selfish nodes where the QoE-driven cross layer optimization techniques are applied. Here we aimed to maximize the throughput however, in a realistic setting, the maximization was modeled to be most efficient with minimum cost and maximum throughput and QoE. Therefore, a joint utility function was formulated. The natural tool to model and assess such a setting is game theoretical tool. Therefore, in order to model a transmission between the nodes which are selfish, autonomous entities, we have formulated a non-cooperative two player game between the two nodes exchanging the video streams. We have proved that if the interaction between the nodes occurs for a finite period of time and the nodes are aware of this time period, no selfish user will exchange the video at a rate higher than a fixed trivial rate. However, if the period of interaction is uncertain and not known to the nodes, exchanging the videos using QoE driven rate adaptation is beneficial for both users. It is also concluded that as the probability of the interaction to continue increases, the nodes adapt to the channel conditions faster. We have analyzed the variation in the behavior of the speed of adaptation of both users when the users' preference varies in terms of expected quality of service, quality of experience and cost. Moreover, an important conclusion drawn from our work is that using the qualitative factors like the rationality of the users and the tolerance of the users towards the benefits of the other users in a video exchange scenario, it would be possible to identify the video QoE and QoS which are more likely to be sustained in long term out of all possible sustainable video QoE and QoS. The quantization of user tolerance also provides benefits from the perspective of the service provider which prefers to satisfy the users' demands in best possible way while ensuring the most efficient utilization of the system resources. Overall, in this work, we have proved that there is an immense potential in the use of repeated game theoretical tools to communication framework and it should be explored further to optimize the network services.

## 6.2 Future Work

As a final note, it is worth mentioning the key future works directly related to the work presented in this document:

- **Chapter 3 (Physical Layer):** From chapter 3 we consider:
  - Extension of implementation of IFP with the suitable constellations for CF implementation.
  - Development of more efficient IFP besides based on partial Zero-forcing Precoding technique.
- **Chapter 4 (Link and Network Layer):** From Chapter 4, we have following lines of work:
  - The trade-off study of Model NC+CL applied to asymmetric two way relay channels.
  - The implementation of location aware Model NC+CL for general wireless networks and complex model maps derivation.
  - The study and developments of practical methods to allow cognition in satellite systems.
- **Chapter 5 (Transport and Application Layer):**
  - The game theoretical analysis of QoE driven rate adaptive video exchange with variable gradient during a cycle.
  - The allowance of network coding at the relay node in the current study.

Overall, our study provides insights towards development of novel methods over current networking techniques. Our results show what possibilities arise by considering various micro issues towards building up robust and efficient next generation wireless networks.

## Appendix A

# Proof of Proposition 16

We will once again use the one-step deviation principle to prove the SPE. Consider the players following the agreement profile  $(\alpha_1^*, \alpha_2^*)$  satisfying (5.10). Let us consider two cases: (i) Case I: there has been no deviation from  $(\alpha_1^*, \alpha_2^*)$  (ii) Case II: there has been a deviation from  $(\alpha_1^*, \alpha_2^*)$  and now both players are using the threat point  $(0, 0)$ .

**Case I:** Let no player deviate from the agreement till the stage  $\tau$ . Let, at stage  $\tau + 1$ , the player  $k$  deviate from the strategy and transmits at  $\tilde{s}_k^{(\tau+1)} = \bar{\alpha}_k \neq \alpha_k^*$ . There are two sub-cases: (1) First we consider the case when  $\bar{\alpha}_k < \alpha_k^*$ . In this case, from stage  $\tau + 2$  onwards, the node A conforms to the initial strategy again. Since there has been a deviation, conforming to the strategy implies  $\tilde{s}_k^{(t)} = 0$  for all  $t \geq \tau + 2$ . We calculate the discounted payoffs in two cases: firstly, in case there is no deviation, the discounted payoff is given by

$$v_k(s^*) = \frac{(1-\delta)}{(1-\delta^T)} \sum_{t=1}^T \delta^{t-1} u_k(\alpha_k^*, \alpha_{-k}^*) = u_k(\alpha_k^*, \alpha_{-k}^*) \quad (\text{A.1})$$

Secondly, in case there is a deviation by player  $k$  at stage  $\tau + 1$ , the discounted pay-off is given by

$$v_k(\tilde{s}) = \frac{(1-\delta)}{(1-\delta^T)} \left( \sum_{t=1}^{\tau} \delta^{t-1} u_k(\alpha_k^*, \alpha_{-k}^*) + \delta^{\tau} u_k(\bar{\alpha}_k, \alpha_{-k}^*) + \sum_{t=\tau+2}^T \delta^{t-1} u_k(0, 0) \right)$$

where  $\bar{\alpha}_k < \alpha_k^*$ . We simplify the above expression as

$$\begin{aligned} v_k(\tilde{s}) &= \frac{(1-\delta)}{(1-\delta^T)} \left( u_k(\alpha_k^*, \alpha_{-k}^*) \left[ \frac{1-\delta^{\tau}}{1-\delta} \right] + \right. \\ &\quad \left. \delta^{\tau} u_k(\bar{\alpha}_k, \alpha_{-k}^*) + u_k(0, 0) \left[ \sum_{t=1}^T \delta^{t-1} - \sum_{t=1}^{\tau+1} \delta^{t-1} \right] \right) \\ v_k(\tilde{s}) &= \frac{(1-\delta)}{(1-\delta^T)} \left( u_k(\alpha_k^*, \alpha_{-k}^*) \left[ \frac{1-\delta^{\tau}}{1-\delta} \right] + \right. \\ &\quad \left. \delta^{\tau} u_k(\bar{\alpha}_k, \alpha_{-k}^*) + \left( u_k(0, 0) \left[ \frac{1-\delta^T}{1-\delta} - \frac{1-\delta^{\tau+1}}{1-\delta} \right] \right) \right) \end{aligned}$$

Applying the limit  $T \rightarrow \infty$ ,  $\delta^T \rightarrow 0$ . Further, taking the fractional term into the brackets, we get

$$v_k(\tilde{s}) = (u_k(\alpha_k^*, \alpha_{-k}^*)(1 - \delta^\tau) + (1 - \delta)\delta^\tau u_k(\bar{\alpha}_k, \alpha_{-k}^*) + u_k(0, 0)\delta^{\tau+1}) \quad (\text{A.2})$$

The discounted pay-off of the strategy with one step deviation should be lesser than the strategy with no deviation, for the later to be SPE. Therefore, we identify the  $\delta$  such that

$$v_k(s^*) > v_k(\tilde{s})$$

Inserting the values from (A.1) and (A.2), we get

$$\begin{aligned} u_k(\alpha_k^*, \alpha_{-k}^*) &> (u_k(\alpha_k^*, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)\delta^\tau + \\ &\delta^\tau u_k(\bar{\alpha}_k, \alpha_{-k}^*) - \delta^{\tau+1} u_k(\bar{\alpha}_k, \alpha_{-k}^*) + \delta^{\tau+1} u_k(0, 0)) \end{aligned}$$

By (5.10),  $u_k(\alpha_k^*, \alpha_{-k}^*) > 0$  because  $u_k(0, 0) > 0$  for sustainable points<sup>1</sup>, we get

$$0 > \delta^\tau (-u_k(\alpha_k^*, \alpha_{-k}^*) + u_k(\bar{\alpha}_k, \alpha_{-k}^*) - \delta u_k(\bar{\alpha}_k, \alpha_{-k}^*) + \delta u_k(0, 0))$$

Since  $\delta > 0$ , therefore,

$$(u_k(\alpha_k^*, \alpha_{-k}^*) - u_k(\bar{\alpha}_k, \alpha_{-k}^*) + \delta u_k(\bar{\alpha}_k, \alpha_{-k}^*) - \delta u_k(0, 0)) > 0$$

Consequently,

$$[u_k(\alpha_k^*, \alpha_{-k}^*) - u_k(\bar{\alpha}_k, \alpha_{-k}^*)] + \delta [u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(0, 0)] > 0$$

Collecting the terms and rewriting, we get

$$\delta > \frac{[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)]}{[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(0, 0)]}$$

Therefore, the player  $k$  has no incentive to deviate to any  $\bar{\alpha}_k < \alpha_k^*$  with the above discount factor condition in order to decrease its cost. Under the following sufficient condition on the discount factor

$$1 > \delta > \max_{k \in \mathcal{P}, \bar{\alpha}_k < \alpha_k^*, \bar{\alpha}_k \in \mathcal{A}_k} \frac{[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)]}{[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(0, 0)]} \quad (\text{A.3})$$

we can see that the discounted payoff  $v_k(\tilde{s})$  is less than  $v_k(s^*)$ . Additionally, it can be seen from (5.10), that  $[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)] < [u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(0, 0)]$ . Therefore,  $[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)]/[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(0, 0)] < 1$ . So the condition (A.3) does not imply any supplementary condition on system parameters. To further simplify the expression, we find the value of  $\bar{\alpha}_k$  which maximizes the expression in (A.3). Using (5.7) and (5.5), we know that  $\frac{du_k(\alpha_k, \alpha_{-k})}{d\alpha_k} < 0$ . In other words, the utility of player  $k$  monotonically decreases with increasing  $\alpha_k$  for a fixed  $\alpha_{-k}$ . We will prove that the condition in (A.3) is strictly decreasing with increasing  $\bar{\alpha}_k$ , and reaches a maximum at  $\bar{\alpha}_k = 0$ . Taking the derivative of right hand side in (A.3), we get

$$\begin{aligned} &\frac{d}{d\bar{\alpha}_k} \left\{ \frac{[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)]}{[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(0, 0)]} \right\} = \\ &\left( \frac{du_k(\bar{\alpha}_k, \alpha_{-k}^*)}{d\bar{\alpha}_k} \right) \times \frac{[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(0, 0)] - [u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(\alpha_k^*, \alpha_{-k}^*)]}{[u_k(\bar{\alpha}_k, \alpha_{-k}^*) - u_k(0, 0)]^2} \end{aligned}$$

With  $\frac{du_k(\bar{\alpha}_k, \alpha_k^*)}{d\bar{\alpha}_k} < 0$ , and the condition (5.10), the above expression is strictly negative. Hence, the lower limit of  $\delta$  in (A.3) is strictly decreasing with increasing  $\bar{\alpha}_k$ . Therefore, it maximizes at minimum value of  $\bar{\alpha}_k$  which is 0. Thus,

$$\max_{k \in \mathcal{P}, \bar{\alpha} < \alpha^*, \bar{\alpha} \in \mathcal{A}_k} \left\{ \frac{[u_k(\bar{\alpha}_k, \alpha_k^*) - u_k(\alpha_k^*, \alpha_k^*)]}{[u_k(\bar{\alpha}_k, \alpha_k^*) - u_k(0, 0)]} \right\} =$$

$$\max_{k \in \mathcal{P}} \left\{ \frac{[u_k(0, \alpha_k^*) - u_k(\alpha_k^*, \alpha_k^*)]}{[u_k(0, \alpha_k^*) - u_k(0, 0)]} \right\}$$

Inserting the values of  $u_k$  from (5.7) we get,

$$\delta > \frac{-[w_3 f_k^{COST}(0) - w_3 f_k^{COST}(\alpha_k^*)]}{[w_1 f_k^{QOS}(\alpha_k^*) + w_2 f_k^{QOE}(\alpha_k^*) - w_1 f_k^{QOS}(0) - w_2 f_k^{QOE}(0)]}$$

We have the limits of  $\delta$  as

$$1 > \delta > \delta_{asym}^{min}(\alpha_k^*, \alpha_{-k}^*)$$

where  $\delta_{asym}^{min}(\alpha_k^*, \alpha_{-k}^*)$  is given by

$$\max_{k \in \mathcal{P}} \frac{w_3 [f_k^{COST}(\alpha_k^*) - f_k^{COST}(0)]}{w_1 [f_k^{QOS}(0) - f_k^{QOS}(\alpha_{-k}^*)] + w_2 [f_k^{QOE}(0) - f_k^{QOE}(\alpha_{-k}^*)]}$$

Now we consider the the second sub-case.

(2) Let us assume that  $\bar{\alpha}_k > \alpha_k^*$ . This is a trivial case because transmitting at one stage at  $\bar{\alpha}_k$  is not a deviation from the strategy as the player still transmits at least  $\alpha_k^*$ . However, for completeness, we provide the analysis. Now, the player  $k$  conforms to the agreement point until the stage  $\tau$ . At  $\tau + 1$ , it deviates from agreement point and transmits at an  $\bar{\alpha}_k$  greater than  $\alpha_k^*$ . From  $\tau + 2$ , it again conforms to the strategy. Conforming to the strategy implies sending at  $\alpha_k^*$ . In this case, the discounted payoff without deviation, given by  $u_k(\alpha_k^*, \alpha_{-k}^*)$ , is strictly greater than the discounted payoff with deviation. This is because, at stage  $\tau + 1$ , the payoff  $u_k(\bar{\alpha}_k, \alpha_{-k}^*) < u_k(\alpha_k^*, \alpha_{-k}^*)$  for any  $\bar{\alpha}_k > \alpha_k^*$ . Further, from stage  $\tau + 2$  onwards the payoff at each stage after deviation is equal to the payoff at each stage without deviation, i.e,  $u_k(\alpha_k^*, \alpha_{-k}^*)$ . Therefore, overall, the discounted payoff without deviation is strictly greater than discounted payoff with deviation. We now consider case II.

**Case II:** We now consider the case when there has been a deviation from the agreement point  $(\alpha_k^*, \alpha_{-k}^*)$  and now both players are using the threat point  $(0, 0)$ . We would now examine if it is possible for any player to deviate from this strategy and play  $\alpha_k > 0$ . We note that  $u_k(\alpha_k, 0) < u_k(0, 0)$ . Hence, if any player deviates at any intermediate stage from  $(0, 0)$  and then conforms to the strategy by playing  $(0, 0)$ , the stage payoff at the deviation will be lesser than the stage payoff if there was no deviation. Therefore, the overall payoff without deviation would be higher than the payoff with deviation. Therefore, the player has no incentive to deviate from  $(0, 0)$ . This completes the proof.

# Bibliography

- [1] Y. Wu, “Information exchange in wireless networks with network coding and physical-layer broadcast,” 2004.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 1991.
- [3] M. Chiang, S. Low, A. Calderbank, and J. Doyle, “Layering as optimization decomposition: A mathematical theory of network architectures,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, Jan 2007.
- [4] A. Tanenbaum, *Computer Networks*, 4th ed. Prentice Hall Professional Technical Reference, 2002.
- [5] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, “Network information flow,” *Information Theory, IEEE Transactions on*, vol. 46, no. 4, pp. 1204–1216, 2000.
- [6] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [7] R. Bassoli, H. Marques, J. Rodriguez, K. Shum, and R. Tafazolli, “Network coding theory: A survey,” *Communications Surveys Tutorials, IEEE*, vol. 15, no. 4, pp. 1950–1978, Fourth 2013.
- [8] T. Matsuda, T. Noguchi, and T. Takine, “Survey of network coding and its applications.” *IEICE Transactions*, vol. 94-B, no. 3, pp. 698–717, 2011.
- [9] S. Deb, M. Effros, T. Ho, D. R. Karger, R. Koetter, D. S. Lun, M. Medard, and N. Ratnakar, “Network coding for wireless applications: A brief tutorial,” in *In IWWAN*, 2005.
- [10] C. Fragouli and E. Soljanin, “Network coding fundamentals,” *Foundations and Trends® in Networking*, vol. 2, no. 1, pp. 1–133, 2007. [Online]. Available: <http://dx.doi.org/10.1561/13000000003>
- [11] T. Ho and D. Lun, *Network Coding: An Introduction*. New York, NY, USA: Cambridge University Press, 2008.
- [12] M. Medard and A. Sprintson, Eds., *Network Coding: Fundamentals and Applications*. Academic Press, 2011.
- [13] S.-Y. Li, R. Yeung, and N. Cai, “Linear network coding,” *Information Theory, IEEE Transactions on*, vol. 49, no. 2, pp. 371–381, Feb 2003.
- [14] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, “Xors in the air: Practical wireless network coding,” *Networking, IEEE/ACM Transactions on*, vol. 16, no. 3, pp. 497–510, 2008.

- [15] C. Fragouli, J.-Y. Le Boudec, and J. Widmer, "Network coding: An instant primer," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 1, pp. 63–68, Jan. 2006. [Online]. Available: <http://doi.acm.org/10.1145/1111322.1111337>
- [16] S. Zhang, S. C. Liew, and P. P. Lam, "Hot topic: Physical-layer network coding," in *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '06. New York, NY, USA: ACM, 2006, pp. 358–365. [Online]. Available: <http://doi.acm.org/10.1145/1161089.1161129>
- [17] T. Koike-Akino, P. Popovski, and V. Tarokh, "Optimized constellations for two-way wireless relaying with physical network coding," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 5, pp. 773–787, June 2009.
- [18] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *Information Theory, IEEE Transactions on*, vol. 57, no. 10, pp. 6463–6486, oct. 2011.
- [19] C. Feng, D. Silva, and F. Kschischang, "An algebraic approach to physical-layer network coding," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7576–7596, Nov 2013.
- [20] U. Niesen and P. Whiting, "The degrees of freedom of compute-and-forward," *Information Theory, IEEE Transactions on*, vol. 58, no. 8, pp. 5214–5232, aug. 2012.
- [21] J. C. Belfiore, "Lattice codes for the compute-and-forward protocol: The flatness factor," in *Information Theory Workshop (ITW), 2011 IEEE*, Oct 2011, pp. 1–4.
- [22] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 438–460, March 2011.
- [23] C. Feng, D. Silva, and F. Kschischang, "Blind compute-and-forward," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 403–407.
- [24] S.-N. Hong and G. Caire, "Reverse compute and forward: A low-complexity architecture for down-link distributed antenna systems," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 1147–1151.
- [25] T. Yang, X. Yuan, L. Ping, I. Collings, and J. Yuan, "A new physical-layer network coding scheme with eigen-direction alignment precoding for mimo two-way relaying," *Communications, IEEE Transactions on*, vol. 61, no. 3, pp. 973–986, March 2013.
- [26] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.
- [27] A. B. MacKenzie and L. A. DaSilva, *Game theory for wireless engineers*. Morgan & Claypool Publishers, 2006, vol. 1, no. 1.
- [28] S. Lasaulce and H. Tembine, *Game Theory and Learning for Wireless Networks: Fundamentals and Applications*, 1st ed. Academic Press, 2011.
- [29] Y. E. Sagduyu and A. Ephremides, "Power control and rate adaptation as stochastic games for random access," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 4. IEEE, 2003, pp. 4202–4207.
- [30] E.-V. Belmaga, S. Lasaulce, and M. Debbah, "Power allocation games for mimo multiple access channels with coordination," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 6, pp. 3182–3192, 2009.



- [31] J. Ye and M. Hamdi, "Priority-based rate adaptation using game theory in vehicular networks," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, 2011, pp. 1–6.
- [32] P. Chaporkar, A. Proutiere, and B. Radunovic, "Rate adaptation games in wireless lans: Nash equilibrium and price of anarchy," in *INFOCOM, 2010 Proceedings IEEE*, 2010, pp. 1–9.
- [33] H. Ren and M.-H. Meng, "Game-theoretic modeling of joint topology control and power scheduling for wireless heterogeneous sensor networks," *Automation Science and Engineering, IEEE Transactions on*, vol. 6, no. 4, pp. 610–625, 2009.
- [34] M. A. Vazquez-Castro, Z. Han, A. Hjørungnes, and N. Marina, "Rate allocation for satellite systems with correlated channels based on a stackelberg game," in *Game Theory for Networks, 2009. GameNets '09. International Conference on*, 2009, pp. 638–645.
- [35] M. L. Treust and S. Lasaulce, "A repeated game formulation of energy-efficient decentralized power control," *Trans. Wireless. Comm.*, vol. 9, no. 9, pp. 2860–2869, Sep. 2010. [Online]. Available: <http://dx.doi.org/10.1109/TWC.2010.072610.091472>
- [36] C. Pandana, Z. Han, and K. Liu, "Cooperation enforcement and learning for optimizing packet forwarding in autonomous wireless networks," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 8, pp. 3150–3163, 2008.
- [37] A. Wiesel, Y. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *Signal Processing, IEEE Transactions on*, vol. 56, no. 9, pp. 4409–4418, Sept 2008.
- [38] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm i. expected complexity," *Signal Processing, IEEE Transactions on*, vol. 53, no. 8, pp. 2806–2818, Aug 2005.
- [39] R. Zamir, "Lattices are everywhere," in *Information Theory and Applications Workshop, 2009*, Feb 2009, pp. 392–421.
- [40] J. Forney, G.D., "Coset codes. i. introduction and geometrical classification," *Information Theory, IEEE Transactions on*, vol. 34, no. 5, pp. 1123–1151, Sep 1988.
- [41] F. Monteiro and I. Wassell, "Dual-lattice-aided mimo detection for slow fading channels," in *Signal Processing and Information Technology (ISSPIT), 2011 IEEE International Symposium on*, Dec 2011, pp. 502–507.
- [42] V. Tarokh, A. Vardy, and K. Zeger, "Universal bound on the performance of lattice codes," *Information Theory, IEEE Transactions on*, vol. 45, no. 2, pp. 670–681, Mar 1999.
- [43] M. Shub and S. Smale, "Complexity of bezout's theorem i: Geometric aspects," *Journal of the American Mathematical Society*, vol. 6, no. 2, pp. pp. 459–501, 1993. [Online]. Available: <http://www.jstor.org/stable/2152805>
- [44] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. Mc-Graw Hill, 1984.
- [45] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, June 2010, pp. 1022–1026.
- [46] V. Ntranos, V. Cadambe, B. Nazer, and G. Caire, "Integer-forcing interference alignment," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 574–578.
- [47] J. Zhan, B. Nazer, M. Gastpar, and U. Erez, "Mimo compute-and-forward," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, June 2009, pp. 2848–2852.

- [48] S. Gupta and M. A. Vázquez-Castro, "Physical-layer network coding based on integer-forcing precoded compute and forward," in *Wireless and Mobile Computing, Networking and Communications (WiMob), 2012 IEEE 8th International Conference on*, Oct 2012, pp. 600–605.
- [49] S. Gupta and M. A. Vázquez-Castro, "Physical layer network coding based on integer forcing precoded compute and forward," *Future Internet*, vol. 5, no. 3, pp. 439–459, 2013. [Online]. Available: <http://www.mdpi.com/1999-5903/5/3/439>
- [50] A. Sakzad, E. Viterbo, J. J. Boutros, and Y. Hong, "Phase precoded compute-and-forward with partial feedback," *CoRR*, vol. abs/1401.7074, 2014.
- [51] J. Zhu and M. Gastpar, "Asymmetric compute-and-forward with csit," *arXiv/1401.3189*, 2014.
- [52] K. Pappi, G. Karagiannidis, and R. Schober, "How sensitive is compute-and-forward to channel estimation errors?" in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 3110–3114.
- [53] M. A. Pimentel-Niño, M. A. Vázquez-Castro, and H. Skinnemoen, "Optimized ASMIRA Advanced QoE Video Streaming for Mobile Satellite Communications Systems," in *30th AIAA International Communications Satellite Systems Conference*, 2012.
- [54] A. Khreishah, C.-C. Wang, and N. Shroff, "Cross-layer optimization for wireless multihop networks with pairwise intersession network coding," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 5, pp. 606–621, 2009.
- [55] H. Seferoglu and A. Markopoulou, "Distributed rate control for video streaming over wireless networks with intersession network coding," in *Packet Video Workshop, 2009. PV 2009. 17th International*, 2009, pp. 1–10.
- [56] E. Magli, M. Wang, P. Frossard, and A. Markopoulou, "Network coding meets multimedia: A review," *Multimedia, IEEE Transactions on*, vol. 15, no. 5, pp. 1195–1212, 2013.
- [57] Y. Kim and G. De Veciana, "Is rate adaptation beneficial for inter-session network coding?" in *Military Communications Conference, 2008. MILCOM 2008. IEEE*, Nov., pp. 1–7.
- [58] M. Proebster, M. Kaschub, and S. Valentin, "Context-aware resource allocation to improve the quality of service of heterogeneous traffic," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1–6.
- [59] S. Gupta, M. A. Pimentel-Niño, and M. A. Vázquez-Castro, "Joint network coded-cross layer optimized video streaming over relay satellite channel," in *3rd International Conference on Wireless Communications and Mobile Computing (MIC-WCMC 2013): 14-16 June 2013, Valencia, Spain*, 2013.
- [60] S. Gupta and M. A. Vázquez-Castro, "Location-adaptive network-coded video transmission for improved quality of experience," *31st AIAA International Communications Satellite Systems Conference (ICSSC), Florence*, 2013.
- [61] H. Cukurtepe and I. Akgun, "Towards space traffic management system," *Acta Astronautica*, vol. 65, no. 5, pp. 870 – 878, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0094576509001532>
- [62] D. Christopoulos, S. Chatzinotas, and B. E. Ottersten, "User scheduling for coordinated dual satellite systems with linear precoding," *CoRR*, vol. abs/1211.5888, 2012.

- [63] K. Larson and E. Al, "Spectrum Policy Task Force: Report of the Interference Protection Working Group," FCC, Tech. Rep., 2002. [Online]. Available: <http://transition.fcc.gov/sptf/files/IPWGFinalReport.pdf>
- [64] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *Communications Surveys Tutorials, IEEE*, vol. 11, no. 1, pp. 116–130, 2009.
- [65] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Comput. Netw.*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006. [Online]. Available: <http://dx.doi.org/10.1016/j.comnet.2006.05.001>
- [66] P. Ren, Y. Wang, Q. Du, and J. Xu, "A survey on dynamic spectrum access protocols for distributed cognitive wireless networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, pp. 1–21, 2012. [Online]. Available: <http://dx.doi.org/10.1186/1687-1499-2012-60>
- [67] X. Hong, C.-X. Wang, and J. Thompson, "Interference modeling of cognitive radio networks," in  *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, 2008, pp. 1851–1855.
- [68] A. Rabbachin, T. Quek, H. Shin, and M. Win, "Cognitive network interference," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 2, pp. 480–493, 2011.
- [69] K. Ruttik, K. Koufos, and R. Jantti, "Computation of aggregate interference from multiple secondary transmitters," *Communications Letters, IEEE*, vol. 15, no. 4, pp. 437–439, 2011.
- [70] A. Rabbachin, T. Quek, H. Shin, and M. Win, "Cognitive network interference- modeling and applications," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1–6.
- [71] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 5, pp. 1813–1827, 2006.
- [72] M. A. Vázquez-Castro, D. Belay-Zeleke, and A. Curieses-Guerrero, "Availability of systems based on satellites with spatial diversity and haps," in *IEE Electronic Letters, Vol. 38 No 6*, pp. 286–288, March 2002.
- [73] S. Gupta, M. A. Vázquez-Castro, and R. Alegre-Godoy, *Cooperative and Cognitive Satellite Systems*. Elsevier North-Holland, Inc., 2014, ch. Cognitive Dual Satellite Systems.
- [74] J. Mitola and J. Maguire, G.Q., "Cognitive radio: making software radios more personal," *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13–18, 1999.
- [75] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 2, pp. 201–220, 2005.
- [76] F. Vatalaro, G. Corazza, C. Caini, and C. Ferrarelli, "Analysis of leo, meo, and geo global mobile satellite systems in the presence of interference and fading," *Selected Areas in Communications, IEEE Journal on*, vol. 13, no. 2, pp. 291–300, 1995.
- [77] K. Sharma, S. Chatzinotas, and B. Ottersten, "Inline interference mitigation techniques for spectral coexistence of geo and ngeo satellites," *to be submitted to Int. J. Satellite Commun. and Networking, under preparation*, 2013.
- [78] S. Pratt, R. Raines, C. Fossa, and M. A. Temple, "An operational and performance overview of the iridium low earth orbit satellite system," *Communications Surveys Tutorials, IEEE*, vol. 2, no. 2, pp. 2–10, 1999.

- [79] P. Noschese, S. Porfili, and S. Di Girolamo, "Ads-b via iridium next satellites," in *Digital Communications - Enhanced Surveillance of Aircraft and Vehicles (TIWDC/ESAV), 2011 Tyrrhenian International Workshop on*, 2011, pp. 213–218.
- [80] M. A. Vázquez-Castro, F. Perez-Fontan, and S. R. Saunders, "Shadowing correlation assessment and modeling for satellite diversity in urban environments," *International Journal of Satellite Communications*, vol. 20, no. 2, pp. 151–166, 2002. [Online]. Available: <http://dx.doi.org/10.1002/sat.718>
- [81] M. Rieche, D. Arndt, A. Ihlow, and G. Del Galdo, "Image-based state modeling of a land mobile dual satellite system," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2013.
- [82] B. Wang and K. Liu, "Advances in cognitive radio networks: A survey," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 1, pp. 5–23, 2011.
- [83] S. Kandeepan, L. De Nardis, M. Di Benedetto, A. Guidotti, and G. Corazza, "Cognitive satellite terrestrial radios," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, 2010, pp. 1–6.
- [84] S. Sharma, S. Chatzinotas, and B. Ottersten, "Satellite cognitive communications: Interference modeling and techniques selection," in *Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC), 2012 6th*, 2012, pp. 111–118.
- [85] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "Cognitive beamhopping for spectral coexistence of multibeam satellites," in *Future Network and Mobile Summit (FutureNetworkSummit), 2013*. IEEE, 2013, pp. 1–10.
- [86] J. E. Barcelo-Llado, M. A. Vázquez-Castro, L. Jiang, and A. Hjørungnes, "Distributed power and carrier allocation in multibeam satellite uplink with individual sinr constraints," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, 2009, pp. 1–6.
- [87] L. Jiang and M. A. Vázquez-Castro, "Interference management versus interference cancellation: Satcom case," in *PSATS*, 2011, pp. 260–273.
- [88] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "Interference alignment for spectral coexistence of heterogeneous networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, p. 46, 2013. [Online]. Available: <http://jwcn.eurasipjournals.com/content/2013/1/46>
- [89] P. Jain and M. A. Vázquez-Castro, "Subspace interference alignment for multibeam satellite communications systems," in *Advanced satellite multimedia systems conference (asma) and the 11th signal processing for space communications workshop (spsc), 2010 5th*, 2010, pp. 234–239.
- [90] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the k -user interference channel," *Information Theory, IEEE Transactions on*, vol. 54, no. 8, pp. 3425–3441, 2008.
- [91] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "Frequency packing for interference alignment-based cognitive dual satellite systems," in *Vehicular Technology Conference, VTC 2013-Fall, Las Vegas, USA*, September 2013.
- [92] S. Kaoprakhon and V. Visoottiviset, "Classification of audio and video traffic over http protocol," in *Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on*, 2009, pp. 1534–1539.

- [93] G. Kassem, I. Ahmad, F. Hameed, and A. Zakariyya, "Tcp variants: An overview," in *Computational Intelligence, Modelling and Simulation (CIMSIM), 2010 Second International Conference on*, 2010, pp. 536–540.
- [94] M. Proebster, M. Kaschub, and S. Valentin, "Context-aware resource allocation to improve the quality of service of heterogeneous traffic," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1–6.
- [95] M. A. Pimentel-Nino, M. A. Vázquez-Castro, and H. Skinnemoen, "Optimized ASMIRA Advanced QoE Video Streaming for Mobile Satellite Communications Systems," in *30th AIAA International Communications Satellite Systems Conference*, 2012.
- [96] A. ParandehGheibi, M. Medard, S. Shakkottai, and A. Ozdaglar, "Avoiding interruptions - qoe trade-offs in block-coded streaming media applications," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, 2010, pp. 1778–1782.
- [97] S. Khan, Y. Peng, E. Steinbach, M. Sgroi, and W. Kellerer, "Application-driven cross-layer optimization for video streaming over wireless networks," *Communications Magazine, IEEE*, vol. 44, no. 1, pp. 122–130, 2006.
- [98] R. Mok, E. Chan, and R. Chang, "Measuring the quality of experience of http video streaming," in *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*, 2011, pp. 485–492.
- [99] S. Gupta, E. V. Belmega, and M. A. Vázquez-Castro, "A game theoretical analysis of qoe-driven adaptive video transmission," *Submitted to Springer Journal on Multimedia Systems.*, Nov 2013.
- [100] —, "Game-theoretical analysis of the tradeoff between qoe and qos over satellite channels," in *Submitted to ASMS/SPSC*, 2014.
- [101] E. Belmega, L. Sankar, and H. Poor, "Repeated games for privacy-aware distributed state estimation in interconnected networks," in *Network Games, Control and Optimization (NetGCooP), 2012 6th International Conference on*. IEEE, 2012, pp. 64–68.
- [102] T. A. Trinh and S. Molnár, "A game-theoretic analysis of tcp vegas," in *Quality of Service in the Emerging Networking Panorama*. Springer, 2004, pp. 338–347.
- [103] Z. Liu, X. Zhang, D. Wang, Y. Zhao, and H. Guan, "A game-theoretic analysis of tcp vegas and fast tcp," *IET*, 2008.
- [104] M. Handley, S. Floyd, J. Padhye, and J. Widmer, "Tcp friendly rate control (tfr): Protocol specification," *RFC 3448 Proposed Standard*, 2003.